Assessing the influence of phonetic variation on the perception of spoken threats

James Andrew Tompkinson

PhD

University of York
Language and Linguistic Science

October 2018

Abstract

In spite of the belief that there is such a thing as a 'threatening tone of voice' (Watt, Kelly and Llamas, 2013), there is currently little research which explores how listeners infer traits such as threat from speakers' voices. This thesis addresses the question of how listeners infer traits such as how threatening speakers sound, and whether phonetic aspects of speakers' voices can play a role in shaping these evaluations. Additionally, it is sometimes the case that a victim of a crime will never see the perpetrator's face but will hear the perpetrator's voice. In such cases, attempts can be made to get the witness or victim to describe the offender's voice. However, one problem with this is whether phonetically untrained listeners have the ability to accurately describe different aspects of speakers' voices. This issue is also addressed throughout this thesis.

Over five experiments, this thesis investigates the influence of a range of linguistic and phonetic variables on listeners' evaluations of how threatening speakers sounded when producing indirect threat utterances. It also examines how accurately phonetically-untrained listeners can describe different aspects of speakers' voices alongside their evaluative judgements of traits such as threat and intent-to-harm. As well as showing that a range of linguistic and phonetic variables can influence listeners' threat evaluations, results support the view that caution should be adopted in over-reliance on the idea that people will "know a threat when they hear one" (Gingiss, 1986:153). This research begins to address the phonetic basis for the perceptual existence of a 'threatening tone of voice', along with how listeners evaluate and describe voices in earwitness contexts. Suggestions are made at the end of the thesis for improvements in the elicitation and implementation of accurate, meaningful information about speakers' voices from linguistically-untrained listeners in evaluative settings involving spoken threats.

Contents

List of Tables

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Acknowledgements

I could not have completed this thesis without support from so many people. Firstly, I cannot express enough gratitude for the guidance provided by my supervisor, Dr Dominic Watt. Thank you for first introducing me to the idea of studying this topic and for being a sounding board for advice ever since. Thank you for helping me become a better academic, for supporting me through every stage of my PhD, and for your amazing proofreading and phrasing skills. I'm not sure we'll ever agree about whether 'data' should be used as a singular noun, but the knowledge I've gained from you about shaping ideas and arguments has been invaluable. I simply could not have asked for a better mentor.

I must also thank my Thesis Advice Panel member, Dr Carmen Llamas, for her expert guidance and support throughout my PhD studies. Carmen has always provided a perfect blend of supportive and insightful feedback which has undoubtedly helped this thesis be the best that it could be. Thanks also to Professor Peter French, Professor Paul Foulkes and Dr Vince Hughes for continuing to teach me about the analysis of speech and audio for forensic purposes over the last few years, and for helping me to become a fully integrated member of the York FSS community. I am also hugely grateful to Vince and Professor Tim Grant for filling my viva examination with engaging discussions and useful feedback which has unquestionably improved the presentation of the work in this thesis.

One of the best aspects of my PhD studies has been the collaborative projects I have been involved in with researchers in the University of York's Department of Psychology. Thanks to Dr Mila Mileva and Professor Mike Burton for all the side projects we have worked on, and for helping me discover just how terrible my face recognition skills are! I have also been lucky enough to be given some great teaching opportunities over the last three years and have benefited from outstanding mentorship in this area. Special thanks to Dr Ghada Khattab for taking a chance on me and giving me my first lecturing job at Newcastle University, and to Professor Richard Ogden and Dr Marton Soskuthy for supporting me in my York-based teaching.

Enormous thanks must go to the Economic and Social Research Council for providing me with a 1+3 award to pursue my postgraduate studies at York. This thesis would not have been possible without the financial support provided by the ESRC. I was also lucky enough to benefit from an ESRC Overseas Institutional Visit during my PhD, and am indebted to Professor Tammy Gales for hosting me at Hoftstra University for five weeks in 2017. From providing me with new ideas and alternative perspectives on my thesis to taking me for my first elk burger, Tammy's mentorship and friendship during my New York visit is something I will be forever grateful for.

Finally, the real heroes behind this thesis are my family and friends who have stoically put up with me over the last few years. Thanks to both my parents for their support and encouragement to pursue this PhD. And Dad, I'm so happy and proud that you made it this far through your cancer and transplant battle to see me progress through to this stage. Thanks also to my brother Dave and best friend Dan for being there for those essential non-academic chats! And finally, the biggest thank you goes to my fiancée Holly. You mean more to me than you will ever know. Thank you for being with me on every step of this journey and for providing me with the day-to-day love and support that I needed to make it to the end.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. Supervision for this project has been provided by Dr Dominic Watt, with additional feedback from Dr Carmen Llamas. This work has not previously been presented for an award at this, or any other, University. This research was funded by the Economic and Social Research Council (ESRC) through the White Rose Doctoral Training Partnership (WRDTP). All sources are acknowledged as references.

Some of the work in Chapter 6 has already been published in the following paper:

Tompkinson, J., and Watt, D. (2018). Assessing the abilities of phonetically untrained listeners to determine pitch and speaker accent in unfamiliar voices. *Language and Law=Linguagem e Direito*, 5(1), 19-37.

Some of the ideas discussed in Chapter 2 were published in the following paper. I was responsible for producing the arguments relevant to this thesis:

Mileva, M., Tompkinson, J., Watt, D., and Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of experimental psychology: Human perception and performance*, 44(1), 128-138.

Parts of the thesis introduction presented in Chapter 1 have also been published in the following article:

Tompkinson, J. (2016). Bomb scares: Can you judge a threat from the voice on the phone? *Article published in The Conversation*. Retrieved from https://theconversation.com/bomb-scares-can-you-judge-a-threat-from-the-voice-on-the-phone-60073 [Accessed 5th September 2018].

Chapter 1 - Introduction

1.1. Introduction

In 2016, approximately 20 secondary schools in various locations across the United Kingdom received threatening phone calls from anonymous callers who stated that there were bombs inside the school buildings. One particular school in Cambridgeshire was told by the caller that the bomb would "take children's heads off" (Sykes and Perring, 2016). Upon receipt of the threats, the schools were evacuated, resulting in widespread panic and disruption for both staff and students – some of who were scheduled to take GCSE and A-Level exams at the time. These types of incidents raise many questions for those interested in forensic language analysis. Arguably the most important of these would be to outline what useful investigative information could be obtained from both the words spoken by a given threatener and the threatener's voice. This is the primary issue addressed through the research presented in this thesis.

It is difficult to provide an accurate estimate of exactly how many people are prosecuted or investigated for threat crimes each year, given that there are many different types of threats and that threats can appear as a facet of lots of different types of crimes. However, the Crime Survey for England and Wales for 2018 (Office for National Statistics, 2018) states that a total number of 27,025 threats to kill were reported to police between April 2017 and March 2018, and that this figure represented an increase of 5,276 on the total number recorded for the previous year. Rather alarmingly, the number of reported threats to kill in the 2009 Crime Survey for England and Wales was 9,448, which means an increase of 17,577 reported threats to kill in two single years which are less than a decade apart. While this may reflect a greater degree of reporting of threat crimes, alongside an increase in frequency, the trend would provide evidence

to suggest that the numbers of serious threats been dealt with by the police in the United Kingdom is growing.

According to Shuy (1993:97), threats are one of the "most negatively received" speech acts. Storey (1995:74) further notes that the linguistic complexity surrounding threats is often overlooked when they are discussed or analysed. The goal of this thesis is to present a body of work examining how listeners perceive threat, intent-to-harm and other traits from the voices of speakers they are exposed to. In doing so, it aims to address one area of potential complexity surrounding this, as-of-yet, under-researched type of spoken language crime.

There are myriad reasons why a speaker may threaten someone, including to show anger, intimidate a hearer, get help, show seriousness of purpose, warn, harass, frighten, alarm or manipulate a hearer (Fraser, 1998:160; Douglas et al., 2013:369). Threats are of particular interest to those working in the legal process owing to the fact that they can both serve as standalone crimes and form part of other serious crimes such as robbery and extortion (Yamanaka, 1995:38). Additionally, while threats are not a mandatory feature of serious crimes, Solan and Tiersma (2015:224) state that threats are often used to accomplish such offences, with Greenawalt (1989:92) further pointing out that criminal acts frequently involve threats which aim to get an innocent victim to commit to an unfavourable course of action.

Of course, not all threats express criminal intent. Consider a mother who threatens her child that their favourite toy will be taken away unless the child puts their shoes on and leaves the house quickly. Although this interaction is not illegal, a clear threat is made by the mother towards her child. Another example of an authentic but non-illegal threat,

highlighted by Solan and Tiersma (2015:223), is of a boss threatening an employee that they will be fired should they make another rude gesture at a customer. Again, a threat would have been made in a case such as this, but it would usually fall within the boundaries of the law. Storey (1995:74) goes as far as to say that threats are simply "a way of life", with Milburn and Watman (1981:2) adding that they provide speakers with a way of exerting personal and social control in unpredictable situations or environments. However, as Fraser (1998:160) highlights, some threats are illegal, and these are therefore suitable for analysis by linguists working in the forensic domain. Solan and Tiersma (2015:233) state that threats become illegal when they are designed to achieve certain, often criminal, goals. Threats can also be illegal if they are directed towards certain people, such as the President of the United States or members of the United Kingdom's royal family (Solan and Tiersma, 2015:233).

The study of threatening language for forensic purposes can be considered to fall under the branch of "investigative forensic linguistics" (Larner, 2015:132). However, it arguably also transcends the boundary between investigative and "descriptive forensic linguistics" (Larner, 2015:132) depending on the question being asked. Investigative forensic linguistics involves both the analysis of language crimes such as threats, and the use of linguistic analysis to assist law enforcement agencies with either investigative or evidential matters. This contrasts with descriptive forensic linguistics, which instead focusses on analysis of language use in the legal system, including the language used within courtrooms, police interviews, Language Analysis for the Determination of Origin (LADO) interviews during the asylum process, and the analysis of the meaning of specific terms or phrases which may be legally relevant or consequential.

Despite the potential for verbal threats to function as serious language crimes (Shuy, 1993), Gales (2016:3) warns of both a current lack of understanding about what threatening language "actually is", and of the potential dangers when those tasked with assessing linguistic aspects of threats rely on personal or stereotypical assumptions rather than findings from empirical research. This potential problem is further compounded when the modality of a threat is spoken rather than written, owing to a current shortage of research examining how spoken threats are perceived by listeners (Watt, Kelly and Llamas, 2013).

While drawing on many of the same analytical techniques used by forensic phoneticians in their everyday work, the analysis in this thesis is predominantly focussed on how people without advanced qualifications in phonetics or linguistics evaluate and describe the voices of unfamiliar speakers with respect to threats. Following the work of Griffiths (2012), it is argued that attempting to understand how those without linguistic training describe and evaluate voices is an important consideration for professional linguists working on the analysis of forensic data. The vast majority of police officers and lawyers do not have any formal linguistic or phonetic training, nor do members of jury panels who are required to analyse, interpret and evaluate evidence in cases involving language crimes such as threats. In such instances, it can be argued that it is the job of forensic linguists and forensic phoneticians to assist courts by providing analytical skills which are beyond the abilities of non-linguists. Additionally, it can also be argued that it is incumbent on such professionals to better understand the motivations upon which which non-linguists base their evaluations about linguistic evidence which may present at any stage of the legal process. A 2015 report produced by the UK Parliamentary Office of Science and Technology entitled "Forensic Language Analysis" (Parliamentary Office of Science and Technology, 2015)

highlighted the issue of a disjointed relationship between jurors and linguistic experts, stating that "jurors expect certain procedures to be possible which experts assert are not, such as personality analysis, determining truth and falsity, and assessing threat in speech intonation (although this is a research interest)" (Parliamentary Office of Science and Technology, 2015:3). The last point in this list highlights both an unrealistic expectation on the part of linguists by jurors, alongside an underlying belief on the part of non-linguists that aspects of voice can be used to determine threat. A core goal of the research presented in this thesis is to explore whether there is any phonetic basis for these kinds of underlying beliefs. In attempting to gain a better understanding of both how listeners infer threat in speakers' voices, and of the other types of inferences that listeners are willing to make about others based on aspects of speakers' voices, it is hoped that this research will help support the work of forensic linguists tasked with assisting in cases involving either spoken threats, voice descriptions or a combination of the two.

There are frequent examples from both forensic phonetic casework and the media which highlight the need for further research into listener evaluations of spoken threats, and of speakers' voices more generally. One such case is documented in Watt, Kelly and Llamas (2013) and comes from a 2012 crown court trial in Middlesbrough, UK, during which the defendant was accused of reiterating a previously unrecorded threat to kill a judge by uttering the words "*I will do summat* [a northern English dialect term for 'something'] *about it when I get out and it won't be with guns or anything like that*". This utterance was produced following a situation where the defendant had been held in a police cell and was to a custody officer that he wanted to be released. This custody officer was the hearer of the alleged reiteration of the previous unrecorded threat to kill.

This case provides one example of what Gales (2010) terms an indirect threat, where a threat is judged to have been uttered, yet the wording of the utterances does not explicitly signal intent-to-harm on the part of the speaker. In the example above, the vague nature of the phrase "*I will do summat about it*" meant that listener interpretation was required to determine what that 'something' was, and by extension whether the speaker had criminal intentions or not. But from the wording alone, as Watt, Kelly and Llamas (2013) point out, the speaker could have simply been signalling his intention to launch a formal complaint, or to write to his local Member of Parliament to voice his displeasure at being kept in a police cell for longer than he deemed was necessary. The interpretation that "*I will do summat about it when I get out and it won't be with guns or anything like that*" constituted a serious threat would require listener inference of the speaker's intentions. The speaker's words in this particular case, if taken in their most literal interpretation, specifically ruled out the use of guns or similar weapons, and yet the utterance was still interpreted as a reiteration of a serious death threat. Watt, Kelly and Llamas (2013) also point out that during the subsequent trial, the custody officer's testimony identified that the defendant's behaviour, the surrounding context and the fact that he used an aggressive tone of voice, served as evidence which supported the interpretation of the utterance in question as a serious death threat.

Another example of aspects of voice influencing a trial involving spoken threats is taken from the Danish Supreme Court (case number U.2016.1939H - TfK2016.491H)[1]. In this case, a man was accused of threatening to cut a fellow employee's throat. As part of the defence offered in this case, the accused threatener stated that because he has a low-pitched voice, he is often perceived as sounding angry. The translated and original text from the court report is produced below:

[1] Many thanks to Tanya Karoli Christensen for providing relevant background information and translations on this particular case.

English

The defendant is very careful with how he phrases things since he is sometimes misunderstood and perceived as angry because he has a very deep voice. He never raises his voice since nothing good comes from it anyway. He can, however, be somewhat direct in his demeanour.

Danish

Tiltalte passer meget på, hvordan han formulerer sig, idet han sommertider bliver misforstået og opfattet som sur fordi han har en meget dyb stemme. Han hæver aldrig stemmen, da man sjældent får noget ud af det alligevel. Han kan dog somme tider godt være lidt kontant i sin fremtræden.

English to Danish translations were provided by Dr Tanya Karoli Christensen, Associate Professor of Nordic Studies and Linguistics, University of Copenhagen.

The interesting aspect of this case is that the defendant's perception of his own voice was offered as a mitigating circumstance in court in an attempt to absolve responsibility for committing a threat crime. Furthermore, throughout the case, the defendant was described by the hearer of the threat as sounding both angry and frustrated. Ultimately in this case, the defendant was found guilty and sentenced to a fine and 30 days imprisonment. However, the potentially complex relationship between language perceptions and threats is highlighted by this particular trial, and this is something that will be further explored through the research in this thesis.

A further example of lay-listener voice evidence being used in a serious criminal investigation is taken from a series of aggravated burglaries which took place in 2018 in the South East of England. In January 2018, the media reported repeated incidents where a masked intruder broke into properties, physically assaulted victims and robbed them of high-value possessions such as jewellery (BBC News, 2018). When asked to provide a description of the perpetrator, one victim described him as follows: *"I would say he spoke well, he had no accent, he didn't have bad grammar, he's an intelligent man, he knows how to assess the situation and carry this out."* Examples of this kind illustrate some of the difficulties that witnesses may have when asked to provide linguistically precise descriptions of the speech of a criminal who provided few or no other useful clues to identity, e.g. from his face, while the offence was in progress. Under such circumstances, the description of the offender's voice provided by the witness may become a highly valuable source of evidence. Examining the description provided by the victim in this particular case, there was an assessment of the speaker's accent, the speaker's intelligence and both the speaker's ability and intention to carry out a violent act. All of this information was obtained from a combination of both the speaker's voice and the situational context. However, more research is arguably needed to gain a better understanding of the accuracy of such descriptions, and of the linguistic and/or phonetic factors that could motivate an earwitness to reach conclusions about factors like a speaker's emotional state or their intention to carry out a given act. Attempting to understand how such decisions integrate with contextual information is also important.

In an attempt to address the current lack of phonetic and linguistic research in these areas, this thesis explores how linguistically-untrained listeners evaluate the voices of unfamiliar speakers producing spoken threat utterances. The aim is to explore, using

empirical data, how different phonetic aspects of voice may cause listeners to infer greater or lesser levels of conveyed threat on the part of a given speaker. Related to this, the research presented in this thesis also examines how listeners who do not have any advanced linguistic or phonetic training describe the voices of unfamiliar speakers producing spoken threats. By combining these two strands of research, it is hoped that the as-of-yet unexplored area of the phonetic cues to inferred threat will be more comprehensively understood. It is also hoped that research presented in the following chapters could help to facilitate an improvement in how anonymous spoken threats are evaluated and assessed by those faced with such tasks in their everyday professional lives.

The overarching research questions addressed in this thesis are as follows:

1. What is the relationship between measurable phonetic aspects of speakers' voices, such as pitch, speaker accent, voice quality, tempo and emphasis patterns, and listeners' inferences of how threatening a given speaker sounds?

2. How successfully can listeners describe the phonetic aspects of speakers' voices detailed in question 1, above?

3. What is the relationship between listeners' own perceptions of certain aspects of speakers' voices and judgements made about those speakers with respect to the inference of traits such as threat and intent-to-harm?

By addressing these three research questions, an additional aim of the thesis is to provide a body of research aimed at improving understanding about threats as a type of language crime.

1.2. Thesis overview

Following this general introduction, Chapter 2 presents background research relevant to the thesis. It firstly considers the issue of threats as speech acts by outlining previous commentary on what constitutes a threat, before subsequently presenting an evaluation of the relative roles of both the speaker and the hearer in the successful communication of a spoken threat. The existing body of literature examining various linguistic aspects of threats is then discussed. Existing linguistic research in this area has primarily focussed on written threats as opposed to spoken threats, with research on how aspects of a speaker's voice can be used to convey greater or lesser levels of threat to harm being comparatively sparse. Following this, the discussion in Chapter 2 considers the legal status of threats in both England and Wales, alongside certain other overseas jurisdictions, before turning to a discussion of how different aspects of speakers' voices can be associated with a range of different emotional and affective speaker states. Background research in this area was used to identify, and formulate hypotheses about, these aspects of voice that may influence listeners' perceptions of how threatening speakers sounded. This was necessary given the absence of an existing body of work exploring this issue in a direct way. Research on how linguistically-untrained listeners evaluate the voices of unfamiliar speakers is then presented, and the research review ends by formulating a restricted working framework for the linguistic analysis of spoken threats that is subsequently adopted for the remainder of the work presented in the thesis.

Chapter 3 presents two initial experiments which aim to address how phonetic variables may influence listeners' evaluations of spoken threats. Both of these experiments are exploratory in nature and were conducted to form a basis for empirically-driven hypotheses which could be developed through the remainder of the thesis. Using the 'frequency code' hypothesis (Ohala, 1984) as a basis for the expectation that lower-pitched vocalisations may cause listeners to perceive a more dominant, threatening and aggressive speaker, Experiment 1 examines the relative effects of average fundamental frequency (F0) on threat evaluations provided by a group of listeners. Following research showing that a speaker's accent can shape listeners' evaluations in legally-relevant settings (Dixon, Mahoney and Cocks, 2002; Dixon and Mahoney, 2004), the effect of speaker accent is also explored alongside F0 in Experiment 1. Additionally, Experiment 1 also examines the potential associations between phonetic variables, perceived threat and judgements made about speakers' body size in the absence of visual cues. This work was conducted to further assess any 'frequency code' associations within the data and how they may link to perceptions of how threatening speakers sound. Experiment 2 presents an extension to the work of Watt, Kelly and Llamas (2013) by using utterances in unfamiliar languages as experimental stimuli, and assessing the relative influence of average F0, F0 range and speech rate on listeners' evaluations of how threatening speakers sounded.

Chapter 4 presents Experiment 3, which aims to build on the work in Chapter 3 by considering the role of phonation quality in listeners' evaluations of how threatening speakers sounded. This experiment also assesses the influence of providing two contrasting contextual environments for listeners' assessments, given the relative importance of context in threat evaluations. In this experiment, listeners were presented with utterances in an unfamiliar foreign language and were instructed to assess how

threatening the speaker sounded in each case. The utterances contained five separate phonation qualities (*modal*, *falsetto*, *harsh*, *creaky* and *whispery*), and listeners were assigned to one of two context groups; either a group in which they were told that the utterances they would hear were bomb threats targeted at a local football stadium, or a group in which they were given no contextual information about the utterances. The experiment in Chapter 5 also explores how listeners describe the voices that they heard using an adapted version of the National Counter Terrorism Security Office (NCTSO) bomb threat checklist (NCTSO, 2016). This document is designed to elicit useful information about a speaker from their voice for use for investigative purposes following the receipt of a bomb threat. The document includes a section which instructs listeners to describe aspects of the voice of the speaker they heard, including descriptors related to pitch, speaking tempo, disfluencies and voice quality.

Chapter 5 builds on the work presented in Chapters 3 and 4 by examining listeners' evaluations of the voices of multiple speakers producing the indirect bomb threat utterance *"I'm warning you about a bomb at York Station, which will go off this afternoon"*. The analysis in this chapter assesses whether phonetic realisations of specific individual tokens can act as markers of speakers aiming to produce utterances in what they considered to be a 'threatening tone of voice' compared with utterances produced in what the speaker considered to be a 'neutral tone of voice'. This is examined with respect to both differences in speakers' productions and differences in listeners' perceptions of the stimuli. The work in Chapter 5 also further extends the research presented in Chapter 4 which evaluates how listeners describe the voices of unfamiliar speakers. It assesses the accuracy of listeners' judgements of vocal pitch and speech rate with respect to measured values for average F0 and the average number of syllables produced per second of speech. Unconstrained descriptions of speakers' voices

are also analysed to assess their potential usefulness in investigative work which requires linguistically-untrained listeners to provide descriptions of a threatener's voice. More generally, the work in Chapter 5 also begins to address the issue of how separate listeners' descriptions of a threatener's voice are from their evaluations of how threatening that speaker sounds. It also provides data to critically assess areas of potential weakness in the NCTSO bomb threat checklist, and offers a view about how such a document may be amended to obtain more linguistically-accurate information from those tasked with using it.

Chapter 6 attempts to collate and develop the findings of the research presented in Chapters 3, 4 and 5. The research in Chapter 6 analyses the effect of multiple linguistic and phonetic variables on listeners' evaluations of how threatening a given speaker sounds, alongside judgements of how much intent-to-harm was conveyed through speakers' utterances. Variables considered in this chapter include average F0 and F0 range, alongside speaker accent, the utterance spoken and whether primary emphasis was placed on the modal verb 'will' or not. Additionally, the effect of listeners' own evaluations of how high-pitched speakers' voices were, using a gradient scale, was also considered. In analysing all of these variables within a single experiment that contained multiple listeners and multiple speakers, the work in Chapter 6 aims to further investigate the underlying factors behind listeners' judgements about how threatening speakers sound and how much intent-to-harm they conveyed through their speech. In addition to this, the work in Chapter 6 also considers the accuracy of listeners' judgements of two specific aspects of voice: pitch and speaker accent. Both of these vocal feature serve as voice description options on the NCTSO bomb threat checklist and were therefore considered worthy of investigation within an experiment assessing listeners' responses to spoken threats. Following from the work in Chapter 5, the study

in Chapter 6 also assesses whether using a gradient scale to elicit listener judgements of how high-pitched a speaker's voice would induce more accurate responses than the check-box system adopted by the NCTSO bomb threat checklist. It also assesses whether listeners were able to accurately classify three different accents of English, including Standard Southern British English (SSBE), Northern Irish English and foreign-accented speakers of English.

Finally, Chapter 7 presents a general discussion of the results in light of the research questions set out in Section 1.1. It brings together the individual pieces of research outlined in each of the experimental chapters and assesses the overall usefulness of the findings in advancing theoretical knowledge and understanding of spoken threats. The work in Chapter 7 also addresses certain practical implications for how research of this kind can aid those tasked with working with evidence provided by lay-witness voice analysis, evaluations and descriptions in order to aid the delivery of justice at every stage of the legal process, focussing on instances involving language crimes such as spoken threats. To conclude, an overview of the key findings are once again considered, alongside suggested directions for further research.

Chapter 2 – Research review

2.1. Introduction

This chapter presents background research relevant to the analyses conducted throughout this thesis. Various definitions of threats are initially considered, with research examining threats under a speech act framework subsequently explored. Previous linguistic and legal analysis of various aspects of threats is presented, along with research on phonetic variation, affective speech and tone of voice. Finally, the chapter builds on existing research to outline a framework for the analysis of spoken threats in a criminal context which will be adopted through the experiments and analysis presented in this thesis.

2.2. Defining and classifying threats

2.2.1. Types of threats

An initial distinction can be made between specific threats communicated by a threatener towards a target, and the general threat posed by a person, organisation or institution towards others. Meloy et al. (2013:3) highlight this difference, stating that 'threat' can refer to both "the perceived possibility of harm" and "a statement conveying an intention to cause harm (i.e., a menacing utterance)". Linguistic research on threats has predominantly focussed on the second definition by examining aspects of the language used in a threatening utterance. In such cases, the goal of the analyst can either be to infer clues to speaker intention through linguistic means (see, for example Gales, 2015), or to examine different linguistic properties of threatening utterances (see, for example, Fraser 1998). The term 'communication threat' is adopted by Douglas et al. (2013:367) to refer to an attempt to inflict psychological harm or distress on a particular

target by a threatener through a statement of intent. Illegal communication threats are categorised as 'non-lethal crimes' (Douglas et al., 2013:367). These threats are placed as part of a series of offences where there is no physical contact between the offender and the victim but potentially strong and damaging psychological trauma inflicted as a consequence of the threat (Douglas et al., 2013).

Shuy (1993) classifies a threat as a type of language crime, and places threats as part of a group of other language crimes which includes bribery, extortion, defamation, perjury, impersonation and incitement to racial hatred. However, threats can also be made through non-verbal signals such as drawings, gestures or body movements. For example, Douglas et al. (2013:373) discuss a case involving a hospital patient who repeatedly and silently greeted his nurse with direct eye contact followed by a hand motion which resembled firing a gun. Such cases are problematic for the classification of threats as language crimes, owing to a debate over whether forms of non-verbal communication can be subcategorised under the umbrella term of 'language'. Whilst acknowledging the existence of threats communicated via a non-verbal means, the focus of this thesis is to examine threats as a means of verbal communication. Shuy's (1993) definition of threats as language crimes holds for such verbal threats. It is also noted here that verbal threats can refer to threats communicated through both writing and speech.

2.2.2. Threats as speech acts

One framework which has been used to analyse threats is Speech Act Theory (Austin, 1962; Bach and Harnish, 1979; Searle, 1979). Verbal threats have been classified as illocutionary speech acts which are intentionally designed by speakers to send a given message (Fraser, 1998:160). Threats have been further defined as 'situation-altering'

utterances owing to the fact that they do not state factual information, but are instead designed to bring about a change in the world or achieve a specific purpose (Greenawalt, 1989:58).

Gales (2010) identifies two main types of verbal threats: direct and indirect. A direct threat makes clear the permutations that may arise as a result of the threatened action. Direct threats state that something unfavourable will happen and potentially also include information about the time, place and people that will be involved in the threatened action. By contrast, indirect threats do not overtly make clear that a threat is being made, and could, on wording alone, be classified as other types of speech acts including warnings, insults, complaints or promises. Any type of sentence can form an indirect threat as the speaker is under no obligation to reveal specific information about the threatened action (Fraser, 1998:168). Both indirect and direct threats can also be worded conditionally. These conditional threats are created through the incorporation of an if-clause into the design of the threat (Gales 2010:9). Milburn and Watman (1981:11) highlight that conditional threats most commonly take the form; "if you don't do X, then I will do Y". They argue that for these threats, the probability of misunderstanding is low because a high level of clarity exists over the attitude or position expressed by the speaker (Milburn and Watman, 1981:11).

For the most basic type of direct threat, or 'pure threat' (Greenawalt, 1989:89), as illustrated in (1), the speaker presents information and the listener has no control over the outcome, with the speaker remaining in the position of power over the threatened action.

(1) *"I'll break your legs for sleeping with my girlfriend"*

Contrast (1) with a conditional version of the same direct threat, as shown in (2), and the illocutionary point of the utterance becomes ambiguous.

(2) *"If you don't pay me the money I'm owed, I'll break your legs"*

The conditionality of the threat in (2) suggests that the purpose of the utterance is to get the hearer to pay the owed money to the speaker. Direct conditional threats are identified by Fraser (1998:168) as the most common type of direct threat, who further argues that in such threats, the addressee has control over the outcome. However, as Gales (2010:11) highlights, just because a speaker factors a condition into the design of their threat, there is no obligation on the speaker's part to uphold the stated condition. This is because the speaker remains in a position of power over the hearer throughout. While the wording of the direct conditional threat in (2) suggests that the speaker will not break the hearer's legs if money is paid, there is no guarantee that the speaker would not carry out the threatened action regardless of whether the condition was met or not. However, for this type of conditional threat to be successful, the target must believe that he or she can gain control over the situation by complying with the threatener's demands, regardless of whether the speaker intends to uphold the condition or not (Milburn and Watman, 1981:10). The key factor for conditional threats is, therefore, whether the listener *believes* they have control over the outcome, rather than whether they actually have any control or not. This is linked to the notion of credibility, with Milburn and Watman (1981:17) arguing that a threat is only credible if the target believes that the threatener intends to carry out the stated demands or the attached consequence should the listener not conform to the conditional element of the threatening utterance.

When a speaker threatens, they express their intention to carry out a given act (Searle, 1979:4). The statement and expression of intention in a threat is sufficient to make a threat a performative speech act which expresses some psychological state, even if the speaker has no intention to carry out the threatened action. One recent example of this is the actions of EgyptAir flight MS181 hijacker Seif Eldin Mustafa. Mustafa caused the hijacked plane to be diverted from its intended flight path in March 2016 after threatening passengers and cabin crew that he would blow up the aircraft using a belt containing explosives (BBC News, 2016). After capture, it was revealed that the belt contained no explosives, but, out of necessity, the threat was considered real by security staff and those on board the plane. It is also noted here that conveying intent-to-harm is not equivalent to having a specific motive for carrying out a threatened action. This difference is outlined by Culpeper, Igansky and Sweiry (2017:5), who propose that "intention involves a plan to direct actions towards particular ends, whereas motivation involves reasons why one might have the intention".

While aspects such as stated conditions and credibility may be clear in direct and direct conditional threats, indirect threats are problematic because any unfavourable act or intention to intimidate must be inferred by the hearer (Fraser, 1998:168). In such cases, the hearer is forced to use other available information to decide upon the meaning of a given utterance. Searle (1979:30) argues that when a speaker produces an utterance and means what they say, the speaker's intention is to produce an illocutionary effect in the hearer which involves recognition of the speaker's intention. Searle (1979:30) subsequently highlights this as a problem for indirect speech acts as they inherently involve the speaker communicating more information than is contained in the words alone. Therefore, the intention behind an indirect threat is left for the hearer to infer, with Searle (1979:31) further stating that the main concern with indirect speech acts is

how speakers say one thing and mean another. The potential for either misinterpretation or misunderstanding is also heightened in indirect threats owing to a lack of expressed clarity over speaker attitude and speaker position. This is further discussed by Kaplan (2016:275), who argued that indirect illocutionary speech acts, such as threats, require inference on the part of the hearer using Gricean reasoning and it is, therefore, very difficult to categorically prove that a threat was indeed a threat and not another type of speech act with a more non-threatening, neutral or alternative meaning.

Consider, for example, the indirect threat in (4): "I know where you live". Based on wording alone, there is no expression of intention to perform an unfavourable act, yet the utterance could plausibly be interpreted as a threat. This point is highlighted by Watt, Kelly and Llamas (2013:106), who designed an experiment building "I know where you live" into two contrasting scenarios. The first involved the speaker inviting the hearer to a picnic with friends, and stating "I know where you live" to express that they could pick the hearer up. The second scenario used "I know where you live" as a threat made by a speaker against a hearer who owed them a large amount of money. However, the expression of intention on the part of the speaker of this utterance is veiled and therefore relies on recipient inference.

Sentences of every type of syntactic form can count as indirect threats (Fraser, 1998:169), and indirect threats can also be masked as other type of speech acts including statements, as in (4), questions, as in (5), promises, as in (6) and warnings, as in (7).

(4) - *"I know where you live"*

(5) - *"Do you want to get hurt?"*

(6) - *"I promise you'll get what's coming to you"*

(7) - *"I'm warning you, I'll never forget this"*

A further complication with the type of utterance in (7) is that it is also possible for a warning to be designed as another speech act without the use of 'warn' as a performative verb. An example of this is shown in (8).

(8) – Are you sure you want to do that?

In (8), the utterance is designed as a question which, if interpreted literally, would evoke a yes/no response from the hearer. However, it could equally be used to warn if the goal is not to question the hearer but rather to get them to reflect on whether to do something potentially unfavourable. Equally, it could also be used as a threat if the unfavourable action was to be performed by the speaker and they were acting to the hearer's detriment. This is summarised by Fraser (1998:165), who states that warnings become threats when the speaker is the agent of the unfavourable action. Crucially, however, for the indirect examples in (7) and (8) this would be left for the hearer to infer rather than being stated explicitly by the speaker.

Two further types of indirect threat are discussed by Kaplan (2016) in relation to the Elonis v. United States trial, which centred around potentially threatening Facebook messages posted by Anthony Elonis towards his ex-wife, who had taken out a protection of abuse order against Elonis following their split. A discussion of the trial and linguistic analysis of Elonis' Facebook posts is presented in Kaplan (2016). Examples of two of Elonis' posts are presented in (9), replicated from Kaplan (2016:276; 278).

(9) – *"Fold up your PFA [Protection from abuse order] and put it in your pocket. Is it thick enough to stop a bullet?... And if worse comes to worse, I've got enough explosives to take care of the state police and the sheriff's department."*

(10) – *"Did you know it's illegal for me to say I want to kill my wife?*

It's illegal.

It's indirect criminal contempt.

It's one of the only sentences I'm not allowed to say.

Now it was okay for me to say it right then because I was just telling you that it's illegal for me to say I want to kill my wife.

I'm not actually saying it.

I'm just letting you know that it's illegal for me to say that."

In this case, Eloinis' defence argued that neither of the posts in (9) or (10) conveyed any intention to cause harm. They argued that the post listed in (9) was "verbal art akin to rap lyrics" (Kaplan, 2016:276), and that the post in (10) did not express direct intention to harm but rather mentioned the idea of Eloinis killing his wife. Kaplan (2016:281) highlights parallels between this and a comedy sketch by Trevor Moore, in which Moore mentions the illegality of threatening to kill the President of the United States without ever directly expressing any intent to cause harm. Of course, the surrounding context of the Elonis and Moore utterances differs substantially. Crucially, and most significantly Elonis' utterances produced an elocutionary effect of fear on the part of his ex-wife. This point is highlighted by Kaplan (2016:282), who argues that in the context of an ex-husband who is angry at his former wife, simply mentioning the idea of killing the ex-wife is unlikely to sufficiently mitigate threat interpretation of that utterance.

However, based on wording alone, the devices used in the Elonis posts show how an indirect interpretation can arise from utterances which more closely resemble direct threats than those listed in (4), (5), (6) and (7).

Searle (1979:7) outlines differences between performative illocutionary speech acts which have a corresponding illocutionary verb and those which do not. Searle (1979:7) points out that speakers do not threaten by stating "I threaten X". This is in contrast to warnings and promises, where it is perfectly plausible to declare "I warn/promise X". Fraser (1998:168) also emphasises that threats are not constructed performatively, except in extremely rare cases where a speaker may say "I threaten you with X". Fraser (1998:168) argues that this means the intention in a threat can never be fully guaranteed, as 'threaten' is hardly ever used performatively.

The issue of delimiting indirect threats from other types of speech acts such as warnings or orders is described by Yamanaka (1995:38) as "the linguistic notion of threats". Gingiss (1986:153) argues that the assumption that both a speaker and a hearer will "know a threat when they hear one" is insufficient for courtroom purposes, despite its status as "the majority view". The position forwarded by Danet et al. (1980) that if a reasonable person would interpret an utterance as a threat then a threat has been made, is rejected by Gingiss (1986) as it does not attempt to define a threat, nor does it highlight the grounds upon which a so-called 'reasonable person' would interpret an utterance as being threatening. Futhermore, Kaplan (2016:283) highlights that while a reasonable inference of the posts in (9) and (10) would lead to the conclusion that Anthony Elonis was threatening his wife, we cannot state this with absolute clarity as "humans are not mind readers".

Issues such as these are particularly problematic with respect to indirect threats, which require a greater amount of interpretation on the part of the hearer than is required for direct threats. Gingiss (1986:155) argues that "the problem of indirect threats is one that the courts must deal with". The value in providing more research on indirect threats is further identified by Gales (2010:97) who reports that 62% of threats in her Communicated Threat Assessment Reference Corpus (CTARC) of 470 threatening letters (totalling 152,000 words) examined by the FBI were indirect, in comparison to 26% conditional and just 12% direct.

In an attempt to better define indirect threats, Gingiss (1986) applies Labov and Fanshel's (1977) framework for the classification of indirect requests to Fraser's (1975) framework for threat classification. The resulting conditions for defining indirect threats are:

> *If A makes an assertion to be about:*
>
> *a) the existential status of an action p*
>
> *b) the time, T, of a future action p*
>
> *c) other preconditions for a valid threat as given in the rule of threats (see Fraser, 1975)*
>
> *and all other preconditions are in effect, then A is heard as making a valid threat.*
>
> Taken from Gingiss (1986)

Gingiss (1986:156) argues that a framework such as this allows for utterances like "this gun is loaded" to be classified as a threat, given that the utterance indirectly asserts the speaker's ability to shoot the gun. However, such a statement could be interpreted

differently depending on the shared understanding between speaker and hearer over the meaning of the utterance and the situation in which it was uttered. In a critique of Gingiss' (1986) formation of rules for the classification of indirect threats, Al-Shorafat (1988) argues that a logical flaw exists in applying a formula designed for requests to threats as they are two different speech acts with fundamentally different functions. Al-Shorafat (1988) also explicitly argues for the inclusion of prosodic factors into a working set of conditions for defining threats, although offers no further analysis about how this should take place or which prosodic variables should be incorporated. Further analysis of indirect threats in relation to Gingiss' (1986) criteria is provided by Yamanaka (1995), who argues that aspects such as the reference to the time of an action (point 'b' in Gingiss' (1986) classification) alone would rarely constitute a threat. Yamanaka (1995:52) states that any criteria for defining indirect threats should be grounded in a set of criteria for defining direct threats, and proposes the following:

If A makes an assertion to B (not necessarily explicitly or in a declarative sentence) about.

> ***a.*** *A's ability to carry out an action X*
>
> ***b.*** *A's intention to carry out an action X*
>
> ***c.*** *the consequences of performing an action X or of a previously performed similar action Y*
>
> ***d.*** *the occurrence of an action X in the near future*
>
> ***e.*** *A's suspending of an action X in return for the satisfaction of A's demands of B*
>
> *and all other preconditions for a threat are in effect, then A is heard as making a valid threat.*

Taken from Yamanaka (1995:52)

While this definition improves on Gingiss' (1986) criteria on account of a greater level of comprehensibility and a defined link to direct threats, it is difficult to envisage how one would set about demonstrating to a court that conditions were upheld by a threatener unless they were stated explicitly. As Searle (1979) argues, the problem with any type of indirect speech act is that it requires the hearer to infer the speaker's meaning or intention. Given Greenawalt's (1989) assertion that threats are situation-altering speech acts, one way to approach indirect threat analysis is to examine what an indirect threat, or a threat of any kind, does to a hearer. The following section further explores the respective roles of speaker and hearers in the communication of spoken threats.

2.2.3. The role of the speaker and hearer

Differences in the way an utterance relates to what is, or is not, in the interests of the speaker and hearer can help to further distinguish between different speech act types such as threats, warnings and promises (Searle, 1979:6). Shuy (1993:98) argues that threats are made for the speaker's benefit and to the hearer's detriment, and have an outcome which is controlled by the speaker. This contrasts with warnings and promises, which are made for the hearer's benefit, with warnings also having an outcome which can be controlled by the hearer (Shuy, 1993:98). Fraser (1998:164) further distinguishes the between threats and warnings by arguing that threats are unfavourable acts designed to impose fear, whereas, when making a warning, the speaker typically acts in the addressee's best interests by informing them before a harmful effect takes place. In direct verbal threats, information about how the utterance is made to the hearer's detriment and for the speaker's benefit is encoded in the words, whereas in indirect verbal threats this information is coded in the hearer's inference.

Working to the principle that threats are primarily concerned with bringing about a negative effect in a hearer, the direct threat in (1) - *"I'll break your legs for sleeping with my girlfriend"* - could be re-analysed under the assumption that the utterance is only a threat if it causes a negative psychological effect in the recipient. This is not to say that such an utterance would not benefit the speaker, but the primary purpose is to cause the hearer to believe that harm will befall them as a result of the threat. It can be further argued that the benefit to the speaker in this case would only come as a result of the threat recipient being negatively affected by the utterance. Furthermore, threatening to break someone's leg is no guarantee that the action will take place or that the speaker will attempt to follow their words with physical action. Gingiss (1986:156) highlights that it is not necessary for the speaker to believe they are capable of carrying out a threatened action so long as their actions convince the target that the threat is real, citing an example of someone who threatens using a water pistol, knowing that they are not carrying a real gun, but convincing a target that the gun is real and could cause serious harm.

Placing greater emphasis on bringing about a desired, unfavourable effect in the hearer also allows for so-called 'empty threats' to be factored into an analytical framework for threats. One such example highlighted by Watt et al. (2016) is of a speaker stating *"I'm going to kill you"* to a friend who they have just beaten in a game of Scrabble. Here, the words form a direct threat, but it is unlikely that the utterance would bring about an unfavourable negative effect in the listener owing to the context in which it was made. A threat is therefore uttered but it is empty because it does not bring about an unfavourable effect in the hearer due to the surrounding context. Additionally, Fraser (1998) highlights that the unfavourable effect on the listener separates threats from promises, as promises are designed to be favourable to the addressee, and therefore

"sanctions may be imposed for a broken promise, but no such sanctions for a broken threat" (Fraser, 1998:164).

Fraser (1998:162) proposes the following conditions for a threat to be made, and states that these criteria serve to form a "context independent definition of a threat" (Fraser, 1998:162):

1. *The intention to perform an act;*
2. *The belief that the state of the world resulting from that act is unfavourable to the addressee;*
3. *The intention to intimidate the addressee.*

Taken from Fraser, (1998:162)

This position is at odds with the work of Storey (1995) and Gales (2010), who argue that threats are bound by a relationship of shared understanding between speaker and hearer. Storey (1995:75) argues that threats, by definition, are a two-way process and must be either accepted or acknowledged by a hearer to carry meaning. Gales (2010) also accepts this definition, arguing that "threats are socially constructed acts of power between two parties – the threatener and the threatened" (Gales, 2010:2). Milburn and Watman (1981:7) also advocate a model for threats which places the listener in a key position, arguing that listeners modify the meaning of a given threat depending on both situational and individual factors, and therefore play an important role in both the meaning and interpretation of a threat. This is particularly relevant for indirect threats, with Searle (1979:30) pointing out that shared understanding of the intention behind the utterance between speaker and hearer is essential for indirect speech acts to be

communicated successfully. This proposition is further supported by Watt, Kelly and Llamas (2013:100), who argue that for an indirect threat to achieve its desired effect, there must be shared understanding between speaker and hearer of both the content and context. This links to Storey's (1995) assertion that shared understanding between speaker and hearer is crucial for the successful communication of a threat, and further validates the idea that producing a desired, usually unfavourable, effect in the hearer is a key criterion in threat communication. This is the view that this research is framed upon.

2.2.4. Making and communicating threats

Based on previous definitions and descriptions of threats, it can be argued that a lack of clarity exists as to the relative roles of the speaker and hearer. For example, while Fraser (1998:162) argues that it is necessary for the speaker to intend to intimidate the addressee in a threat but not for the addressee to feel intimidated, Storey (1995:75) argues that the "degree of criminality of a threat depends upon the effect that threat has on the victim". One distinction that can be proposed to better define the role of the speaker and the hearer in verbal threats is the difference between *making* a threat and *communicating* a threat. Milburn and Watman (1981:8) state that a threat is "the communication of one's intention to take an action harmful to another party". The emphasis is placed on not only making a threat, but communicating the intention behind an utterance to a hearer or target. Fraser (1998:163) argues that aspects like ambiguity, a threat not being heard by the recipient or the recipient not understanding the words used within a threat can all serve as example of a threat being made, but not communicated. Here, it is argued that the expansion of this concept to include the acceptance or acknowledgement of the unfavourable effect of a threat on the recipient would better define the importance of the listener's role within the interaction. It is further argued that

a speaker can make a threat by fulfilling all the necessary criteria proposed by Fraser (1998), but for a threat to be communicated, the intended psychological effect on the recipient must be accepted or acknowledged by either that recipient or the hearer of the threat.

Consider, for example, a situation in which the status of an indirect threat is disputed in court, such as the dispute over whether Don Tyner threatened Vernon Hyde with the utterance "[H]ow's David? [Hyde's son]", outlined by Shuy (1993:108). Acting for the prosecution, a Federal Bureau of Investigation (FBI) analyst in this case stated that he considered the utterance to be a "serious and real threat" (Shuy 1993:109). However, contrasting linguistic analysis provided by an expert witness acting for the defence stated that the structure of the interaction and aspects of the surrounding context meant that the utterance should have been interpreted in its literal sense. Milburn and Watman (1981:11) state that context can record to both the situation in which a threat is made and the wider social and personal norms which underpin both the threatener's and target's behaviour. They further argue that context can serve as a mediating variable between speaker intention and how a target responds to the threat (Milburn and Watman, 1981:11). Milburn and Watman (1981:11) add that in order to examine the linguistic nature of threats, characteristics of both the threatener and the target need to be considered, as threats are bound to the status and position of both the threatener and the threatened along with the legitimacy of the sanction being threatened.

The courtroom dispute in the "How's David" case centred around the shared understanding between speaker and hearer over the communication of a potential indirect threat. A figure originally created by Shuy (1993:17), and reproduced in Figure 2.1 can used to demonstrate the shared knowledge that the speaker and hearer bring to

the understanding, and potential misunderstanding, of a given utterance. The original figure is replicated below.

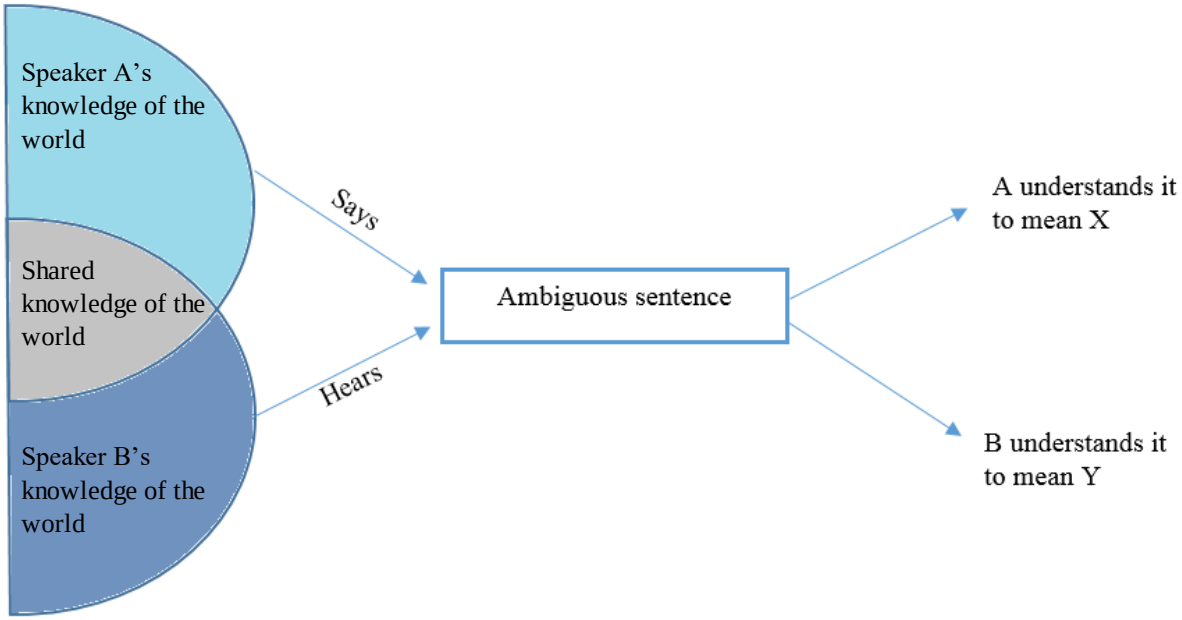


Figure 2.1 – Replication of figure designed by Shuy (1993:17) to illustrate how speaker and hearer can arrive at different interpretations of an ambiguous sentence

Figure 2.1 illustrates that when an utterance is produced by a speaker, the intended meaning is not always necessarily shared between speaker and hearer. Shuy (1993:17) shows how a speaker (A) can produce an utterance which results in the hearer (B) perceiving a contrasting meaning to the intended meaning. The notions of making and communicating a threat, along with the roles of the speaker and hearer in threatening communications, can also be considered using a comparable approach. Figure 2.2 is a replication of the figure produced by Shuy (1993) to show how differences in the interpretation of a threatening utterance can arise on the part of speakers and hearers.

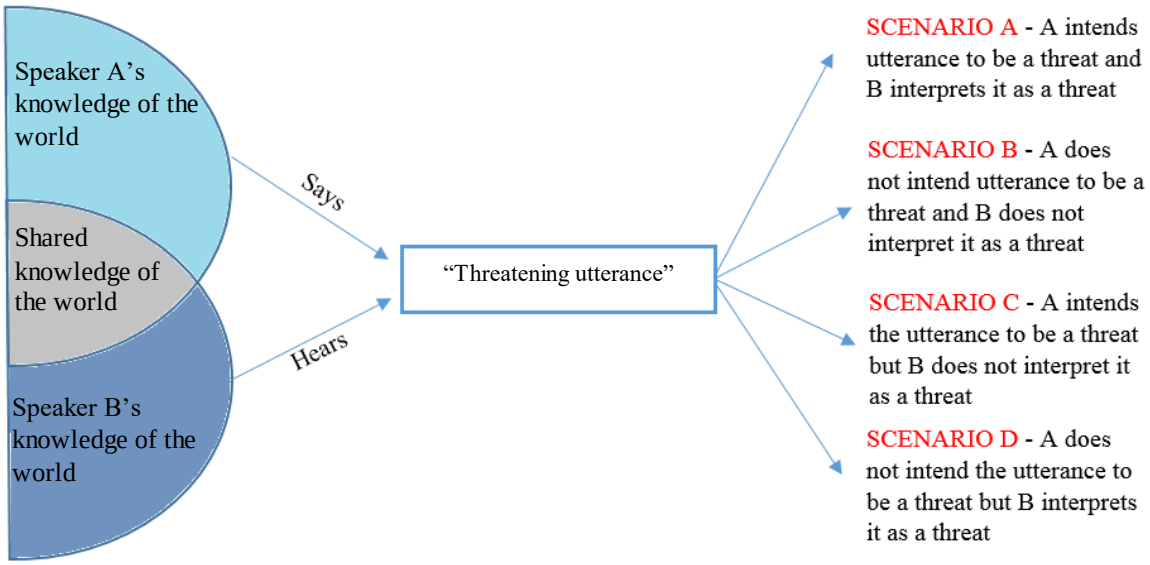


Figure 2.2 – Illustration of the different potential outcomes arising when a threatener (Speaker A) produces a threat directed towards a hearer (Speaker B)

The contrasting scenarios in Figure 2.2 illustrate the potential outcomes of a threatening utterance, depending upon how the meaning is interpreted by the speaker and the hearer. The notions of making and communicating a threat can also be considered under this framework. In Scenario A, both speaker and hearer share acceptance of the utterance as a threat. Given that the speaker intends the utterance to be a threat and the hearer interprets it as a threat, it can be said a threat has been successfully made and also successfully communicated. In Scenario B, the speaker does not intend the utterance as a threat, so a threat has not been made, and the hearer does not interpret it as a threat so a threat has not been communicated. In both Scenario A and Scenario B, the making and communication of the threat are consistent with one another. However, it is possible for mismatches to occur, as shown in Scenario C and Scenario D. In Scenario C, while the speaker intends the utterance to be a threat, the hearer does not interpret it as a threat. Under this scenario, a threat has been made but not communicated as it is not accepted

as a threat by the hearer. The reverse applies to Scenario D, where the speaker does not intend the utterance to be a threat but the hearer interprets it as a threat. Under this condition, it can be argued that a threat has been communicated, but not made.

Re-examining the "How's David" example discussed by Shuy (1993), I argue that the dispute over whether the utterance was a threat was a contrast in interpretations between Scenario A and Scenario D in Figure 2.2. In this case, the prosecution argued for Scenario A, under which the speaker meant the utterance as a threat, and the defence argued for Scenario D, under which the speaker did not make a threat even though the hearer interpreted the utterance as a threat. Moreover, commenting on this particular case, Fraser (1998:169) states that the court heard "[H]ow's David" as a serious threat. This highlights that in criminal trials where threats are disputed, the role of the speaker and hearer is reduced, with the perception and judgements of courtroom triers of fact playing a pivotal role in the decision making process. In such cases, a third-party listener bringing new perspectives and differing knowledge of the world to both the speaker and the hearer assumes primary responsibility for assessing and evaluating the legality of an alleged threatening utterance.

2.3. Threats and the law

The dual nature of threats as standalone crimes and as an integral part of other crimes is captured by the definition of 'threat' provided by the Oxford Dictionary of Law (Law and Martin, 2009), which states that a threat is "the expression of an intention to harm someone with the object of forcing them to do something" and that threats are "an ingredient of many crimes". The Oxford Dictionary of Law provides a more detailed definition for 'threatening behaviour', which is listed as the use of "threatening, abusive

or insulting words or behaviour" towards another which is punishable by either a fine or up to 6 months imprisonment. For threats embedded within other crimes, the punishment could be substantially larger.

Watt, Kelly and Llamas (2013:100) highlight that it can be difficult to prove that a threat was meant as a threat, even if a hearer interpreted the utterance as one. In order for a successful prosecution in court for the offence of 'threatening behaviour', it must be proved that "the accused person had the specific intent to cause the other person to believe that immediate unlawful violence would be used against him or, simply, that the threatened person was likely to believe that violence would be used against him" (Law and Martin, 2009). This legal position is further clarified by Watt, Kelly and Llamas (2013:101) who state that threats are defined by the fact that they cause the target to believe that the threatener carries an intention to harm, not necessarily whether the threatener either has the ability or the intention to do so.

A further aspect of interest in the definition of 'threatening behaviour' provided by the Oxford Dictionary of Law is that a police officer can lawfully arrest anyone who is reasonably suspected of uttering verbal threats. Again, this further emphasises the potential role of a third-party listener in the legal process surrounding threats, as police officers can legally make decisions about whether they consider a given utterance to be threatening, and make arrests based on such decisions.

In the UK, illegal verbal threats are covered under the 1986 Public Order Act, which states:

(1) A person is guilty of an offence if he—

(a) uses towards another person threatening, abusive or insulting words or behaviour,

or

(b) distributes or displays to another person any writing, sign or other visible representation which is threatening, abusive or insulting, with intent to cause that person to believe that immediate unlawful violence will be used against him or another by any person, or to provoke the immediate use of unlawful violence by that person or another, or whereby that person is likely to believe that such violence will be used or it is likely that such violence will be provoked.

(Public Order Act 1986, Ch. 64, Section 4.1).

Watt, Kelly and Llamas (2013:102) argue that detailed analysis of the intricacies of legal interpretations of the Public Order Act is best left to those with specific expertise in the application and interpretation of legal language. However, a more general analysis of the wording used in Section (b) further highlights that there must be the intention on the part of the speaker to cause the threatened party to believe that a negative consequence will befall them as a result of the threat. Solan and Tiersma (2015:223) state that "threats provide a basis for criminal liability if they instil fear or violence as a retribution for failing to comply with a demand", again emphasising a listener-oriented approach to the analysis of threats rather than one which focussed entirely on the threatener.

Although not taken from the UK, the aforementioned *Elonis v. United States (2015)* case resulted in several developments for the legal treatment of threats in the USA. During the trial, the Supreme Court justices argued that the threatener's mental state should be factored into judgements about threats. The case highlighted three important states to consider: intent, recklessness and negligence. Intent concerns intention to intimidate via language use; recklessness refers to whether a defendant is aware that their words would be interpreted as a threat and was indifferent to the idea; and negligence concerns whether a defendant should have known that there was a risk that their utterance would be interpreted as a threat but did not. While this reduces the need for a 'reasonable person' interpretation, Kaplan (2016) highlights that to successfully ensure that a defendant was found to be reckless, prosecution lawyers must prove that defendants were aware that their utterances would be interpreted as threats but decided to utter then anyway. Kaplan (2016:288) highlights that this judgement poses issues for the linguistic analysis of threats, which has previously been seen as a binary concept which requires little understanding of human motivations. Arguing against the idea that an utterance is either a threat or not a threat, Kaplan (2016:289) further asserts that in the case of verbal threats linguistic analysis cannot always reveal everything about linguistic phenomena. How linguists factor human motivations into working taxonomies for threatening communications remains a challenge which should be addressed through further research and analysis of cases involving verbal threats.

2.4. Linguistic analysis of threats

In addition to work examining the threats as speech acts and the classification of different threat types (Al-Shorafat, 1989; Fraser, 1998; Gingiss, 1986; Storey, 1995;

Yamanaka, 1995), research also exists on specific linguistic properties of threats. This work has primarily focussed on threats as a form of written communication. For example, Carter (2010) presented a corpus analysis of sentence type and pronoun usage in threats delivered by terrorist and non-terrorist groups. Carter's (2010) research found that declarative sentences and first person nominative pronouns are frequently used by both groups, with terrorist threats favouring use of the second person nominative 'we' and non-terrorist threats favouring the first person nominative 'I'.

A body of work on the role of stance markers in written threats has been conducted by Gales (2010; 2011; 2012; 2015; 2016). Gales (2015:171) highlights that while there is no one-to-one mapping between linguistic markers and actions taken by threateners, linguistic analysis of features such as stance markers can contribute greater understanding of threats and help to substantiate victims' claims of feeling afraid. Stance is defined as the "personal feelings, attitudes, value judgements or assessments" that speakers express through their utterances (Biber et al., 1999:966). An investigation of grammatical stance markers in the CTARC corpus is presented by Gales (2010). Gales (2010) reports, among other features, that the presence of the non-contracted modal verb 'will' was identified by both threat assessment professionals and scholars as a marker of an increased level of commitment to the threatened action, whereas the use of possibility modals such as 'may' weakened commitment and the speaker's overall stance. Nini (2017) further argues that prediction models such as 'will' emphasise certainty on the part of the threatener, and Napier and Mardigian (2003:18) identified 'will' as a linguistic feature in high-level threats such as "I will shoot him between the eyes" and "If I can't find him at the casino, I will find him at his residence on Townsend Avenue". However, an examination of stance markers in realised and non-realised threats in the CTARC corpus is presented by Gales (2016), and shows trends between

the two groups including, but not delimited to, prediction modals such as 'will' and 'shall' occurring more frequently in non-realised threats than realised threats, and certainty adverbs occurring more frequently in realised threats than non-realised threats.

Gales' (2016) work shows how perceptions of features of 'threatening language' can often be at odds with the realities of the way that threats are actually uttered. For example, in the CTARC, threateners who acted on their threats were often found to use mitigating language to either displace responsibility or allow for negotiation to take place (Gales, 2016:19). However, in an earlier community of practice survey among threat assessment researchers and practitioners, Gales (2010) found that mitigating language was often identified as a property of non-realised threats. This further illustrates the gap that can exist between the actual meaning behind a threat and the way it is interpreted, even by professionals and other experts. It also reinforces the potential differences between production and perceptual aspects of spoken threat utterances.

Gales (2016:21) further states that linguistic research on threats to date is yet to address the difference between spoken and written threats, including research on stance markers. Biber et al (1999) argue that in addition to grammatical and lexical markers, speakers can display what they term a "linguistically covert stance" (p.967) through aspects of voice such as pitch, loudness and utterance duration. While Gales (2010:58) highlights the potential for the inference of prosodic cues in written threats through aspects such as capitalisation, emboldening and the use of emojis in computer mediated communication, prosodic factors remain a primary property of spoken language as opposed to written communication.

However, as paralinguistic stance is not marked by grammatical or lexical aspects of speech, listeners must instead infer the attitudes being expressed by the speaker (Biber et al., 1999: 967). This shifts the analytic emphasis from speakers' productions onto listeners' perceptions of speakers' language use. The relationship between paralinguistic stance marking and the inference of threat in spoken language is currently an underexplored area. Biber et al (1999) do not present a detailed analysis of this aspect of stance marking, other than to state examples where fictional writers use dialogue tags such as "'Do you?' Helen spoke angrily" (Biber et al., 1999:967), where the attached feeling is placed alongside speech marks to denote an attitude which would not be automatically marked by the words alone. Biber et al (1999:968) further add that readers have "no difficulty in imagining the tone of voice and body gestures that could accompany these attitudes", although offer no empirical analysis of the phonetic markers associated with different paralinguistic stances.

In comparison to the body of research that exists on written threats, fewer studies have investigated how aspects of speakers' voices could affect listeners' perceptions of spoken threats. Watt, Kelly & Llamas (2013:100) state that a speaker's 'tone of voice' – however this might be defined phonetically – is the term used by the police, the courts and the general public to capture properties of the speech signal that listeners may use to infer threat. Watt, Kelly and Llamas (2013) also note that empirical research on how specific phonetic variables may contribute to listener perceptions of a so-called 'threatening tone of voice' was at that time still lacking. Milburn and Watman (1981:55) further argue that "if a threat is uttered in a warm and friendly tone of voice, what might otherwise have seemed hostile or fearsome may be perceived as being humorous and acceptable". Investigating this further, Watt, Kelly and Llamas (2013) found that listeners inferred greater levels of threat from productions of the indirect

threat "*I know where you live*" when it had been designed by the speaker to sound threatening, compared with productions of the same sentence that had been designed by the speaker to convey no threat or intent-to-harm. This study challenges the idea that only the words used in a spoken threat can influence either its meaning or interpretation, particularly when the utterance in question is indirect, vague, or could be interpreted as another speech act such as a warning or a promise.

Watt, Kelly and Llamas' (2013) study helps to illustrate that listeners can use multiple channels when inferring threat from spoken utterances. It can, therefore, be argued that their work begins to challenge the notion that threats should be treated as a purely verbal content-driven phenomenon. However, they acknowledge that their work does not begin to analyse how specific phonetic cues may cause listeners to infer greater or lesser levels of threat in a talker's utterance (Watt, Kelly and Llamas, 2013:100).

In a follow-up experiment to Watt, Kelly and Llamas (2013), Kelly (2014) investigated the plausibility of finding common phonetic cues adopted by speakers when making verbal threats. This research found that, of the features and spoken threats investigated, there were no consistently significant phonetic differences between utterances designed to be interpreted neutrally and those designed as threats. Although changes were made by speakers when wishing to sound threatening, the manner of achieving such a 'threatening tone of voice' was not consistent across the sample of speakers. Kelly's (2014) findings suggest caution in assuming cross-speaker commonalities when considering the phonetic basis of a 'threatening tone of voice'. Considering the wide array of possible reasons why somebody might choose to make a verbal threat, Kelly's (2014) results are not surprising, and the conclusion calling for "a less simplistic consideration of threatening language" (Kelly, 2014:29) is a valuable assertion. Kelly's

(2014) study also showed limited phonetic correlation between induced threat utterances and previously documented reports on phonetic cues to anger. However, results in Tompkinson (2016) showed strong correlations between perceptual listener ratings for how angry, aggressive and threatening speakers sounded when producing a range of simulated direct and indirect threat utterances. This highlights the lack of one-to-one correspondence between speakers' production and listeners' perceptions with regard to phonetic aspects of spoken threats in a comparable way to the results presented by Gales (2010; 2016) when examining written threats.

According to the framework outline by Agha (2005; 2007), the term 'threatening tone of voice' can be classified as a metalinguistic label (Agha, 2007). Metalinguistic labels link speech registers with specific linguistic features or properties, with the use of specific language features also indexing personal or social characteristics (Agha, 2007:145). Agha (2007:145) states that the existence of metalinguistic labels acts as evidence for cultural speech models that link features of speech with "typifications of actor, relationship and conduct". The concept of metalinguistic labels can be further extended to incorporate metapragmatic stereotypes (Agha, 2007:148), which develop when speech types become culturally linked to personality traits. Agha (2005:38) terms this "enregisterment", which is the process by which "distinct forms of speech come to be socially recognised (or enregistered) as indexical of speaker attributes by a population of language users." Gales (2010) argues that assumed linguistic markers of threats, despite their potential inaccuracy, become enregistered in the minds of listeners, leading to stereotypical assumptions made by untrained lay-listeners about the nature of threatening language. Linked to this, Agha (2005:39) further argues that voices can become "characterisable" from their association with linguistic forms, and that these so-called "enregistered voices" index specific social personae and characteristics. These

personae are not necessarily confined to a single individual, and can reflect wider social groups as well as specific people (Agha, 2005:40). Given the lack of direct correspondence between the linguistic patterns adopted by threateners and their subsequent actions (Gales, 2010:262), relying on folk-linguistic assumptions of threateners' intent inferred by non-linguists through speakers' language is potentially dangerous. The assumption that language users simply "know a threat when they hear one" (Gingiss, 1986:153) is fundamentally insufficient for any purpose with legal implications or consequences.

The potential dangers of over-reliance on lay-listener threat perception are particularly well-illustrated when members of the public are required to analyse spoken threats made by unknown threateners. One situation in which this takes place is when linguistically untrained lay-listeners working in places such as schools, hospitals and businesses are required to evaluate bomb threats made via the telephone. The UK National Counter Terrorism Security Office (NCTSO) issues a bomb threat checklist document designed to elicit information about both the verbal content of the threat and the threatener's voice. Users are required to provide a description of what the speaker sounded like, along with information about the 'threat language' used. This section of the document is shown in Figure 2.3.

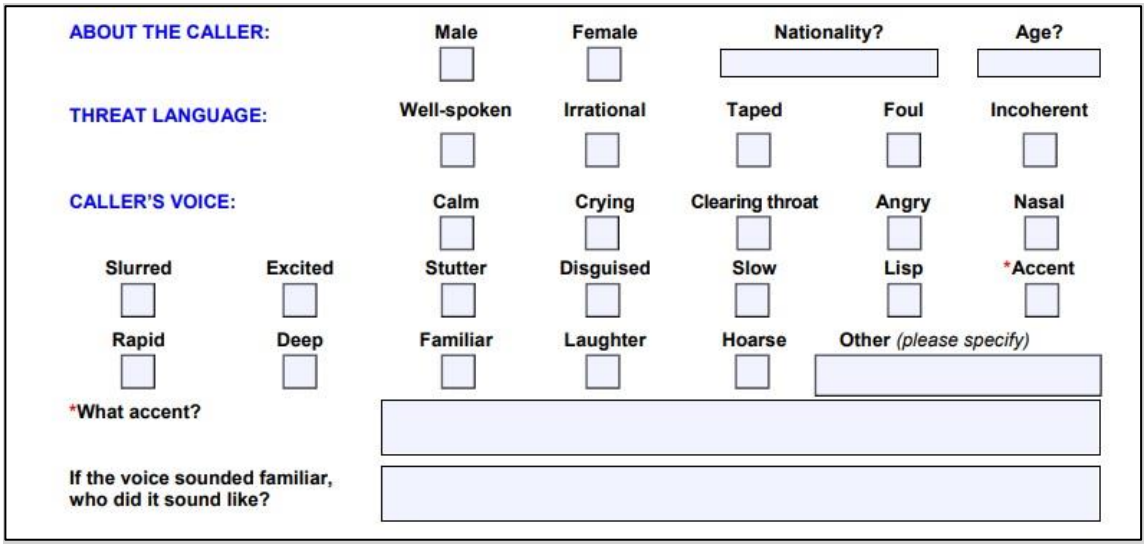


Figure 2.3 – Extract from UK National Counter Terrorism Security Office bomb threat checklist (NCTSO, 2016)

However, the great majority of earwitnesses to crimes will have had no formal linguistic training (Griffiths, 2012), and, according to Shuy (1993), will almost always lack both the ability and the vocabulary needed to give accurate descriptions of other speakers' language behaviour. Added to this, it is unlikely that the majority of earwitnesses will have voice description skills comparable to those who have received specialised training in phonetics or linguistics (Watt and Burns, 2012). These issues present an ongoing problem to police officers and security personnel, who from time to time will wish to elicit meaningful descriptions of the voices of criminals from earwitnesses.

Sherrin (2015) documents two examples of cases in Canada in which unreliable earwitness voice identification led to wrongful convictions, and also cites 17 US cases of wrongful imprisonment that were based, at least in part, on faulty earwitness testimony. Although speaker identification by earwitnesses and earwitnesses' descriptions of offenders' voices are not equivalent, dependent as they are upon different sorts of memory recall, they are closely related. Broeders and van Amelsvoort (2001) state that the foil (non-suspect) samples in a voice identity parade should match

as closely as possible the verbal description given by the witness, although they also point out that such descriptions are often fraught with complications and do not necessarily form a sound basis upon which foil selection should take place. Furthermore, the UK guidelines on constructing voice lineups (Nolan, 2003:288) explicitly state that the identification officer in charge should obtain a detailed statement from the witness which "should contain as much detail and description of the [offender's] voice as is possible". This emphasises the need for voice descriptions to be promoted as best practice in the UK as a part of eliciting earwitness evidence.

It has also been argued that the process by which linguistically-untrained listeners identify voices operates below the level of consciousness (Broeders and van Amelsvoort, 2001; Watt, 2010), making it difficult for an earwitness to introspect about and verbally externalise what can essentially be viewed as an automatic process. The problem is further compounded by the often highly technical nature of the terminology used by expert phoneticians to capture aspects of a speaker's voice, much of which – in spite of the relative transparency of labels like 'creaky', 'whispery' or 'breathy' for certain voice quality attributes – is unlikely to be contained in the non-linguist's lexicon. This was commented on by Yarmey (2001), who obtained voice descriptions of unfamiliar speakers using an open-ended question format in which listeners were free to provide as many or as few descriptors as they considered appropriate. Yarmey (2001) observed that listeners provided, on average, between 4 and 5 descriptors, but that these were often non-technical and somewhat limited in their usefulness.

However, despite warnings from researchers that linguistically-untrained listeners perform poorly when tasked with describing the voices of speakers, some research has shown that listeners appear to be able to identify some aspects of speakers' voices with

relative accuracy. In an investigation into listener accent attribution, Griffiths (2012) found that lay listeners were able to label speakers' regional accents relatively accurately, although descriptions of the voices of speakers with marked regional accents were more accurate than those for speakers with less marked regional accents. Additionally, Watt and Burns (2012) found that listeners were able to provide phonetically interpretable descriptions of voice quality with a relative degree of accuracy, and in a way that was compatible with expert terminology. Furthermore, the study reported by Dixon, Foulkes and LaShell (2013) showed positive correlations between listeners' perceptions of pitch and measured average fundamental frequency (F0). These studies also highlighted voice quality as a potentially influencing factor in listeners' judgements of how high-pitched a speaker's voice was. This finding was supported by further research conducted by Fisher (2018), who found significant strong correlations between F0 and listeners' judgements of pitch in short clips of both studio and telephone-quality recordings. The correlations were, however, stronger in the studio-quality speech in Fisher's (2018) work.

Both Griffiths (2012) and Watt and Burns (2012) stress the importance of further research on how non-linguists describe voices in forensically relevant contexts. Griffiths (2012:76) specifically warns that this research is needed because "non-linguist members of the general public are appointed to elicit the best possible linguistic evidence, from other non-linguist members of the general public, which other non-linguists then represent in law courts." This call is addressed by a more recent research projects outlined by Smith et al. (2018). Smith et al's. (2018) work explored the merits and weaknesses of free recall descriptions compared with structured responses based on a series of questions provided to listeners. The research found that while the descriptions

provided across the two forms of information elicitation, using structured questions created more comparable results across the different participants within the experiment.

2.5. Phonetic variation and the perception of affective speech

Given previous assertions (Gales, 2010; 2016) that linguistic features can become enregistered in the minds of listeners as markers of threatening language, and that a speaker's 'tone of voice' can influence threat perception, it seems plausible to suggest that certain phonetic aspects of speech could act as enregistered markers of threat. However, given the lack of a specific body of research on vocal cues and perception of threat, it is necessary to draw inferences from literature on perceptions of emotional and affective speech in order to hypothesise how phonetic variables may influence threat perception. This section examines previous research on the perception of speaker characteristics from aspects of voice.

Research spanning a period of over 80 years has illustrated that listeners willingly form impressions of unknown speakers based on their vocal characteristics. Pear (1931) and Allport and Cantril (1934) were among the earliest researchers to illustrate this phenomenon, using radio broadcasts to obtain listener evaluations of presenters' voices. Tusing and Dillard (2000:148) argue that given their primitive origins, vocal cues may have a more important role in social perception than either linguistic content or other non-verbal cues including, for example, facial characteristics or expressions. Dimos et al. (2015) highlight that F0 and speech rate are among the most perceptually salient acoustic cues used by listeners to infer emotion and affect in speech, although other potentially important cues could include voice quality (Xu et al., 2013) and intonation patterns (Scherer, 2003). Relevant research on the link between each of these phonetic

variables and the perception of both affective and emotional states in human speech is examined in this section.

2.5.1. Fundamental frequency

Fundamental frequency (F0) is the rate at which the vocal folds vibrate, and is measured in Hertz, corresponding to the number of vocal fold vibrations per second. For a typical male speaker speaking with modal voicing, the vocal folds will vibrate regularly around 120 times per second (Laver, 1994:193) with an average female voice having a mean F0 of approximately 210Hz. F0 is linked to pitch, which is a perceptual property that has F0 as its acoustic correlate (Laver, 1994:450). Although the link between pitch and F0 is strictly non-linear, Laver (1994:451) argues that at the low frequencies relevant for the perception of pitch in both male and female voices, a linear relationship can be assumed.

Of the phonetic parameters that have been investigated by researchers in relation to perceptions of attributes such as threat, dominance and aggression, along with various emotional states, F0 is the most common (Bachorowski, 1999; Ohala, 1984). Building on work by Morton (1977), who argued that lowered pitch marks aggression and dominance across a variety of animal species, Ohala (1984) showed that when listeners heard low-pass filtered human speech with spectral details removed, low-pitched recordings were rated as sounding more dominant than high-pitched recordings when all other aspects remained constant. Ohala (1984) further argues that the lowering of mean pitch to signal dominance is related to the idea that lower pitch signals a larger person; a phenomenon known as either the ‘frequency code’ or ‘size code’ hypothesis.

Gussenhoven (2004:79) links the correlation between larynx size and vocal fold vibration to the communication of power dynamics, and associates the frequency code with other seemingly universal aspects of pitch variation including, for example, high pitch with utterance beginnings and lowered pitch with utterance endings. However, while highlighting the universal nature of the frequency code hypothesis, Gussenhoven (2004:79) argues that when coupled with meaningful speech, this universal aspect of human vocalisation acquires a more defined "affective" role in order to signal particular speaker or utterance attributes. Equally, emphasis is also placed on social functions when analysing the affective nature of the frequency code. For example, while the biological nature of sex differences between men and women results in male speakers having lower pitch compared to female speakers, the degree to which this gap exists varies as a function of the social constraints placed on language use within particular groups or communities (Gussenhoven, 2004:80). Gussenhoven (2004:82) highlights a range of affective Frequency Code associations, including higher pitch with appearing as submissive, friendly, polite and vulnerable, and lower pitch with appearing as dominant, aggressive and scathing.

Based on the frequency code hypothesis, speakers can produce utterances with lowered F0 should they wish to appear more dominant, larger and physically imposing (Ohala, 1984:5). This would be particularly relevant in cases where an unfamiliar speaker was heard but not seen, as is often the case in interactions over the telephone; a common method for the delivery of threats (Eriksson, 2005). However, there is no requirement for the speaker in question to have a large build. Ohala's (1984) hypothesis instead centres around the ways in which a speaker may manipulate their pitch in order to appear more physically dominant. Furthermore, research by both Künzel (1989) and Gonzalez (2003) found no statistically significant relationship between speaker height,

weight and F0 in either running speech (Künzel, 1989) or single vowels (Gonzalez, 2003). This highlights the lack of one-to-one mapping between pitch and body size, and suggests that the relationship between pitch and dominance is more closely tied to listeners' perceptions of speaker size than it is in the biological relationship between body size and F0.

A range of perceptual studies have shown a link between perceptions of body size, personality judgements and the lowering or raising of F0. For example, Feinberg et al. (2005) found that female listeners rated male voices with lowered F0 as being more masculine, physically larger, older and more attractive than those with raised F0. Xu and Kelly (2010) found that lowered F0 projected a larger and angrier speaker, with raised F0 signalling a smaller, happier speaker. Xu et al. (2013) further examined the relationship between F0 and attractiveness, finding that male listeners showed preference for female voices with a breathy voice quality, higher F0 and more widely distributed formants, whereas female listeners found male voices with lower F0, breathy voice quality and denser formant distribution to be more attractive.

Tusing and Dillard (2000) further highlight the contrast between the association of low-pitched vocalisations with hostility and aggressiveness and the association of high-pitched vocalisations with non-aggression and submissiveness. Puts et al. (2006) found that a one-semitone increase or decrease in mean pitch caused listeners to perceive significant differences in both social and physical dominance for male speakers, with lowered pitch resulting in increased dominance ratings. This result was also replicated in Puts et al. (2007), which found effects for both mean pitch and formant dispersion independently of one another on listener perceptions of dominance. McAleer et al. (2014) also found that listeners rated lower pitched male voices as sounding more

dominant that higher pitched male voices, but that the opposite pattern held for female voices, with higher-pitched female voices rated as sounding more dominant than lower-pitched female voices. However, Borkowska and Pawlowski (2011) illustrated an effect for mean pitch on listener evaluations of dominance in female voices, with lower pitched voices rated as sounding more dominant than higher pitched voices. Tsantani et al. (2016) highlight that a perceptual link between dominance and lowered pitch is less well-established for female voices than for male voices. In a forced choice task where listeners were instructed to pick the most dominant-sounding voice from a pair of vocal stimuli, Tsantani et al. (2016) found that while both lower-pitched male and female voices were picked more frequently as the dominant-sounding voice by listeners, the preference was only significantly greater than chance for the male voices.

Both Chuenwattanapranithi et al. (2009:3) and Gussenhoven (2002) argue for the frequency code to be labelled as the size code in order to acknowledge that other aspects of the vocal channel aside from F0 could contribute to listener perceptions of body size, aggression and threat. Across five experiments examining F0 and vocal tract length in relation to perceptions of body size and emotions such as happiness and anger, Chuenwattanapranithi et al. (2009) found that a statically longer vocal tract and lower F0 projected a larger person, with a dynamically lengthened vocal tract and lower F0 signalling anger. These results led the authors to argue that the size code is involved in the perception of both emotions and body size (Chuenwattanapranithi et al., 2009:17). In further discussion about the relationship between phonetic cues, emotion, threat and body size perception, Chuenwattanapranithi et al. (2009:4) argue that "intuitively, if a vocal expression sounds angry, it also feels aggressive and threatening". It is further argued that the expression of emotions such as anger correlate with displays of aggressiveness and link to the size code. Linked to this, Vaissiere (2005:251) also

highlights the potential for other aspects of the vocal channel to influence affective perceptions, arguing that anger and hostility in speech are characterised by "F0 irregularities, forceful innervation of the glottal muscles, narrow constriction of the glottal space as well as retracted lips and tongue retraction".

Furthermore, Banzinger and Scherer (2005:257) highlight the subjective nature of pitch perception, and argue that expectation biases on the part of both speaker and hearer can influence perceptions of pitch. This further links to Gussenhoven's (2004:80) assertion that affective use of F0 is socially constrained. It also highlights the importance of considering the range of expectation biases that may arise as a result of social, emotional and other constraints on the part of both speaker and particularly hearer when examining any link between pitch or laryngeal aspects of voice and the perception of spoken threats.

2.5.2. Voice quality

Voice quality is defined by Laver (1994:153) as the general phonetic settings used by an individual. A setting is defined as "any tendency for the vocal apparatus to maintain a given configuration or featural state over two or more segments in close proximity" (Laver, 1994:153). Laver (1994:184) defines phonation quality as the use of the laryngeal system to create audible acoustic energy, which can be subsequently modified by the higher part of the vocal tract. Laver (1994:153) argues that phonetic settings can form the basis of a particular tone of voice.

The human larynx is capable of producing a wide array of contrasting phonation qualities depending upon how the vocal folds are held in place in the larynx (Laver, 1994:186). During modal voicing, the vocal folds are brought together and are set in

regular, rhythmic vibration by pulmonic airflow (Catford, 2001:37). Modal voicing can be considered as a neutral phonation setting (Laver, 1980:110). In contrast, creaky voice is characterised by "a very low frequency 'creaking' or 'crackling' sound" (Catford, 2001:51) produced with a predominantly closed glottis that has a small portion at the front of the vocal folds open and vibrating slowly. The frequency range for creak can be as low as 25-50Hz (Laver, 1994:195). Falsetto voice involves sound production at frequencies extending beyond a speaker's modal range (Laver, 1994:197), with a male falsetto range reported between 275-634Hz (Hollien and Michel, 1968:602). Whispered phonation is described as having a hissing quality caused by turbulent airflow through the glottis (Laver, 1994:190), while harsh voice is identified as involving a "severely constricted" larynx caused by extreme laryngeal hyperextension (Laver, 1994:420).

According to Laver (1994:197), different phonation qualities can create both phonological and paralinguistic meaning, depending on both the particular quality and the conventions of the language in which it is being used. However, Gobl and Ni Chasaide (2003:191) identify the lack of empirical work exploring links between voice and phonation qualities and affective speech, with the majority of knowledge about these links formed from impressionistic observations rather than empirical analysis. Gobl and Ni Chasaide (2003:192) further argue that alongside stronger emotions such as anger, joy and fear, phonation and voice qualities can also signal milder states, moods, attitudes and feelings.

Laver (1994:196) highlights that creaky voice is used habitually by English speakers at turn endings and can be used as both an identity and social class marker among certain parts of the English-speaking world. He (Ibid.) argues that in English, creaky voice also has the paralinguistic function of signalling "bored resignation" when it is used by a

speaker across an entire utterance. Gobl and Ni Chasaide (2003:206) illustrate that listeners rated creaky voiced utterances as signalling of a range of both positive and negative affective states with low activation, including friendliness, relaxation, contentedness, intimacy, sadness and boredom. Furthermore, in an investigation of how paralinguistic and prosodic parameters were used in 100 conversational deliveries of good and bad news, Freese and Maynard (1998) found that creaky and breathy phonation settings were used when speakers delivered bad news, whereas delivery of good news showed preference towards the use of modal phonation settings.

Laver (1994: 197) argues that falsetto has no phonological purpose in language, but can have paralinguistic functions, including being a signal of excitement or a vocal method of mocking speakers. Podesva (2007) explored intra-speaker variation in use of falsetto as a style marker by a gay speaker, "Heath". Podesva (2007:480) argues that to use falsetto voice constitutes performance of "socially marked behaviour at odds with more culturally normative pitch practices for men." Podesva (2007:486) observed that falsetto voice was adopted to varying degrees in different social settings. This lead to the assertion that falsetto voice is used to signal expressiveness and as a method of identity and persona construction. Podesva (2007:486) further highlights that the use of falsetto voice by male speakers extends beyond the standard mean F0 used by female speakers (around 200Hz). While not critically explored by Podesva (2007), this assertion opens the idea that in a forensic context where a male speaker might want to disguise his voice, a falsetto quality may be used in an attempt to sound more stereotypically female. Indeed, I am aware of forensic cases in which this type of disguise has been attempted, albeit with often limited success.

Laver (1994:190) states that in a large number of cultures, whispered phonation is used to signal "secrecy or confidentiality". Gobl and Ni Chasaide (2003:205) found that some listeners in their sample associated whisper with fear, although the authors argue that a better cue for fear for their sample as a whole may be a whispery falsetto phonation quality. Gobl and Ni Chasaide (2003:204) show further associations between whispered phonation, intimacy and timidness, although the authors highlight that whispery phonation was also the most complex to digitally simulate in their study. They therefore advertise caution in making strong inferences about paralinguistic functions of whispery voice from their results (Gobl and Ni Chasaide, 2003:198).

In contrast, harsh voice is described as having an "audibly rough" quality (Laver, 1994:420), and has been labelled as a marker of high activation, high power emotions such as anger, stress, aggression and confidence (Gobl and Ni Chasaide, 2003:204; Laver, 1994:420). Watt and Burns (2012) additionally report that both whispery voice and creaky voice were accurately identified by untrained lay-listeners, and argue that these terms have a greater degree of salience in the minds of lay-listeners compared with terms for other voice qualities such as laryngealisation or velarisation.

2.5.3. Speaking tempo

The speed of a talker's speech can be calculated in two ways; speech rate and articulation rate. Speech rate is generally defined as the number of syllables occurring per second of speech in a given sample (Goldman-Eisler, 1968) irrespective of pausing and hesitations, with articulation rate used to refer to the number of syllables per second of speech per second with pauses of less than 100ms removed (Künzel, 1997). Goldman-Eisler (1968:24) reports typical articulation rate values between 4.4 and 5.9 syllables per second for spontaneous speech, with Gold (2014) reporting mean

articulation rates for spontaneous telephone speech of between 5 and 7.3 syllables per second.

With regards to perceptions of affective states and speaking tempo, Apple et al. (1979) asked listeners to judge interview question responses from male speakers which had been artificially altered for both F0 (30% above and below each speaker's average) and speech rate (20% above and below each speaker's average) for a range of personality traits. Apple et al.'s (1979) results highlighted that both pitch and speech rate appeared to contribute to listener judgements of affective states, even when a verbal channel was present alongside a vocal channel. The authors argue that higher pitched vocalisations can lead to perceptions of a speaker being "less truthful, less persuasive, weaker, and more nervous", with slower speech rate leading to perceptions of a speaker being "less truthful, fluent, emphatic, serious, and persuasive, and more passive", but also "more potent." (Apple et al., 1979:724). Apple et al.'s (1979) results also suggest that the content of an utterance can influence listener judgements alongside vocal parameters. This is an important consideration for the study of spoken threats, given the potential significance of the linguistic content and interpretations of the words used in a threat.

Speech rate has also been investigated in studies on emotion perception, with, for example, Breitenstein et al. (2001) finding an association between slow speech and listener perceptions of sadness. Cruttenden (1986:179) links a faster speaking tempo to increased levels of excitement. Speech rate has also been identified as a linguistic cue to determining whether a speaker is reading aloud, with a slower speech rate generally adopted for read speech than in spontaneous speech. In their report on the speech patterns of Yorkshire Ripper tape hoaxer John Humble, French, Harrison and Windsor Lewis (2006) argue that the speaking rate of 1.64 syllables per second adopted by

Humble in the Ripper tape created "a style of delivery which could be described as measured, slow and purposeful" (French, Harrison and Windsor Lewis, 2006:267). French, Harrison and Windsor Lewis (2006) also cite slow speech rate alongside a range of other prosodic variants including lack of hesitation markers and placement of word stress as linguistic markers which suggest pre-planning, premeditation and read-aloud speech on Humble's part. Applying this to spoken threats, it could be argued that similar slowed speech rate in threatening communications could also lead listeners to infer pre-planning and premeditation surrounding a threatener's intent to harm.

2.5.4. Intonation and lexical stress

Vaissiere (2005:236) highlights that all languages differentiate aspects of meaning through intonation, and that such meaning extends beyond the level of the word and can exist at phrase, paragraph or discourse level. It is further argued that intonation can serve to mark speaker attitudes and emotions, such as arousal, anger, joy or doubt, along with aspects of speaker intention (Vaissiere, 2005:236). Additionally, Gussenhoven (2004:24) argues that phonetic alteration of pitch through the course of an utterance can help signal attitudes such as surprise, excitement and authoritativeness, and that intonation can convey a wide range of attitudes, including emphasis and anger (Gussenhoven, 2004:69).

Banzinger and Scherer (2005:256) argue that there is minimal evidence to suggest that specific emotions, such as anger, happiness, joy, and fear, are directly linked to specific intonation patterns or stylistic manipulation of the F0 contour. However, Banzinger and Scherer (2005:256) also state that intonation appears to vary as a result of emotional speech and that listeners are able to use intonational information to infer emotional and

affective information from the speaker. This somewhat mirrors the current situation regarding phonetic cues to threat, where no specific phonetic markers have been identified despite the more accepted idea that speakers can alter their voice in some way to signal threat, and that these cues would be identifiable by listeners (Watt, Kelly and Llamas, 2013; Kelly, 2014).

Cruttenden (1986:16) states that 'stress' refers to "syllables made prominent for linguistic purposes". While some definitions of 'stress' refer only to pitch prominence, Cruttenden (1986) uses the term in a more general way, referring to the achievement of syllable or word prominence through linguistic means. Vaissiere (2005:249) argues that stress placement links to both focus and emphasis, with and states that displacing sentence stress onto a particular word is one method of marking focus onto that word.

2.5.5. Regional accent

Alongside evaluations of different phonetic parameters, a speaker's accent can also be important in shaping listeners' attitudes and evaluations of speakers. Preston (2002:40) argues that the link between attitudes towards groups and language varieties is "the least surprising thing imaginable", while Watson and Clark (2015) highlight that previous studies (see, for example, Coupland and Bishop, 2007; Giles, 1970; Labov, 1972, Preston, 2002) have shown that accent stereotypes appear to be widely held, with some holding stable across time. Edwards (1982:25) states that speech samples evoke stereotypes that reflect how social groups are viewed, with standard accents in the UK typically rated as having higher status and competence than regional accents. Conceptual accent evaluation studies in the UK by Coupland and Bishop (2007), and Giles (1970) have shown that standard accents are generally rated more positively in

terms of prestige and social attractiveness than urban, non-standard accents. However, Giles and Billings (1999:195) highlight that in many cases, non-standard speakers are perceived more favourably on traits relating to aspects such as solidarity, integrity and benevolence.

The implications of accent evaluation has also been examined in various legal settings. Kalin (1982:148) states that accents are the source of many inferences about speakers, and that this is particularly important in legal settings where a vast array of opportunities for reactions to varieties are available, with potentially life-changing consequences. One such study comes from Dixon, Mahoney and Cocks (2002), who examined the extent to which regional accent could affect listener attributions of guilt alongside the type of crime committed and the race of the speaker. Their study used guises for Birmingham and RP accents, blue collar and white collar crimes, and black and white speakers, and found that the Birmingham accent was generally rated as sounding guiltier than the standard RP accent. The study also found that the Birmingham accent/blue collar crime/black speaker guise was rated as being significantly guiltier than the other five combinations. Dixon, Mahoney and Cocks (2002:166) hypothesise that one reason for this effect may be that speakers with non-standard accents are more commonly associated with negative or criminal stereotypes.

In a follow-up study, Dixon and Mahoney (2004) examined the effect of introducing two contrasting types of evidence, strong and weak, into their guilt evaluation paradigm. Unsurprisingly, this study found that listeners attributed higher ratings of guilt when the evidence against the speaker was strong than when it was weak. However, no effect was found which said that accent evaluation contributed to stronger or weaker attributions of guilt when the evidence was either strong or weak. The Birmingham guise was,

however, rated as being more typically criminal and more likely to be accused of committing a crime than the standard accent guise. Dixon and Mahoney (2004:71) argue that the provision of evidence, be it strong or weak, could cause listeners to focus away from character evaluations of the subject when making assessments of guilt. Their study did, however, still show that broader criminality stereotypes appear to be linked to accent evaluations. Dixon and Mahoney (2004:71) argue that further research is needed to understand how accent evaluation shapes psycho-legal judgements, and the extent to which existing results can be generalised to other psycho-legal areas.

An initial investigation into the idea that a speaker's accent could also contribute to listeners' perceptions of spoken threats is presented in Tompkinson (2016). This study examined a range of direct and indirect threats in three accent guises, London Cockney, Northern Irish and Received Pronunciation. The results showed that in the indirect threat condition, the urban non-standard London Cockney accent guise was rated as being significantly more threatening than the Northern Irish and RP guises, but that this effect was not replicated in the direct condition. There was no effect for accent in the direct threat condition. As would be expected, the results also showed a strongly significant difference between the direct and indirect threat stimuli, with direct threats rated by listeners as sounding more threatening. The research asserts that the overtly non-standard, urban accent guise was rated as sounding the most threatening when the words in the stimuli did not overtly signal a threat (Tompkinson, 2016). There are parallels between this result and those presented by Dixon, Mahoney and Cocks (2002), who argue that non-standard accents can be negatively stereotyped in legally relevant judgements. Results presented in Tompkinson (2016) also suggest that listener age and listener geographical background can influence accent evaluation with respect to spoken threats. The study showed data trends which suggest older listeners rated a Northern

Irish bomb threat as being more threatening than younger listeners who, it is argued, perhaps have less awareness of the link between Northern Ireland and bomb threats than older listeners for historical reasons. It also found that Southern English listeners rated the London Cockney indirect threats as sounding less threatening than Northern English listeners. This links into Coupland and Bishop's (2007) assertion that people have more positive associations towards accents closer to their own, and Montgomery's (2007) argument that geographical proximity can influence accent evaluation judgements.

2.5.6. Methodological considerations in affective speech research

While the link between phonetic variables and the perception of affect, emotion and personal characteristics has been well researched, there are methodological issues with much of the work carried out in this area in relation to examinations of spoken threats. For example, the method adopted by Chuenwattanapranithi et al. (2009) used synthesised vowels which were artificially altered for F0 and vocal tract length. While the use of single synthesised vowels may be acceptable for investigating certain emotions, and indeed has been argued to be a superior method on grounds that it mitigates any effect of verbal content or unwanted prosodic influences, the inference of threat from speech involves aspects of both the verbal and vocal channel. Therefore it is argued here that both of these aspects should be represented in perception experiments which seek to examine how listeners infer threat from speech.

Potential problems also exist with making automatic links between threats and certain emotions. This is highlighted by Kelly (2014:7), who argues that attributing threats to certain psychological and/or emotional states is highly problematic and that the goal of future linguistic research in this field should be to lessen reliance on such links when

attempting to understand or analyse spoken threats. Watt, Kelly and Llamas (2013:100) further highlight this issue, and while acknowledging likely links between anger and threat, they also state that "we must avoid conflating angry speech with threatening speech", as "wishing to threaten someone does not presuppose that the threatener is angry with the recipient" (Watt, Kelly and Llamas, 2013:100). Watt, Kelly and Llamas (2013:101) further illustrate the difference between threat and anger by pointing out that being angry with someone is not an offense, it can be an offense to threaten to harm someone. Here, it is argued that any links between the two emotions are better left to be established through empirical research rather than through general presupposition about automatic associations between threats and affective states.

Chuenwattanapranithi et al. (2009:3) argue that attempts to relate multiple acoustic aspects of the speech signal to different emotions has, as of yet, failed to create appropriate models of emotional speech. However, this approach contradicts research findings that have shown that humans can accurately detect emotion from vocal cues (Chuenwattanapranithi et al., 2009:3). It can be argued that this imbalance between production and perception is appropriate for research on threatening speech. While it may appear improbable that there are direct, one-to-one links between acoustic properties of speech and the production of threats (see Kelly, 2014), there may be more widely held perceptual properties which relate specific acoustic cues to perceptions of threats, as hinted at by Ohala (1984). This further opens the possibility for the exploration of misconceptions related to what a 'typical threatener' sounds like, especially given Gales' (2010) assertion that supposed features of 'threatening language' become enregistered in the minds of listeners. This could link to the Danish case outlined in Chapter 1, where a threatener felt he was been treated unfairly because

he had a deep voice and was therefore, in his opinion, automatically perceived as sounding as though he was angry when he was not.

Bachorowski (1999) argues that previous investigations into the production of emotions via the mechanism of speech are based on the notion that different acoustic cues are used to signal discrete emotional categories. These are then tested through perceptual studies which test listeners' abilities to recognise the emotion played to them (Bachorowski, 1999:55). However, Bachorowski (1999:55) argues that the complex nature of speech makes this approach somewhat simplistic. This is particularly true in the case of threats, which cannot be underpinned to a single emotional state. Furthermore, it can be argued that indirect threats, by nature, cannot be classified as either threats or non-threats with any degree of certainty. Applying Bachorowski's (1999) argument to threat perception, it can be said that it is important to acknowledge that some listeners may find the same utterance to be more or less threatening than other hearers, and that threat perception can be both a relative and absolute process. It is further argued that existing research on listener perceptions of emotions has failed to account for any relationship between speaker and hearer and "the intended impact of vocal signals on the listener's affective states" (Bachorowski, 1999:55), which is crucial for the analysis of spoken threats.

2.6 Threat typology and research framework

2.6.1 - A threat typology and working definition

Gales (2010) argues that a large proportion of work conducted examining threats has focused on behavioural characteristics of the threatener rather than their use of language, and that there is still a lack of understanding about "how threateners successfully threaten" (Gales, 2010:2). This could be considered somewhat surprising given the status and classification of verbal threats as potentially serious language crimes. Furthermore, Gales (2010:27) highlights that the majority of threats analysed by law enforcement agencies and threat assessment professionals are anonymous, leaving language as the main form of evidence which is available for analysis.

It is clear however, that linguistic research into threats should avoid any temptation to move towards making assessments and judgements about a speaker's psychological state. I argue here that this should be considered an issue in psychology and not something that should be commented on by linguists and forensic phoneticians. Indeed, point 9 in the International Association for Forensic Phonetics and Acoustics' code of practice states that "[M]embers should not attempt to do psychological profiles or assessments of the sincerity of speakers (IAFPA, 2004). This point is highlighted by Watt, Kelly and Llamas (2013:103), who state that the role of their perceptual investigation into spoken threats was not to comment on speakers' sincerity or identify phonetic traits which may mark sincerity (see Kirchhübel, 2013), but rather to explore listeners' responses towards speech samples produced in both a 'threatening' and 'neutral' tone of voice. The framework for analysis adopted in this thesis is much the same as that taken by Watt, Kelly and Llamas (2013). Throughout the thesis, listeners'

responses and judgements to various aspects of speech will be examined in order to attempt to gain a fuller understanding of how listeners infer threat from phonetic aspects of speech. In doing so, I argue that this is one way in which linguists can explore human motivations for inferring threat without performing psychological profiling of a 'typical threatener' from vocal cues, or attempting to identify whether it is possible to determine whether or not a threat is 'real' from vocal cues alone.

It is also necessary to consider the scope of this research aim within a wider acknowledgement of what threats are and how they work. Milburn and Watman (1981:10) argue that there are five important elements which all contribute to the system under which threats are communicated:

1. A medium of communication
2. A source
3. A target
4. An audience
5. A situational context

However, the complexity surrounding threats as a type of language crime means that further clarification of a threat typology is needed beyond the five points identified by Milburn and Watman (1981), and listed above. Figure 2.4 details five key criteria which I argue are essential for the communication of a verbal threat. These criteria are consistent with Storey (1995) and Gales (2010), who both argue that shared understanding between speakers and hearer is a requirement for the successful communication of any threat. In Figure 2.4, points 1-3 relate to the use of language, with points 4 and 5 relating to wider situational contextual factors which are also

essential considerations in threat communications. Figure 2.4 has been designed to reflect an increasing level of abstraction away from language use from point 1 through to point 5.

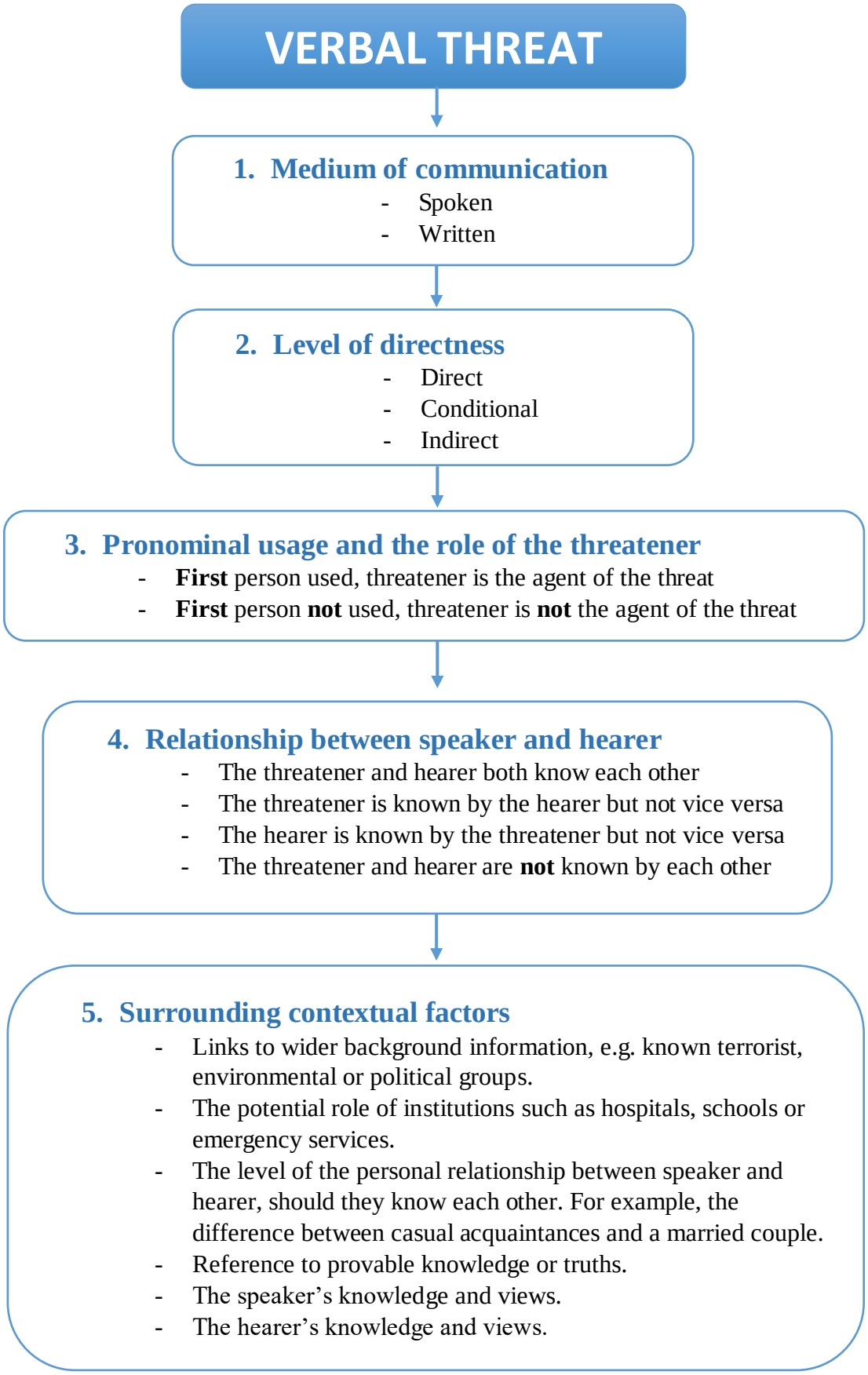


Figure 2.4 – Factors relevant to the communication of a verbal threat

Point 1 in Figure 2.4 relates to the medium of communication used to deliver a verbal threat. In Section 2.2.1, it was identified that verbal threats can be delivered through either writing or speech. It should also be acknowledged that while this is the case, there is a substantial difference between the two, and any linguistic research into threats should make clear whether the focus is on written or spoken threats. The research presented throughout this thesis exclusively focuses on spoken threats and does not consider written threats beyond this chapter.

The second point in Figure 2.4 refers to the level of directness within a given threat. Three different levels of directness were discussed in Section 2.2.2: direct, indirect and conditional. Each of these three levels should be factored into a linguistic typology of threatening language, but again it is important to distinguish between the three when considering a research project of the kind conducted for this thesis. Results from Tompkinson (2016) provide one example of how direct threats were perceived differently to indirect threats, both in terms of the overall level of conveyed threat and the effect of a speaker's regional accent on the level of perceived threat in the two different utterance types. It is also acknowledged that each of these threat types could operate through either spoken or written mediums. However, despite the categorisation of threats as either direct or indirect, there is a great deal of fluidity both within and between the two categories. Consider, for example, the two utterances detailed in (1) and (2), below.

(1) – *I'm warning you about a bomb at York Station. It will go off this afternoon.*

(2) – *I know where you live.*

Both utterances in (1) and (2) can be classified as indirect threats. The utterance in (1) would be classified as an indirect threat owing to the possible interpretations as either a warning or a threat. Gales (2017, personal communication) classifies the type of utterance used in (1) as a direct performative warning, but an indirect threat. If the utterance is interpreted literally, then it is a direct and clear warning owing to the use of 'warn' as a perfomative verb, whereas the threat interpretation requires listener inference as to the speaker's intentions. However, the severity of the action mentioned in (1) is both clear and of a high level, and the utterance also mentions both a clear time and a place. The utterance in (1) is arguably more direct than the utterance in (2), which requires a greater level of listener inference to arrive at a threat interpretation, despite the collective indirect classification. Here, I argue that the classification of a threat as either direct or indirect provides a base level of classification, with more nuanced and fluid classifications present within these overarching categories.

Point 3 in Figure 2.4 relates to the role and position of the threatener within a given threat. The main linguistic feature encompassed within this is the use of either first person or third person pronouns by a threatener. The use of first person pronouns serve to position the threatener as the agent of a threat, whereas the use of third person pronouns position the threatener away from, or independent to the threatened action. Building on the first and second points in Figure 2.4, the use of first or third person pronouns could relate to either written or spoken threats, and also either direct, indirect or conditional threats. This is exemplified below, with (1a) and (1b) illustrating the difference in threatener position and pronominal usage in direct threats, (2a) and (2b) show the difference with indirect threats, and (3a) and (3b) exemplify the difference with conditional threats. In each set of examples, (a) shows a threat where the threatener

is the agent of the threatened action, whereas (b) shows a threat where the threatener is not the agent of the threatened action.

(1a) – *I'll break both of your legs for sleeping with my girlfriend.*

(1b) – *Your legs will be broken for sleeping with Gemma.*

(2a) – *I'm warning you about a bomb at York Station. It will go off this afternoon.*

(2b) – *There's a bomb at York Station. It will go off this afternoon.*

(3a) – *If you don't pay me the money I'm owed, I'll break both of your legs.*

(3b) – *If you don't pay £1000, both of your legs will be broken.*

On the direct-indirect continuum discussed above, it can be argued that direct threats are more likely to place the threatener as the agent of the threatened action, whereas removing the threatener as the agent of the threatened action through the omission of first person pronouns could be part of the range of linguistic features that increases the indirectness of a threatening utterance.

Points 4 and 5 in Figure 2.4 relate to wider contextual factors which could also influence the interpretation of a potential threat. These contextual factors are distinct from the linguistic factors discussed under points 1, 2 and 3, but are nonetheless important when considering how threats are made and communicated. Point 4 in Figure 2.4 details the different possible relationships between the threatener and the hearer. The first of these possible relationships is one where the threatener and the hearer both know each other. When threats are made and the speaker and hearer are familiar with one another, it is likely the case that the contextual information introduced as a product of

the relationship between the speaker and hearer will play a greater role in the interpretation of a potentially threatening utterance. For example, if someone exclaims "*I'm going to punch your head in"* to their closest friend, the relationship between the speaker and the hearer could be sufficient to mitigate the threatening nature of the words used. However, if the same utterance was produced by a speaker to an unfamiliar hearer, the anonymous relationship between speaker and hearer would provide less contextual information to mitigate the interpretation of the utterance as a threat. Two further possibilities are detailed in point 4 of Figure 2.4 with respect to the relationship between the speaker and hearer of a threat. The first is that the hearer knows the identity of the threatener but not vice versa, and the second is that the threatener knows the identity of the hearer but not vice versa. The latter would likely be the type of relationship seen in anonymous stalking cases involving threats, where a stalker threatens a victim who is familiar to them, but the identity of the stalker is not known to the victim.

The relationship between speaker and hearer with respect to threats is further complicated when threats made to institutions such as hospitals, schools or the emergency services are made. In these cases, it is likely that the speaker and hearer will be unfamiliar with one another. However, there is more contextual information introduced into the speaker-hearer relationship in cases involving threats to institutions because of the institution itself. For example, consider the school bomb threat case discussed in Section 1.1. If a speaker makes a targeted bomb threat to a school via telephone communication, it is unlikely, although not impossible, that the person who answers the telephone in the school reception will know the identity of the threatener, or vice versa. However, the fact that the threatener has targeted the school as an institution

reduces the personal nature of the threat and therefore the interpretation is likely to be less affected by the direct relationship between speaker and hearer.

Point 5 in Figure 2.4 details a range of additional contextual factors which could all influence the delivery, communication and interpretation of a potentially threatening utterance. As previously stated, such factors are largely independent from the linguistic features of threatening utterances, but are, nonetheless, important to consider in any taxonomy for threats as a type of communicative language crime. Two of the most unpredictable factors that fall into this category are the knowledge and views of the world brought to the interpretation of an utterance by both the speaker and the hearer. These factors were discussed in Section 2.2.4 in relation to the role of the speaker and the hearer in making and communicating threats, and it should be acknowledged that the different knowledge and views that each individual speaker and hearer has will likely affect the interpretation of a potentially threatening utterance.

A project of the size and scope of the research presented in this thesis cannot consider all types of threats or all the different permutations of influencing factors discussed in this section so far. In order to avoid over-interpretation of any of the research findings, it is therefore necessary to delimit the scope of the research presented in the following chapters in relation to the range of possibilities displayed in Figure 2.4. In order to keep the analysis more tightly constrained, the focus of the research in this thesis will centre on the perception and interpretation of indirect threats by different listeners. The experiments presented in the following chapters were designed to test listener responses to various phonetic aspects of speech and their influence on threat perception. The overarching goal of the work presented is to build towards providing critical analysis of how listeners infer threat from phonetic aspects of a speaker's voice, working within a

communication framework that includes a medium, a source, a target, an audience and some form of situational context. Using the same outline as Figure 2.4, Figure 2.5 shows the delimited framework that will be the focus of the work presented in the remainder of this thesis. This will be taken forward as a working framework, upon which the experimental research is based and designed. Emboldened sections represent inclusion in the working framework, with greyed-out sections denoting categories not included within the working framework. Figure 2.5 highlights that the focus of the thesis will be on spoken indirect threats, produced with both types of speaker positions (agent of the utterance vs not the agent of the utterance). The relationship between speaker and hearer will be one where both parties are unfamiliar with each other in order to mitigate the effect of any personal relationships between speaker and hearer. The research in Chapters 4, 5 and 6 will include the mention of the emergency services in order to provide a situational context, and the later experiments will also examine differences between individual listeners in an acknowledgement of the fact that different listeners will bring different knowledge and views to the interpretation of any potentially threatening utterance.

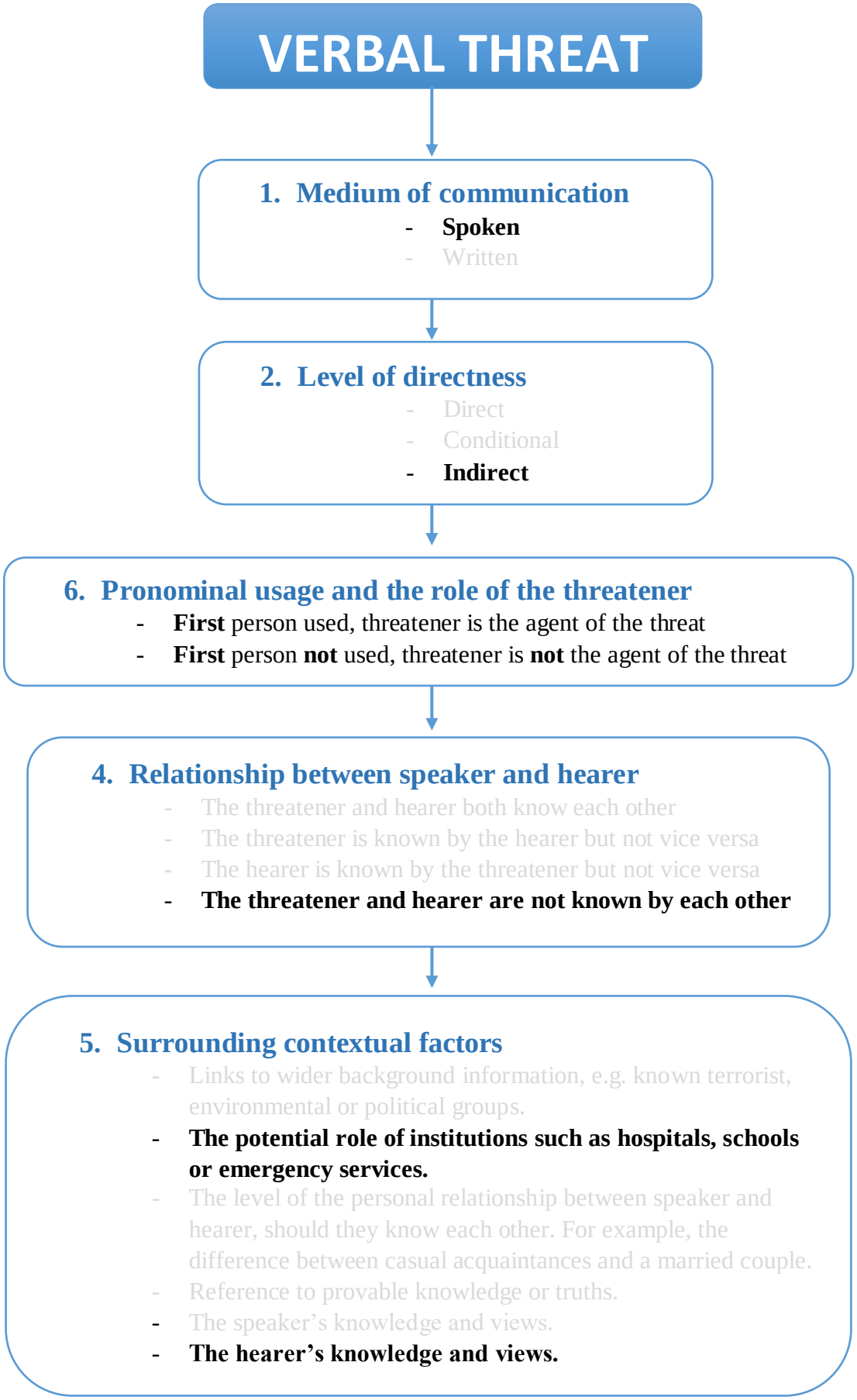


Figure 2.5 – A working framework for the experiments presented in this thesis

The work in this thesis aims to provide a more thorough analysis of the linguistic and phonetic factors that can drive the inference of both perlocutionary and illocutionary effects from potentially threatening utterances within the framework outlined above.

2.6.2 – Experimental research approach

Before presenting the experimental chapters in this thesis, it is necessary to consider both the scope of the research, and the experimental approach used, for the research in this thesis. It should be stated at the outset of the thesis that the work presented is inherently experimental in nature. The stimuli in each experiment were created specifically for use in the experimental research projects presented in each chapter, and do not come from real casework data or any other real-world sources of spoken threats.

While the use of real-world data would have the advantage of being genuinely authentic, the sparsity and lack of availability of such data meant that the approach of using recordings taken from casework was judged to be broadly non-advantageous in order to answer the questions outlined at the beginning of this thesis. In real-world examples of spoken threat recordings, it would also be difficult to ascertain ground-truth knowledge of speakers' backgrounds and the context in which the recordings were made. Furthermore, it was not considered ethically valid or appropriate to play such data to multiple listeners in perception experiments of the type conducted and reported through this thesis. Using an experimental approach mirrors the approach taken in a range of research work on the perception of indirect threats (Watt, Kelly and Llamas, 2013), emotional trait perception, social trait perception and evaluative accent perception.

The majority of the studies presented in this thesis also used speakers who were not trained actors. Instead, participants were volunteer members of the general public, recruited predominantly from the student population at the University of York. Using stimuli produced by non-actors was highlighted by Watt, Kelly and Llamas (2013:118) as potential methodological improvement on studies using trained actors to produce spoken threat stimuli, arguing that "the majority of people who threaten one another with harm are not trained actors". The use of experimental stimuli also facilitated full researcher control over the recordings, allowing for phonetic alterations to be made where necessary, and researcher choice over the appropriateness of both speakers and individual stimuli.

However, while taking an experimental approach to answering the questions set out at the outset of this thesis had certain procedural and research-based advantages; questions remain as to how far the research findings can be expanded to real-world scenarios involving spoken threats. The evaluative settings in which the experiments took place were considerably more favourable than would be expected should spoken threats be heard by either jurors or earwitnesses in real-world evaluative settings. Among other factors, participants were permitted to listen to the recordings as many times as they wished to, and the environment for the experiments in Chapters 4, 5 and 6 was pre-determined by the researcher and designed to be quiet. Listeners were also provided with high-quality closed-cup headphones for the judgement tasks in Chapters 4, 5 and 6. One aim of conducting the experiments in this way was to attempt to reduce the number of potentially extenuating environmental factors that could have influenced listeners' evaluative judgements of speakers' voices. However, while such an environment was considered appropriate for the purpose of the research presented in this thesis, it is

acknowledged that this environment does not accurately reflect real-world earwitness evaluation tasks.

One further issue which should be acknowledged in relation to the design of the experiments presented through this thesis is that the nature of the repeated-stimuli, multiple-speaker designs also does not closely reflect real-world voice evaluation tasks. Having listeners evaluate multiple voices within a single experiment creates a situation where voices and speakers can be directly compared to one another, with relative listening forming part of the evaluative process. This contrasts with real-world earwitness situations, where an earwitness to a bomb threat is only likely to hear the voice of a single speaker in a single instance. However, the amount of participants that would be required to take part should listeners only be exposed to one voice within an experiment would be unrealistically high for a project of this size and scope. For example, the experiment in Chapter 6 elicits over 1000 voice evaluations, and recruiting over 1000 listeners for this project would not have been possible. Furthermore, the repeated-stimuli design also allowed for an assessment of the evaluative patterns of individual listeners.

However, despite these shortcomings, the primary benefits of using experiments to answer the research questions outlined in Chapter 1 are that they allow for a greater level of researcher control and the systematic testing of the influence of specific phonetic parameters on listeners' evaluative judgements. These advantages were judged to be more advantageous than the shortcomings discussed above, and therefore it was decided that an experimental approach was the best research approach to take in this thesis.

2.6.3 – Research impact

It is also important to directly and critically address the potential wider impact of the research presented in this thesis from the outset. The implementation of the Research Excellence Framework in the UK from 2014 explicitly evaluated the so-called 'real-world' impact of academic research (McIntyre and Price, 2018:4). In fact, the mention of impact in the introduction to this thesis arguably reflects the changing, and important, role that it now plays in the delivery and assessment of academic research. Commenting on the role of impact in Forensic Speech Science (FSS) research, French and Watt (2018:150) argue that while research in FSS can often be viewed as having a high degree of real-world relevance, researchers should avoid being both "unduly optimistic" and "too casual" (French and Watt, 2018:161) about the impact that research in forensic phonetics and forensic linguistics can have on the real world. Further to this, French and Watt (2018:161) also warn of the dangers of research in both forensic phonetics and forensic linguistics being seen as areas which automatically "tick the impact box". In other words, the fact that a linguistic research topic sits in the forensic domain does not automatically make the research impactful, and research projects in the area should not be designed with the sole intention of being impactful. As McIntyre and Price (2018:3) argue, there is value in research that is designed to be directly impactful and research which, at the surface level, does not have immediate impact on a given area in the 'real-world'.

Following the working framework outlined above, the research in this thesis aims to begin to address the previously-highlighted disconnect between linguists and non-linguists surrounding the perceptions of spoken threats as an under-researched type of language crime. In exploring the assumptions that people without advanced linguistic and/or phonetic training make about speakers from aspects of voice through a series of

perception experiments, it is hoped that some suggestions for practical and realistic improvements to procedures for the evaluation of certain types of threat could be suggested. Or, perhaps more minimally, that the research undertaken in this thesis could act as a springboard for further discussions with police officers and security policy makers who are tasked with providing advice to those who are required to deal with spoken threat evidence at different stages of the legal process. The focus of the impact of the work presented in this thesis is anticipated to primarily be earwitness contexts in which judgements about potential threateners are made. The focus of the later experiments in the thesis is on earwitness evaluations of indirect bomb threats by anonymous or unknown threateners. This was designed to mirror the context in which the NCTSO bomb threat evaluation checklist is designed to be used, where a listener is unfamiliar with a speaker but could be required to provide inferences about a speaker's intentions or a description of a speaker's voice. It is hoped that the research in this thesis can provide both more knowledge about, and a general assessment of, the evaluative and descriptive tendencies of listeners when they are tasked with describing and making judgements about speakers' voices.

In relation to the different forms of impact discussed previously, this would fall under the definition of anticipatory research (French and Watt, 2018:153). This type of research can be considered as research that has potential for real-world applications, but doesn't respond directly to a specific case or an urgent, immediate need for data.

Chapter 3 – Exploratory experiments

3.1. Introduction

Given the sparse amount of research specifically examining "the phonetics of threat" (Watt, Kelly and Llamas, 2013:100), the goal of this chapter is to present research which begins to critically examine whether specific phonetic aspects of the vocal channel can influence listeners' evaluations of how threatening speakers sound. Two experiments are presented in order to begin to address this issue, with both designed to provide results which will further inform the analyses presented in subsequent chapters of this thesis.

Experiment 1 builds on work presented in Tompkinson (2016), which used a matched guise design to illustrate that indirect threats produced in a non-standard regional London Cockney accent were perceived as sounding more threatening than the same indirect threats produced in a standard Received Pronunciation accent. Experiment 1 extends this finding by examining the relative contributions of speaker accent and mean fundamental frequency (F0) on listener evaluations of two contrasting indirect threats. In doing so, it forms an initial attempt to examine the relative effects of both social and phonetic variation on listener threat perception. Experiment 1 also examines differences between threat ratings assigned to two different indirect threat utterances, and the relative effects of average F0 and speaker accent on these two contrasting indirect threats.

The second experiment presented in this chapter builds on the design of Watt, Kelly and Llamas (2013) and examines the relative effects of three phonetic variables on the perception of indirect threats produced in unfamiliar foreign languages by multiple

speakers. The goal of this experiment is to examine whether phonetic variables can influence listener evaluations of spoken threats in the absence of an interpretable vocal channel. The variables examined in Experiment 2 are mean F0, speech rate and F0 range, and the work presented in Experiment 2 further builds on the analysis presented in Experiment 1 by considering the effect of factoring multiple speakers into the experimental design.

While it is acknowledged that other aspects of the vocal channel such as voice quality (Xu et al., 2013), stress placement and intonation patterns (Scherer, 2003) could also influence listener evaluations, consideration of these factors was beyond the scope of the experiments presented in this chapter. Additionally, despite Milburn and Watman's (1981) assertion that context is a key influencing factor in threat evaluation, it was also not considered in the experiments presented in this chapter. Given the minimal amount of previous research on the phonetic analysis of spoken threats, it was considered necessary to initially examine whether a set of narrowly defined phonetic variables could influence threat perception in a context-less scenario. Analysis presented in subsequent chapters builds on the results of the experiments presented here by considering the effect of both contextual information and other aspects of the vocal channel.

3.2. Experiment 1

3.2.1. Methodology

The data used for Experiment 1 were comprised of modified versions of a subset of data collected for the previously outlined project presented in Tompkinson (2016). The recordings were produced for the experiment in Tompkinson (2016), and modifications

were made to them for the purpose of the current experiment. Two indirect threats – *"I know where you live"* and *"I wouldn't do that if I were you"* – were recorded in RP and London Cockney accent guises by one male speaker and used as experimental stimuli. In order to create three different mean F0 levels for each stimuli, a Praat pitch alteration script (Fecher, 2015) was used to alter the mean F0 level of each recording to 90Hz (low), 115Hz (mid) and 140Hz (high). The low and high values are 25Hz above and below an approximation of an average male mean F0 level, as reported by various phonetic analysts (Hudson et al. 2007; Künzel, 1989; Lindh, 2006). They also represent values in the highest and lowest 10% of population values reported by Hudson et al. (2007) for 100 male speakers of Standard Southern British English. The mid value of 115Hz represents an approximation of the average male F0 level for speakers of English. All recordings were checked post-alteration to ensure that no digital artefacts had influenced the sound quality as a result of the editing process. The alteration procedure also preserved the shape of the intonation contour while altering the average pitch for each recording. Finally, each F0-altered recording was band-pass filtered between 300Hz and 3400Hz in an attempt to replicate the frequency range of the landline telephone channel (Künzel, 2001; Nolan et al., 2013). This was done following Eriksson's (2005:8) assertion that threats encountered in forensic phonetic casework are commonly made over the telephone.

40 participants (age range 18-53, mean age 23, SD 8.4; 29 female) gave informed consent to take part in an online survey designed to obtain attitudes towards the experimental stimuli. Participants were instructed to listen to each recording and provide ratings of how *intelligent*, *aggressive*, *threatening*, *angry*, *friendly*, *menacing* and *agitated* they thought the speaker sounded using a seven-point Likert-type scale (1= "Not at all… X", 7= "Very… X"). The use of this scale reflected the idea that some

listeners may find the same utterance more or less threatening than other listeners. Additionally, the inclusion of other traits alongside *threatening* was designed to ensure that listeners' attentions would not be solely focussed on how threatening speakers sounded. It also facilitated an analysis of the relationship between the attribution of threat ratings and judgements of other potentially related traits such as anger, aggression and menace.

In order to address Preston's (2002:42) argument that it is often unclear whether listeners in accent evaluation experiments assign the accents they hear to appropriate group categories, listeners were asked to state where they thought the speaker in each stimulus they heard was from. Listeners were also asked to describe what they thought the speaker in each recording they heard looked like, with particular reference to height and build. This was done to further examine the potential link between body size perception and vocal pitch, following assertions (see, for example, Ohala, 1984; Puts et al., 2006) that the two are perceptually related.

Foil voices were also incorporated into the experiment, with an equal number of target and foil recordings presented to listeners. Some foils were explicit direct threats recorded for a previous experiment (Tompkinson, 2016) and others were taken from the IViE corpus (Grabe, Post and Nolan, 2001). These two contrasting foil types were chosen to place the indirect threat target utterances somewhere between utterances designed to be overtly threatening and utterances designed to be not at all threatening.

Given Montgomery's (2007) argument that geographical and phonological proximity between speaker and hearer may affect dialect ideologies and perceptions, information was also gathered about listeners' geographical background. Listeners were separated

into Northern (n=30) and Southern (n=10) geographical background categories. While there is much debate over the concept of a linguistic North/South divide (Wales, 2000), Trudgill's (1990) proposal of a dividing line running from The Wash to Shropshire was adopted for the purposes of this experiment. Despite RP's position as a social accent of the UK rather than a regional accent, Hughes, Trudgill and Watt (2012:3) argue that northerners tend to view it as a southern accent, and therefore splitting listeners into Northern and Southern categories was considered an appropriate distinction to make within the design. The purpose of the North/South distinction was to evaluate whether those listeners from the 'linguistic South' evaluated the samples differently from listeners from the 'linguistic North'.

Statistical analysis was conducted in R (R Core Team, 2015) using linear mixed effects regression models (hereafter `lmer`) constructed using the `lme4` package in R (Bates et al., 2015). The data collected for this experiment are from Likert-type scale responses and are therefore not strictly linear; instead being classified as ordinal-level data rather than interval-level data. Langdridge and Hagger-Johnson (2009) state that while Likert scale data are strictly ordinal because it cannot be assumed that the distance between points are equal, this is something that researchers who use such scales frequently do indeed assume. They further argue that it is acceptable to make this assumption provided that the scale being used has at least five points (Langdridge and Hagger-Johnson, 2009:47). Further debates have taken place in the literature surrounding the use of parametric statistical testing with response data collected through Likert scales. A summary is provided by Norman (2010), who argues that it is acceptable to use parametric statistical testing with ordinal-level data collected from Likert-scale response tasks, despite this being a common source of reviewer criticism of such work. For example, Jamieson (2004) criticises research which assumes that, for example, the

distance between ‘disagree’ and ‘strongly disagree’ can be assumed to represent an equal distance in attitudes towards a given statement or stimuli. This is acknowledged here, and an attempt to guard against this was introduced by using a numerical Likert scale which presented listeners with extreme ends of perceptual responses; for example “not at all threatening” and “very threatening”, with a 1-7 numerical scale between these two points. Furthermore, in order to perform statistical analysis on Likert scale data which uses categories such as “strongly agree” and “strongly disagree”, is it necessary to assign numbers to the categories anyway, so I therefore argue that this approach is a logical one that attempts to guard against some of the criticisms proposed by Jamieson (2004).

Norman (2010) highlights that three frequent criticisms of using parametric statistical analysis on data similar to the data collected for the experiments throughout this thesis are small sample sizes, non-normally distributed data and Likert-type scale response data. While acknowledging that there is a theoretical and technical correctness to such criticisms, Norman (2010) argues that parametric statistical tests are powerful enough to ensure that the chances of reaching incorrect conclusions as a result of using a parametric statistical test as opposed to a non-parametric test are small. Norman (2010:7) concludes:

> *“Parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of ‘‘coming to the wrong conclusion’’. These findings are consistent with empirical literature dating back nearly 80 years. The controversy can cease (but likely won’t).”*

In this experiment, *p*-values for the main effects were calculated through model comparisons, constructed using Chi-Square tests with the `anova` function in R. This approach is suggested by Winter (2013) as a way of obtaining *p*-values from linear mixed effect regression models. To test the significance of each variable in the full model, a reduced model was constructed for each variable of interest. These models were identical to the full model with the exception of having the variable of interest removed. A Chi-Square comparison was then conducted between the full and reduced models for each variable. Winter (2013) states that the p-value obtained from the model comparison can be used to provide a measure of statistical significance for the variable of interest. Further analysis of within-variable effects was conducted using Holm-Bonferroni corrected Tukey pairwise comparisons, constructed using the `multcomp` package in R (Hothorn et al., 2008).

Within the model, listeners' ratings of how threatening the speaker sounded formed the dependent variable. In Experiment 1, mean F0, speaker accent guise, listener sex and listener geographical background were included as fixed effects, along with an interaction between mean F0 and speaker accent. Listener and utterance were included as random effects.

3.2.2. Results

3.2.2.1. Listener accent descriptions

Preston (2002:42) argues that one weakness of matched guise designs is that it is often unclear whether listeners assign the voices they are rating to the intended target group categories as opposed to another perceived group. For example, in this experiment it should be clear that listeners were able to identify the London Cockney and RP accent

guises to the appropriate target group, rather than to, say, speakers from the north of England. In order to reach conclusions about potential accent biases in the data, it should be clear that the group of listeners were able to identify the target accents with reasonable accuracy. To address this potential weakness and provide additional validation of the matched guise stimuli, listeners in Experiment 1 were asked to state where they thought the speaker in each recording they heard was from. The ten most frequent answers provided by listeners for both accent guises are presented in Figures 3.1 and 3.2. The London Cockney guise is shown in Figure 3.1 and the RP guise in Figure 3.2. Given the nature of the experiment, multiple responses from each listener are collated in Figures 3.1 and 3.2.

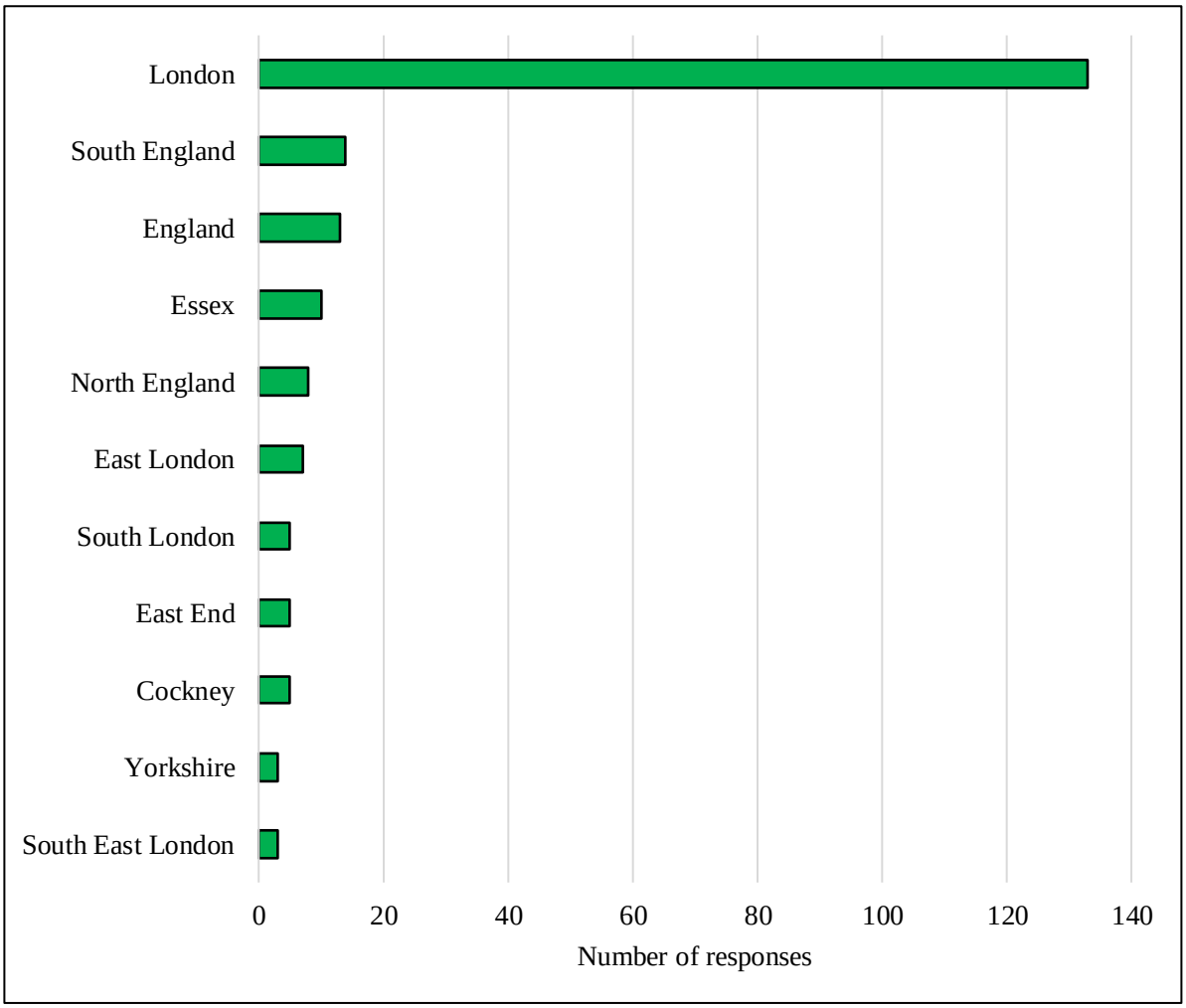


Figure 3.1 – Accent attributions for the London Cockney accent guise (Experiment 1)

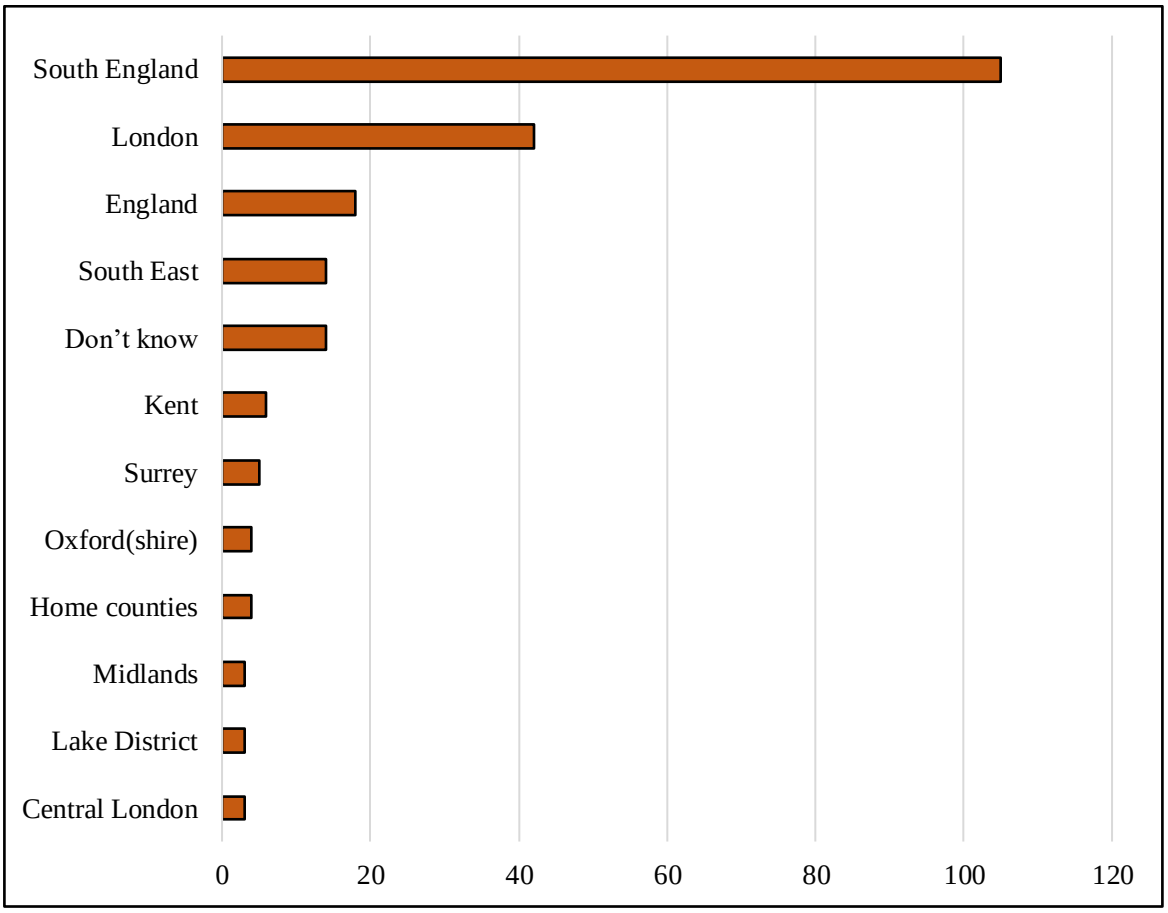


Figure 3.2 – Accent attributions for the Standard Southern British English accent guise (Experiment 1)

Figures 3.1 illustrates that 74% of responses identified the London Cockney accent guise as being from London or the surrounding area, with a further 12% of responses providing other accurate but less locally defined responses, such as *South England* and *England*. Listeners generally identified RP to suitable areas of Southern England by providing either general responses such as *South England, England, the Home Counties* or specific counties such as *Kent* and *Surrey* (Figure 3.2). Although there were some errors in identification, such as the attribution of the London Cockney guise to *Yorkshire*, the overwhelming majority of responses provided an accurate identification. The data in figures 3.1 and 3.2 do, however, highlight a potential issue with the use of *London* as an accent descriptor as it was frequently assigned to both the RP and London

Cockney accent guises. This possibly reflects the fact that the city of London, as Britain's geographically largest and most populated city, is associated with both standard varieties of British English alongside more traditional dialect forms. It can therefore be argued that the listener group as a whole were predominantly able to correctly identify the accent guises to appropriate regions. This provides some validation that the vocal guises accurately represented the accents they were intended to portray.

3.2.2.2. Effects of phonetic variables on listener threat evaluations

Table 3.1 shows significance values calculated for each of the fixed effects on listener perceptions of how threatening speakers sounded from the `lmer` model outlined in Section 3.2.2. As previously stated, mean F0 (*high/mid/low*), speaker accent (*London Cockney/RP*), listener sex (*male/female*) and listener background (*Northern/Southern*) were all included in the statistical model. The effects of listener and utterance were included as random effects, and the model also included an interaction between accent and average F0.

	χ^2	df	P
Mean F0	39.5	4	**<0.001*** **
Speaker accent	66.18	3	**<0.001*** **
Listener sex	2.12	2	0.35
Listener geographical background	0.02	1	0.89
F0 * Speaker accent	4.94	2	0.08

Table 3.1 – Significance values for the fixed effects in the `lmer` model (Experiment 1)

The model output in Table 3.1 shows a significant main effect of mean F0 on listener threat ratings ($\chi^2(4) = 39.5$, $p<0.001$). This effect is illustrated in Figure 3.3, which plots the fitted model output for listener ratings in each F0 category. Figure 3.3 illustrates that listener threat ratings were lowest in the High F0 category and highest in the Low F0 category, with a comparatively smaller difference between ratings in the Low and Mid mean F0 categories than the difference between ratings in the Mid and High F0 categories. A Holm-Bonferroni corrected Tukey pairwise comparison test was subsequently conducted to assess the differences between threat ratings in the different F0 categories. This testing revealed that while the difference between the Low and Mid F0 categories did not reach statistical significance ($z=1.20$, $p=0.23$), the differences between ratings of the Low and High mean F0 stimuli ($z=5.67$, $p<0.001$) , and between ratings of the High and Mid mean F0 stimuli ($z=4.47$, $p<0.001$) were both significant.

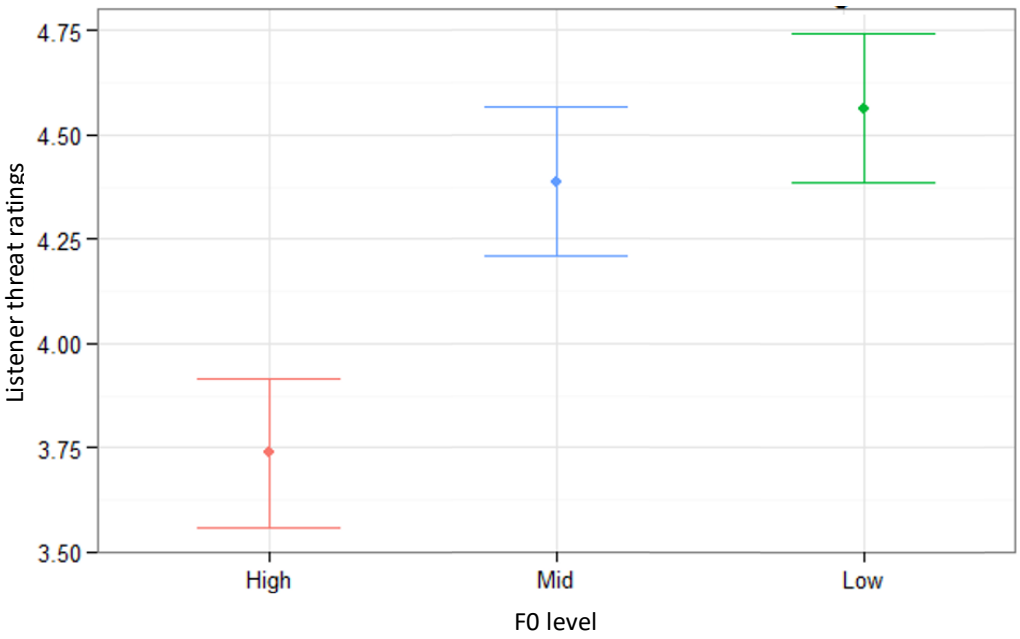


Figure 3.3 – Effect of mean F0 on listener threat ratings (Experiment 1)

In addition to F0, the `lmer` model output shows a statistically significant effect of speaker accent guise on listener ratings of how threatening speakers sounded ($\chi^2(4) = 66.18$, $p<0.001$). This is further displayed in Figure 3.4, which shows that threat rating scores were lower for the RP guise than for the London Cockney guise.

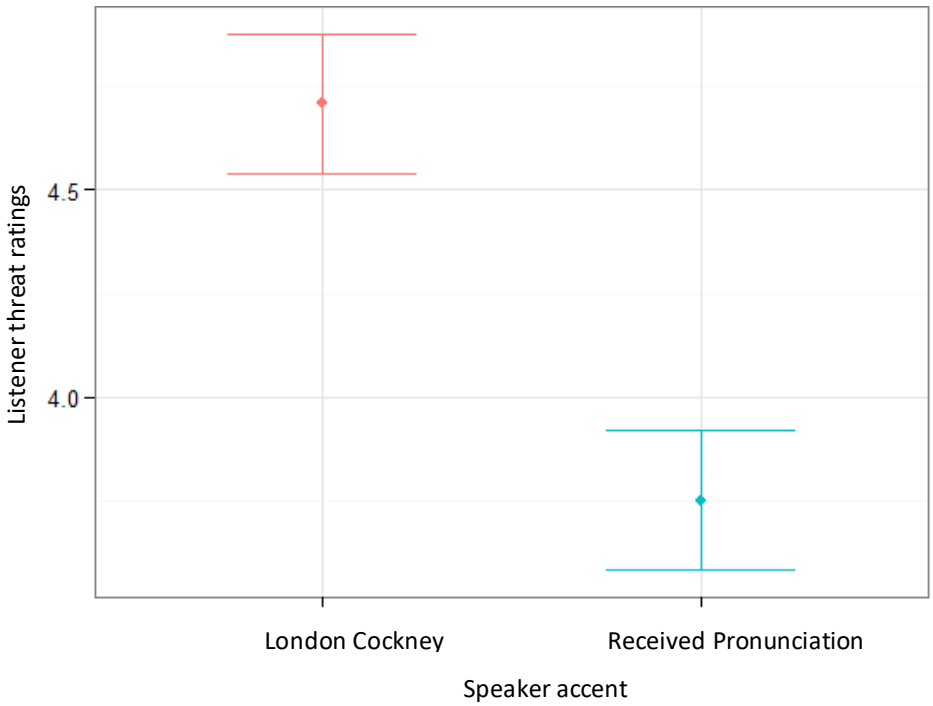


Figure 3.4 – Effect of speaker accent on listener threat ratings (Experiment 1)

The `lmer` model output in Table 3.1 reported that the effect of the interaction between F0 and speaker accent did not reach statistical significance ($\chi^2(2) = 4.94$, $p<0.08$). The relative effect of F0 did not, therefore, differ significantly between each speaker accent group. Figure 3.5 further illustrates the individual effects of speaker accent and mean F0, showing that the RP accent guise was rated as sounding significantly less threatening than the London Cockney accent guise in all three F0 categories.

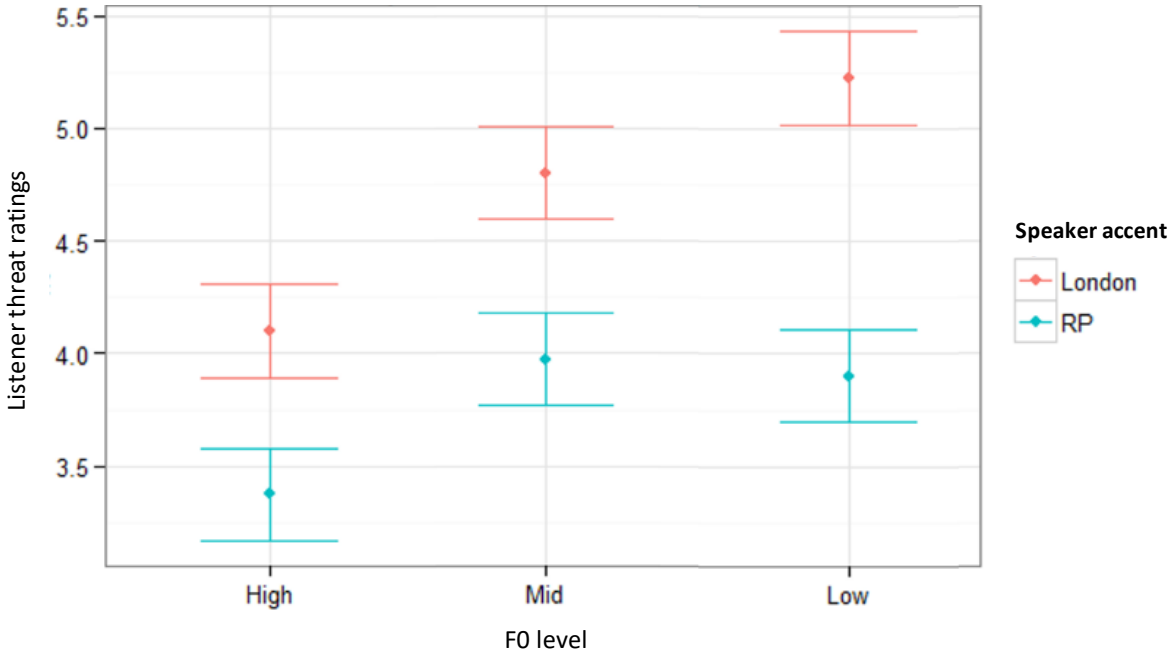


Figure 3.5 – Effect of mean F0 and speaker accent on listener threat ratings (Experiment 1)

The `lmer` output in Table 3.1 shows that there were no statistically significant effects of either listener sex ($\chi^2(2) = 2.12$, $p=0.34$) or listener geographical background ($\chi^2(1) = 0.02$, $p=0.89$) on threat ratings assigned by participants. Listeners from the north of England assigned ratings which did not significantly differ from ratings assigned by listeners from the south of England. Equally, male listeners did not assign significantly different threat ratings to female listeners.

The experimental stimuli used in this experiment included two indirect threat utterances – *"I know where you live"* and *"I wouldn't do that if I were you"*. Figure 3.6 illustrates that the raw data shows that threat ratings were higher for the "I know where you live" utterances than the "I wouldn't do that if I were you" utterances. Figure 3.6 displays boxplots of the threat ratings assigned in each of the utterance categories. This shows

that utterance may also influence listeners' perceptions, but more testing with set hypotheses would be required to develop this idea further.

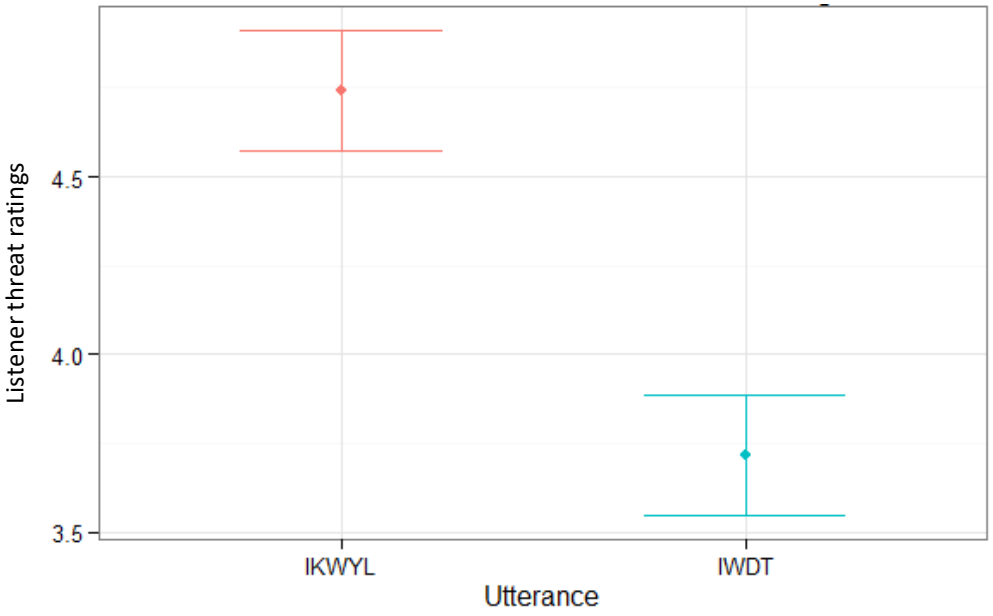


Figure 3.6 – Effect of utterance on listener threat ratings (Experiment 1)

3.3.2.3. Qualitative evaluations of samples

In addition to providing ratings of how threatening the speaker in each recording sounded, listeners were asked to provide evaluations of speakers' body size, with particular reference to height and build. These evaluations were made from vocal information alone as participants were not provided with images of speakers or any other visual cues to speakers' body size. The aim of this analysis was to further explore the previously associated link between F0 and body size perception (see, for example, Ohala, 1984). A breakdown of comments relating to build and height are listed in Tables 3.2 and 3.3 respectively. Following the protocol adopted by Watt and Burns (2012), the three most common descriptors are listed in each table. Instances where more than three descriptors are listed reflect an equal number of listeners providing that

answer. Additionally, where comments were suitably similar (for example, '*large build*' and '*larger build*'), they were collapsed into a single category. The total number of responses for each descriptor is provided in the adjacent brackets.

	RP – Low F0	**RP – Mid F0**	**RP – High F0**
Descriptive labels	Average build (17) Slim (13) Large build (2) Stocky (2)	Average build (21) Slim (9) Slender (2) Small build (2)	Average build (19) Slim (18) Small build (3)
	London – Low F0	**London – Mid F0**	**London – High F0**
Descriptive labels	Large build (14) Average build (9) Stocky (4)	Average build (9) Stocky (7) Slim (7) Large build (5)	Average build (16) Stocky (6) Large build (6)

Table 3.2 – Evaluations of speaker build provided by listeners (Experiment 1)

	RP – Low F0	**RP – Mid F0**	**RP – High F0**
Descriptive labels	Tall (16) 5' 11" (5) 6' (5)	Tall (15) Average (7) 6' (5)	Tall (10) Average (6) 6' (4)
	London – Low F0	**London – Mid F0**	**London – High F0**
Descriptive labels	Tall (11) Average (10) Short (3) 6' 2" (3)	Tall (12) Average (8) 5' 11" (4)	Average (8) 5' 10" (3) Tall (3)

Table 3.3 – Evaluations of speaker height provided by listeners (Experiment 1)

Examining the descriptors in Tables 3.2 and 3.3, it can be argued that a link can be seen between the qualitative results presented in Section 3.3.1 and listeners' evaluations of speakers' body size. Table 3.2 shows that the guise that was rated as sounding the most threatening (London Cockney accent guise with low F0) was also reported as having a '*large build*' more frequently than any other speaker. Equally, the guise rated as sounding the least threatening (RP accent guise with high F0) was reported to be '*slim*' more frequently than any other speaker. The results in Table 3.2 also indicate that speaker accent guise may influence listener comments about speaker build. For example, the London Cockney accent/low F0 speaker is said to have a '*large build*' by 14 listeners, but only 2 listeners rated the RP speaker with the same F0 as having a '*large build*'. The link between height perception and phonetic properties of speech is comparably weaker in this experiment, although the results in Table 3.3 show that guises with lower average F0 were generally rated as sounding taller than guises with higher average F0, but not compared with the mid F0 recordings.

3.2.2.4. Correlations between threat ratings and other traits

In addition to providing evaluations of how threatening speakers sounded, listeners were also instructed to rate for how *angry*, *aggressive*, *menacing*, *agitated*, *friendly* and *intelligent* speakers sounded. Given that prior research has highlighted potential links between anger and threat (Watt et al., 2013), and the link between lowered F0 and increased perception of aggression (Bachorowski, 1999; Ohala, 1984), testing for correlations between threat perception and anger and aggression is merited. Equally, including ratings of friendliness in the experiment allowed for an assessment of whether listeners thought the utterances they rated were produced in the best interests of the hearer or to the hearer's detriment, under the assumption that the former would increase friendliness ratings with the latter decreasing friendliness ratings. Correlations between

listener threat ratings and ratings of the other assessed traits were calculated using Pearson Product-Moment Correlation Coefficients, produced using the `cor.test` function in R, and are presented in Table 3.4. Cohen's (1992) approach is used when assessing the magnitude of these effects, where $r > 0.10$ equates to a small effect size, $r > 0.30$ is the threshold for a medium effect size (classified by Cohen (1992:156) as an effect which would "represent an effect likely to be visible to the naked eye of a careful observer"), and $r > 0.50$ is said to represent a large effect size.

Correlation	***r*=**
Threatening ~ Aggressive	0.80
Threatening ~ Menacing	0.76
Threatening ~ Angry	0.68
Threatening ~ Agitated	0.36
Threatening ~ Intelligent	-0.23
Threatening ~ Friendly	-0.47

Table 3.4 – Correlations between ratings of how threatening speakers sounded and other assessed traits (Experiment 1)

The results in Table 3.4 show that strong positive correlations with a large effect size existed in the data between listener ratings of how threatening speakers sounded and ratings for *aggression*, *anger* and *menace* in the same speakers' voices. A moderate negative correlation existed between listener threat ratings and evaluations of how *friendly* speakers sounded, and a moderate positive correlation was found between judgements of how threatening and *agitated* speakers sounded. Finally, there was a weak negative correlation between ratings for threat and *intelligence*.

3.3 Experiment 2

The study conducted in Experiment 1 examined the relative effects of speaker accent and average F0 on listener threat perception, showing that both factors can influence listener judgements of how threatening a speaker sounds. However, only one speaker was used, producing utterances in two accent guises. Additionally, other potentially-influencing phonetic variables were not factored into the experimental design. The work presented in Experiment 2 further builds on Experiment 1 by examining the relative effects of three phonetic variables – *mean F0*, *speech rate* and *F0 range* – on listener threat evaluations of multiple speakers talking in unfamiliar languages.

3.3.1 Methodology

To obtain the data used for Experiment 2, German and Polish-speaking male and female speakers were recorded producing indirect threats. *"I know where you live"* was recorded in Polish, and *"I wouldn't do that if I were you"* was recorded in German. Given that the listeners used for this perception experiment (see below) had no prior knowledge of German or Polish, it was anticipated that the use of different utterances would not impact negatively on the outcome of the experiment.

A Praat script (Antoniou, 2010) was initially used to alter the mean intensity level of all recordings to 70dB. The recordings were then duplicated and altered to create contrasts for both F0 and speech rate. The same Praat pitch alteration script as used in Experiment 1 (Fecher, 2015) was used to create two contrasting F0 levels for this experiment. For male speakers, the mean F0 of each recording was altered to 90Hz (low) and 140Hz (high), using the same rationale as the alteration used in Experiment 1. For female speakers, the mean F0 of each recording was altered to 170Hz (low) and 250Hz (high).

These values are 40Hz above and below an approximation of an average female F0 level, and reflect the low and high ends of the mean F0 range reported for female speakers (Künzel, 1989; Traunmüller and Erickson, 1995). Following F0 alteration, the tempo of each recording was normalised to an articulation rate of 5 syllables per second, and subsequently tempo-altered ±20% using Audacity software to create slow (-20%) and fast (+20%) speech rate versions of each stimulus. Performing the speech rate alterations in this way allows for tempo to be altered independently of average F0. Once all alterations had been made, each recording was re-checked to ensure that the F0 and speech rate were at the desired levels. As in Experiment 1, all recordings were checked to ensure that no digital artefacts had influenced the sound quality as a result of the editing process. In addition to average F0 and speech rate, F0 range was also considered as a potentially influencing variable, and was taken to represent a measure of how monotonous speakers sounded. F0 range was treated as a continuous variable in this study, calculated as the difference between the maximum and minimum F0 values in each stimuli once measurement errors in Praat had been discarded.

The recording and alteration procedures provided a **2** (slow/fast speech rate) x **2** (low/high F0) experimental design for voice samples within **four** (German male/German female/Polish male/Polish female) speakers. As in Experiment 1, recordings were band-pass filtered between 300 and 3400Hz to provide an approximate replication of the telephone channel frequency band (Künzel, 2001; Nolan, McDougall and Hudson, 2013).

For the perception experiment, 42 British English listeners (age range 18-65, mean age 25, SD 9.4; 33 female) completed an online survey where they were asked to evaluate how threatening they thought the speaker sounded in each recording using a seven-point

Likert-type scale (1= “Not at all threatening”, 7= “Very threatening”). Ten foil recordings were included in the experiment, interspersed between the target stimuli. Information was also collected on whether listeners had any background with foreign languages. To ensure that the verbal channel remained uninterpretable to listeners, any listener who stated they had any prior experience of German or Polish was removed from the sample beforehand, even if this experience only included basic learning at school.

Comparable statistical analysis to the process used in Experiment 1 was used to analyse the data in Experiment 2. Mean F0 (*high/low*), speech rate (*fast/slow*), F0 range (*continuous*), speaker sex (*male/female*), listener sex (*male/female*) and speaker language (*German/Polish*) were all included in the model. Given that the experiment contained multiple speakers and multiple listeners, both speaker and listener were included as random effects, along with an interaction between speech rate and F0 and an interaction between listener sex and speaker sex. This was included to assess whether male and female listeners would evaluate male and female speakers differently.

3.3.2. Results

3.3.2.1. Effects of phonetic variables on listener threat evaluations

Table 3.5 shows significance values calculated for each of the fixed effects on listener perceptions of how threatening speakers sounded from the `lmer` model outlined previously.

	Chisq	Df	Pr(>Chisq) =
Mean F0	70.50	2	**<0.001*** **
Speech rate	12.37	2	**<0.001*** **
F0 range	3.22	1	0.07
Listener sex	1.71	2	0.42
Speaker sex	11.26	2	**0.004** **
Speaker language	11.79	1	**<0.001*** **
F0 * Speech rate	11.48	1	**<0.001*** **
Listener sex * speaker sex	1.50	1	0.22

Table 3.5 – Fixed effects on listener threat ratings (Experiment 2)

The model output in Table 3.5 shows several significant fixed effects on listener perceptions of how threatening speakers sounded. Table 3.5 shows that the difference between listener threat ratings assigned to the low and high mean F0 stimuli was statistically significant ($\chi^2(2) = 70.50$, $p<0.001$). This difference is further illustrated in Figure 3.7, which plots the effect of mean F0 from the `lmer` model, and illustrates that stimuli in the low F0 category were rated as sounding more threatening than those in the high F0 category.

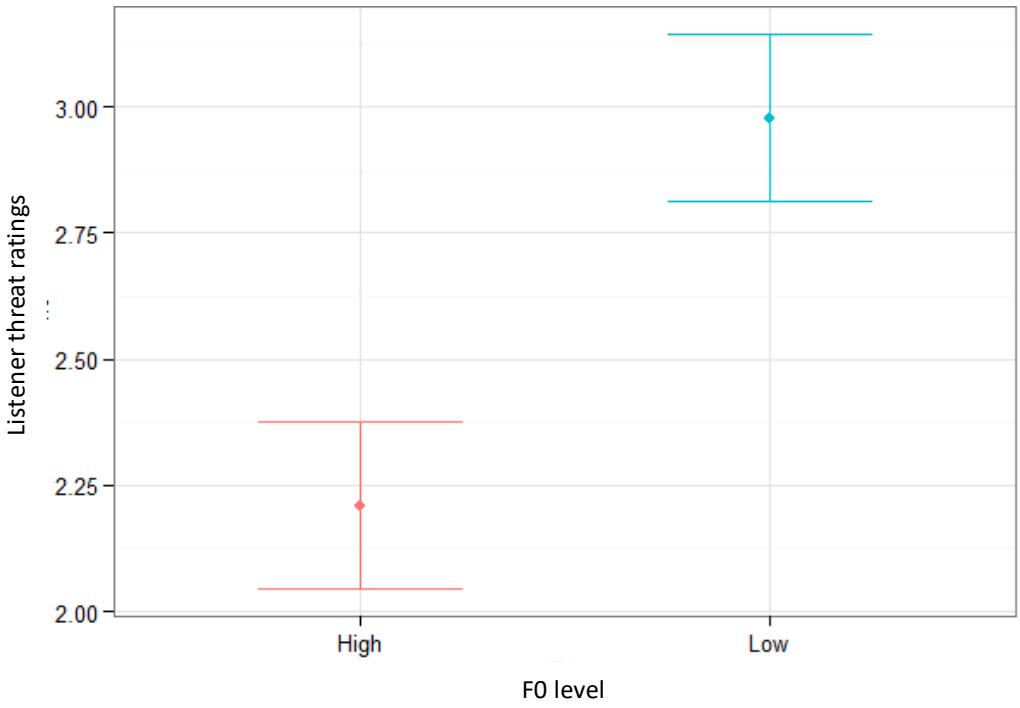


Figure 3.7 – Effect of mean F0 on listener threat ratings (Experiment 2)

The output in Table 3.5 also shows significant effects of speech rate ($\chi^2(2) = 12.37$, $p<0.001$) and the interaction between F0 and speech rate ($\chi^2(1) = 11.48$, $p<0.001$) on listener perceptions of how threatening speakers sounded. This is further illustrated by the model output plot in Figure 3.8, which shows that stimuli in the slow speech rate category were rated as sounding more threatening than stimuli in the fast speech rate category, but that this effect was stronger for low F0 recordings compared with high F0 recordings. In the high F0 category, the fast speech rate stimuli were rated as sounding more threatening, whereas the opposite was true for the low F0 recordings.

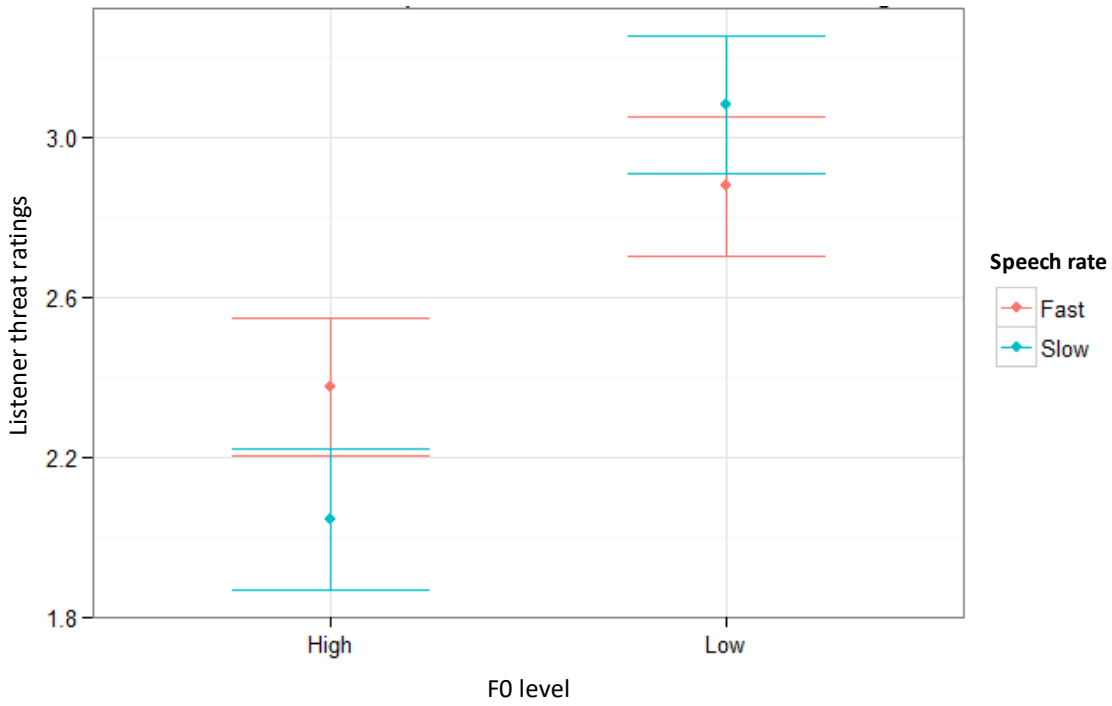


Figure 3.8 – Effect of speech rate and F0 on listener threat ratings (Experiment 2)

Table 3.5 illustrates that both speaker language (χ^2 *(2) = 11.79, p<0.001*) and speaker sex (χ^2*(2) = 11.26, p=0.004*) had significant effects on listener threat ratings. The plots in Figures 3.9 and 3.10 show that the German speakers were rated as sounding less threatening than the Polish speakers, while the two male speakers were rated as sounding significantly more threatening than the two female speakers in the sample.

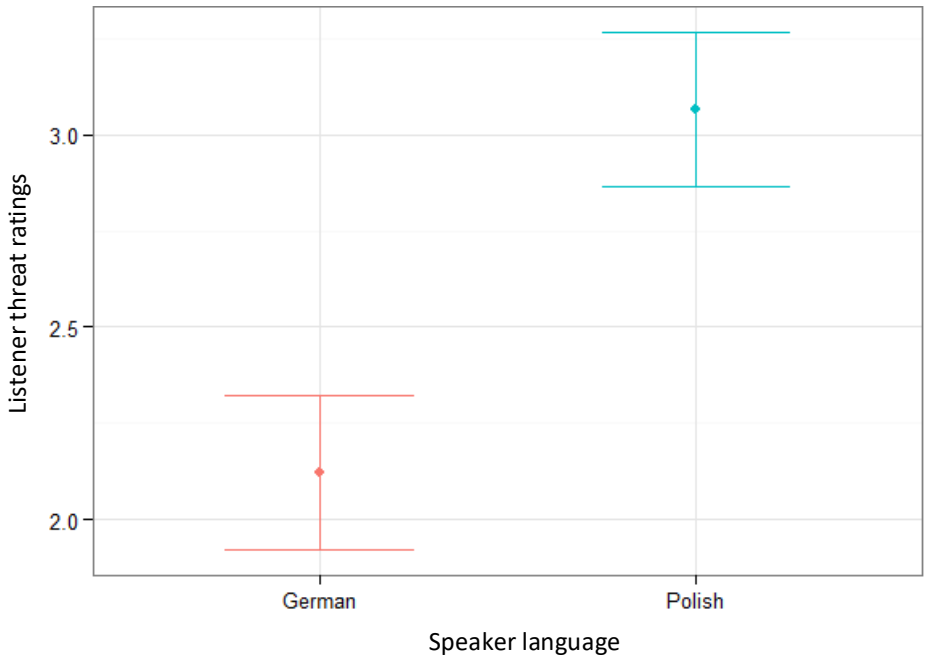


Figure 3.9 – Effect of speaker language on listener threat ratings (Experiment 2)

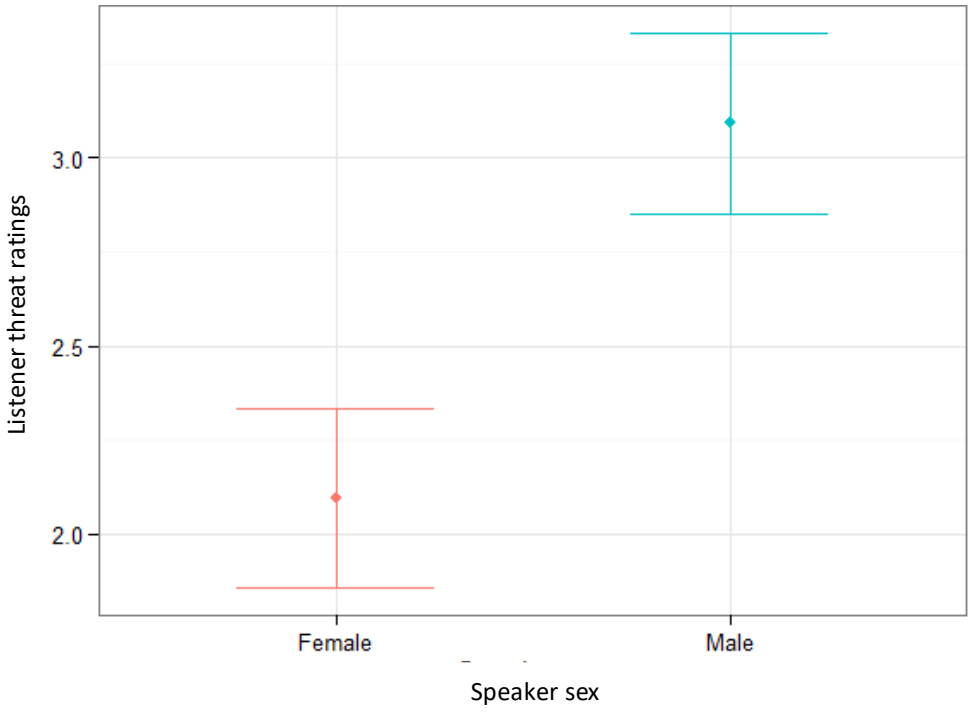


Figure 3.10 – Effect of speaker sex on listener threat ratings (Experiment 2)

The model output in Table 3.5 also reported non-significant effects on listener perceptions of how threatening speakers sounded. There was no significant difference between threat ratings assigned by male and female listeners ($\chi^2(2) = 1.71$, $p=0.42$), or

in the interaction between listener sex and speaker sex ($\chi^2(2) = 1.50$, $p=0.22$). Additionally, the effect of F0 range did not reach statistical significance ($\chi^2(1) = 3.22$, $p=0.07$) in the `lmer` model.

3.4. Discussion

The primary goal of the work presented in this chapter was to provide an initial examination of whether aspects of the vocal channel could influence listeners' perceptions of how threatening speakers sounded.

Experiment 1 explored the effects of mean F0 and speaker accent on listeners' perceptions of two indirect spoken threats. Results showed a significant effect of mean F0 on listener perceptions of how threatening the speaker sounded, with lower F0 recordings rated as sounding the most threatening and high F0 recordings sounding the least threatening. This is consistent with previous literature which predicts that lowered F0 is correlated with listener perceptions of a more dominant and aggressive speaker (Chuenwattanapranithi et al., 2009; Ohala, 1984; Xu et al., 2013). This claim would be further supported by the strong correlation (r=0.8) between listener ratings of how threatening and aggressive speakers sounded in the experiment.

Additionally, the non-standard London Cockney accent guise was rated as sounding more threatening than the standard RP accent guise. This result is consistent with previous research which found that standard accents were perceived more positively than non-standard accents in various legal settings (Dixon, Mahoney and Cocks, 2002; Dixon and Mahoney, 2004). The results from Experiment 1 also showed that the strongest accent effect occurred in the Low F0 category and the strongest F0 effect

occurred in the London Cockney accent guise category. This could suggest that these two individual variables can combine, to some degree, to enhance the overall threat rating provided by listeners, although the statistical testing showed a non-significant interaction between the two variables. The results also support the idea that supra-segmental aspects of speech can combine with segmental phonetic aspects to influence perceptions of how threatening a speaker sounds, lending support to the view that factors associated with both social and acoustic phonetic variation in the vocal channel are worthy of further exploration in work examining how listeners perceive threat in speakers' voices.

Additionally, Experiment 1 also obtained evaluations of speaker height and build from participants, in order to further examine the potential association between low frequency vocalisations, listeners' perceptions of traits such as dominance, aggression and threat, and the perception of a physically larger speaker. The qualitative evaluations of speaker size provided by listeners did, to some extent, validate the hypothesis that speakers who were perceived as being physically larger would also be evaluated as sounding more threatening than speakers who were perceived to be physically smaller. It appeared that both F0 and speaker accent guise influenced perceptions of body size in relation to speaker build in Experiment 1, with the low F0 stimuli in the London Cockney accent guise most commonly associated with perceptions of a larger speaker. This was also the accent guise/F0 combination that received the highest threat rating. While a link between F0 and body size is predicted by previous research (Chuenwattanapranithi et al., 2009; Ohala, 1984; Xu et al., 2013), this has not yet been expanded to include the influence of accent stereotyping or a potential cumulative effect of combining both segmental and prosodic features. The results from Experiment 1 could suggest that listeners may use a more holistic evaluation of how threatening someone sounds when

describing what they think a speaker looks like, rather than solely focussing on aspects such as F0. Furthermore, although only a small number of comments were provided by listeners in Section 3.3.2, the work highlights that eliciting and analysing qualitative evaluations alongside results from quantitative judgement-based tasks could help to provide further useful insights into how listeners evaluate voices.

Although not tested in a formal or structured way, the results from Experiment 1 highlighted the possibility that the individual indirect utterance may also influence listeners' perceptions of how threatening someone sounds. This links to the idea that even within the 'indirect' category, two utterances would likely be perceived differently to one another. However, this would require further development and structured testing beyond that provided by Experiment 1 in order to reach firmer conclusions as to the influence of individual utterances.

Experiment 2 aimed to further examine links between vocal attributes and threat perception by minimising the effect of the verbal channel and presenting stimuli to listeners in unfamiliar foreign languages. This followed the design of Watt, Kelly and Llamas (2013), and aimed to build on their work by examining how manipulations of specific phonetic features could influence listener ratings of how threatening a speaker sounded. In Experiment 2, statistically significant results were found for the effects of F0, speech rate, and the interaction between F0 and speech rate, on listener ratings of how threatening speakers sounded.

Watt, Kelly and Llamas (2013:114) found no observable difference between induced-threat and neutral utterances that were played to listeners in unfamiliar foreign languages. However, the results in Experiment 1 indicate that listeners did assign

greater or lesser levels of threat when specific phonetic parameters were altered in the speech signal, even when the signal contained words in unfamiliar languages. This highlights the possibility that there are certain phonetic parameters that British English listeners interpret as influencing perceptions of how threatening a speaker sounds, even in the absence of an interpretable verbal channel.

Two additional significant effects were found in the results of Experiment 2. Male speakers were rated as sounding significantly more threatening than female speakers, and ratings of how threatening the two Polish speakers sounded were significantly higher than ratings of how threatening the German speakers sounded. This language effect highlights the possibility that listeners recognised the languages, or believed they recognised the languages, without understanding the words being spoken. It can be therefore argued that perceptions of speakers' language backgrounds in unfamiliar languages can influence perceptual judgements. This study did not, however, ask listeners to identify the language they thought they had heard, and it is acknowledged that this limits the scope for any further analysis of this effect.

The work presented in this chapter can be seen as an initial attempt to examine the idea that aspects of the vocal channel such as pitch, speech rate and speaker accent can cause listeners to perceive greater or lesser levels of threat in a speaker's voice. Although the results are derived from two small experiments, the findings indicate that all three variables could contribute to listener perceptions of what Watt, Kelly and Llamas (2013) describe as a 'threatening tone of voice'.

However, the two experiments do have several limitations. Given that Experiment 1 used one speaker under a matched guise design which aimed to model two accents, and

that Experiment 2 only used four speakers, the speaker-dependency of the results can be questioned. It is acknowledged that finding systematic patterns for either a single speaker or a small amount of speakers does not equate to systematic patterns across a population of listeners. It is also acknowledged that in context-independent scenario, analysis of linguistic effects on threat perception can only primarily reveal information about relative threat perception, rather than threat perception based on prior contextual information, which would arguably reflect real-world threat assessment situations more accurately.

However, given the lack of previous research on the link between aspects of the vocal channel and listener threat perception, the work presented in this chapter does begin to address the phonetic basis for the perceptual existence of a 'threatening tone of voice' through controlled experimental analysis. The goal of the work presented in subsequent chapters of this thesis is to examine effects with a wider array of utterances and phonetic parameters, both in and out of certain forensically-relevant contexts (e.g. the effect of hearing a range of 'threatening' utterances within the context of a bomb threat to an emergency service operator). This, it is hoped, will address some of the shortcomings of the analysis derived from these two initial experiments.

Chapter 4 – Incorporating voice quality and listener descriptions

4.1. Introduction

The experiment presented in this chapter aims to build on the research presented in Chapter 3 by further examining the effects of laryngeal and temporal properties of voice on listener threat perception. Section 3.4 of this thesis highlighted that the research presented in Chapter 3 could be further expanded by testing the effect of voice quality on listener evaluations of how threatening speakers sounded. It also stated that incorporating and testing listener evaluations within the bounds of contextual information, rather than in a context-independent scenario, would be a worthwhile pursuit for the research presented in this thesis. These aspects are addressed by the research presented in this chapter. In doing so, this chapter aims to advance the work which addresses Watt, Kelly and Llamas' (2013) broader assertion that a lack of research exists investigating how phonetic aspects of speech can influence how people perceived threat from aspects of the vocal channel.

4.2. Methodology

4.2.1. Materials

The stimuli used for the experiment presented in this chapter consisted of adapted versions of demonstration recordings produced by Hartwig Eckert in a standard North German accent for a reference work on voice quality types (Eckert and Laver, 1994). The experimental stimuli were comprised of the same utterance produced using five contrasting voice qualities – *creak*, *falsetto*, *harsh*, *modal* and *whispery*. These were chosen as they span the frequency range adopted by male speakers (Laver, 1980) and have labels which are arguably more intuitive to lay-listeners than the labels for

qualities such as laryngealized or velarized (Watt and Burns, 2012). The utterance produced by the speaker was *"Beim Fußball können die Sportfreunde immer davon ausgehen, dass die schönsten Tore und die interessantesten Spielzüge abends um zehn Uhr dreißig in der Sportschau übertragen werden"*, which translates as *"With football, sports fans can always count on the finest goals and most interesting play being covered in the sports show at ten-thirty in the evening"* (translation provided by Watt and Burns, 2012). Watt and Burns (2012) highlight that not all the recordings provided by Eckert and Laver (1994) contain labels, making it difficult for subsequent researchers to identify and use the recordings appropriately. Watt and Burns (2012) used the Vocal Profile Analysis scheme (MacKenzie Beck, 2005) to label each of the voice qualities provided by Eckert and Laver (1994), and these were independently verified by an independent phonetician with a high level of expertise in voice quality analysis (Watt, personal communication, 2nd June 2017). In order to maintain consistency with previous research, the descriptive labels used by Watt and Burns (2012) are adopted in this experiment.

The voice quality stimuli were altered to create additional contrasts for both pitch and speech rate. This procedure was conducted using Audacity software. The 'change pitch' function was used to alter the pitch of each sound file by ±10% to create higher-pitched (+10%) and lower-pitched (-10%) pitched stimuli for each voice quality. While altering the stimuli using this function changes the pitch independently of altering tempo, formant frequencies beyond F0 are also affected and shifted by the same magnitude as F0. Hence, 'pitch' will be used to refer to the variable rather than F0. This process was used instead of the F0 alteration Praat script used in Chapter 3 (Fecher, 2015) as it did not produce audibly distorted samples with sound quality compromised by digital artefacts for the non-modal voice quality recordings in the experiment. Following pitch

alteration, each recording was subsequently tempo-altered ±10% using Audacity to create slower (-10%) and faster (+10%) speech rate versions of each stimuli. Once all alterations had been made, each recording was re-analysed to ensure that the pitch and speech rate of each recording were at the desired levels. All recordings were also checked to ensure that no digital artefacts had influenced the sound quality as a result of the editing process. The alteration procedures provided a **5** (voice quality) x **2** (slower/faster speech rate) x **2** (lower/higher pitch) experimental design. Each recording was band-pass filtered between 300-3400Hz to provide an approximate replication of the telephone channel frequency band (Künzel, 2001; Nolan et al., 2013). This was done in order to more accurately replicate the context in which stimuli were played to participants in the experiment (see Section 4.2.3). It also mirrors the work presented in Chapter 3.

4.2.2. Participants

A total of 80 participants (mean age = 19.6, age range = 18-32) provided informed consent to take part in a listening task in which they were required to answer a series of questions about the auditory stimuli outlined in Section 4.2.1. All participants were recruited from the student population at the University of York and received either payment or course credit in exchange for their participation. No participant reported any hearing impairments and all participants had normal or corrected-to-normal vision. All participants were tested in person in the Department of Psychology at the University of York. As the vocal stimuli were in German, listeners were asked to state whether they spoke German or had any past experience with the language. Only participants who stated that they did not speak German and had either limited or no prior exposure to the language were included in the final sample. No participants had received any advanced-level formal training in phonetics, and this type of training was not provided before the

experiment took place. Watt and Burns (2012:3) argue that providing listeners with training before their participation in experiments similar to this study detracts from the forensic realism that such work aims to replicate.

4.2.3. Experiment design

The edited vocal stimuli were embedded within a Qualtrics survey platform but the experiment took place within a lab setting. For each vocal stimulus, listeners were instructed to listen to the recording using closed-cup headphones and then answer a series of questions relating to the voice they heard. So as to avoid placing too heavy a demand on the listeners, each participant heard half of the total number of recordings (n=10). To control for the effect of order as much as possible, stimuli were presented in a computer-generated randomised order to each of the participants, and this order was not known to the researcher until the experiment had been completed.

One advantage of using foreign language stimuli in this experiment is that it allowed for analysis of how contextual evidence influences listener evaluations independently of an interpretable verbal channel. Given that contextual evidence was not accounted for in the experiments presented in Chapter 3, accounting for its influence in lab-based perception experiments was one goal of the research presented in this chapter. To facilitate this, participants were pre-assigned to one of two context groups. Group 1 were given prior instructions that the recordings they would hear were bomb threats received by German emergency service operators. Group 2 were given no contextual information about the origin of the recordings, other than to be made aware that listeners can be asked to provide information about unknown speakers' voices in forensic contexts. This was done to ground both experiments in some degree of forensic realism, but with the intention of only having one group who would explicitly associate

the recordings they heard with threats. 40 listeners were assigned to each group. Given the potential problems surrounding language perception highlighted in Chapter 3, listeners in both groups were instructed that the stimuli they would hear would be in German at the beginning of the experiment. No foil voices were included in this experiment. This was designed to remove the potential for listeners to believe that the samples were taken from a language other than German, or that they would be exposed to more than one language in the experiment. Given the phonetic similarity between the German *"Fußball"* and English *"football"*, listeners in Group 1 were informed that the bomb threat utterance they heard stated that a bomb would go off at a local football stadium. All listeners were provided with full information about the true nature of the recordings following completion of the experiment and were given the option to withdraw their data following this debrief. However, all participants consented for their data to be included in the results following the debrief. Details of the full contextual information and instructions provided to participants are listed in Table 4.1. The emboldening of text in Table 4.1 reflects the emboldening of text on the screen presented to listeners in the experiment.

Group 1 - Contextual information and experimental instructions	Group 2 - Contextual information and experimental instructions
In forensic contexts, listeners are often asked to provide information about an unknown speaker's voice. These evaluations frequently occur when bomb threats are received by companies or emergency service operators. In this experiment, you will hear a series of phone calls which are replications of **bomb threats** received by the **German** emergency services. The words in each call are the same, produced by different speakers. **In each case, the caller states that there is a bomb that will be detonated at a local football stadium unless certain demands are met**. You will hear a series of these calls, which have been designed for training purposes, and in each case will be asked to describe the voice you hear by ticking appropriate boxes. You will also be asked to describe what you think the speaker in each call looks like and give an impression of how intelligent, threatening and friendly you think the speaker sounds.	In forensic contexts, listeners are often asked to provide information about an unknown speaker's voice. In this experiment, you will hear a series of phone calls produced by different speakers. The words in each call are the same. In each case will be asked to describe the voice you hear by ticking appropriate boxes. You will also be asked to describe what you think the speaker in each call looks like and give an impression of how intelligent, threatening and friendly you think the speaker sounds. If you have any questions please ask the researcher.

Table 4.1 – Contextual information provided to Group 1 and Group 2

The experiment was designed to elicit information in a comparable way to the current bomb threat checklist document provided by the UK National Counter Terrorism Security Office (hereafter NCTSO), details of which are presented in Section 2.4. Participants were instructed to detail whether they thought the caller was male or female, along with the caller's age and any noticeable qualities of the caller's voice. The bomb threat checklist provides users with the following options to describe a speaker's voice; *calm, crying, clearing throat, angry, nasal, slurred, excited, stutter, disguised, slow, lisp, accent (if selected, which accent), rapid, deep, familiar, laughter, hoarse, other (please state)*. All of these options were included, with the exception of *accent*,

which was removed given that listeners were explicitly told that the samples were spoken in German. In addition to the options listed on the NCTSO bomb threat checklist, *whispery*, *creaky* and *harsh* were included to assess how accurately listeners would describe the voice qualities they were exposed to. *High-pitched* was also included to provide an opposite to the already-included *deep* and as a more natural classification for the falsetto voice quality stimuli. Listeners were instructed to select all the descriptors that they felt applied to each speaker's voice, with no upper or lower limit placed on how many descriptors could be assigned in each case.

In addition to describing the speakers' voices, participants were instructed to provide descriptions of what they thought the speaker in each recording looked like. As listeners were presented with no visual stimuli or information about the physical characteristics of the speakers, they were required to infer visual information about the speaker from his voice. The work presented in Chapter 3 highlighted potential problems with eliciting free descriptions from listeners about the physical characteristics of speakers. This was particularly problematic with respect to grouping similar but non-identical descriptions, as this process required inference on the part of the researcher. To avoid this issue and to ensure that results were comparable across the sample of listeners, tick-boxes were used to elicit physical descriptions of the speakers. This also mirrored the approach taken for the evaluation of vocal traits. Listeners were presented with the following selection options; *shorter than average*, *taller than average*, *very short*, *very tall*, *average height*, *average build*, *large build*, *small build*, *slim*, *stocky*. Again, participants were permitted to select as many or as few descriptors as they felt appropriate, with no upper or lower limit placed on how many physical attributes could be assigned. As was the case in Chapter 3, the voice was the only source of information available for listeners to assess speakers' physical characteristics.

Alongside describing each voice and the perceived physical characteristics of the speaker, listeners were instructed to rate each voice for how *intelligent*, *threatening* and *friendly* they thought the speaker sounded using a seven-point Likert-type scale. This was included to facilitate an assessment of how the various vocal parameters investigated in the experiment would influence listener judgements of how *threatening* a speaker sounded. The inclusion of *intelligence* and *friendliness* was designed to focus listeners' attention away from only considering threat ratings. This was particularly appropriate for the listeners in Group 2 as they were provided with no background context about the source of the voices.

4.4. Results

4.3.1. Listener descriptions of voices

One aim of the current experiment was to assess how listeners with no prior phonetic training attribute specific speaker traits to voices they hear. The analysis in this section focuses on the responses provided by listeners to each of the voices in the experiment. The total number of 'yes' labels provided by participants for each of the vocal traits from the adapted NCTSO checklist are listed in Table 4.2, split in accordance with the two experimental context groups that listeners were assigned to.

	Number of 'Yes' responses				
	Bomb threat context	**Non-threat context**		**Bomb threat context**	**Non-threat context**
Angry	75	72	**Hoarse**	89	86
Calm	166	148	**Laughter**	0	1
Clearing throat	15	11	**Lisp**	7	1
Creaky	62	82	**Nasal**	22	21
Crying	4	0	**Rapid**	41	42
Deep	139	139	**Slow**	120	95
Disguised	68	68	**Slurred**	15	10
Excited	20	20	**Stutter**	0	1
Harsh	83	81	**Whispery**	69	63
High-pitched	65	70			

Table 4.2 - Number of 'yes' responses for each checklist trait

The results in Table 4.2 show that minimal differences existed between the two context groups with respect to the number of traits chosen by listeners. While some differences did exist, for example a higher number of 'slow' descriptors used by the bomb threat context group, the application of descriptors was relatively consistent across the sample as a whole, especially taking into account the potential for between-listener differences. There was also no consistent direction of the differences within the responses. Furthermore, descriptors such as *angry* and *harsh*, which may be more likely to be associated with threat perception, are used almost equally across the two groups. These minimal differences are perhaps unsurprising given that listeners were only required to describe the voice they had heard for this part of the study, and the voices were identical in each experimental context group. Based on this, descriptions from listeners in the two

experimental contexts are grouped together through the remainder of this section (4.3.1).

Tables 4.3, 4.4, 4.5, 4.6 and 4.7 show the five most popular listener attributions for each of the acoustic stimuli presented in the experiment. The total number of listeners who heard each sample is also presented, and the number of times each descriptor was selected is provided in the brackets next to each label. Each table also details the percentage of listeners who identified the voices as belonging to male or female speakers, along with details about the perceived age of the talker in each of the vocal stimuli.

4.3.1.1. Creaky voice

	Creak			
	Faster speed + Lower pitch	**Faster speed + Higher pitch**	**Slower speed + Lower pitch**	**Slower speed + Higher pitch**
Descriptors assigned	*Deep (35)* *Hoarse (13)* *Creaky (12)* *Calm (12)* *Slow (12)*	*Deep (29)* *Creaky (16)* *Hoarse (13)* *Calm (13)* *Slow (9)*	*Deep (28)* *Slow (25)* *Creaky (20)* *Hoarse (17)* *Calm (14)*	*Deep (29)* *Slow (21)* *Creaky (18)* *Hoarse (18)* *Calm (17)*
Perceived sex	*Male: **97%*** *Female: **3%***	*Male: **100%*** *Female: **0%***	*Male: **92%*** *Female: **8%***	*Male: **98%*** *Female: **2%***
Perceived age	*Mean age: **43*** *Range: **28-80*** *Std. Dev.: **13***	*Mean age: **45*** *Range: **20-70*** *Std. Dev.: **14***	*Mean age: **44*** *Range: **25-70*** *Std. Dev.: **11***	*Mean age: **46*** *Range: **17-80*** *Std. Dev.: **14***
Total no. of listeners	**39**	**38**	**38**	**41**

Table 4.3 – Descriptors assigned to *creak* stimuli by listeners

The results in Table 4.3 demonstrate that listeners identified the creak samples as being *deep* more frequently and consistently than they labelled the voices as sounding *creaky*. Across the four conditions, 78% of listeners labelled the creak stimuli as sounding *deep*, whereas 42% of listeners opted to use the *creaky* label to describe the stimuli. However, *creaky* appears in the top five identified traits in each of the four conditions, suggesting that the label is, to some extent, used correctly by the listener group. Analysis of the *slow* labels illustrates that while all samples had *slow* in the top five identified traits, the label was used more frequently by listeners when hearing the slower samples in comparison to the faster samples. This suggests that the listener group perceived relative speech rate with some degree of accuracy, even if they generally considered both speeds to be slow. The perception that the speaker in the creaky stimuli sounded *calm* was also common enough in the group to ensure the label appeared in the top five attributed traits in each condition.

Analysis of the perceived sex section of Table 4.3 shows near-categorical male identification rates for the creaky voice samples. This is perhaps unsurprising given the association between low F0 and the production of creaky voice in male speakers. Although creaky voice has been noted as a sociophonetic property of female speech (see, for example, Yuasa, 2010), the results of this study would suggest that the use of persistent creak over an entire utterance does not affect listeners' ability to correctly determine the sex of a speaker. With respect to age perception, although the four creaky stimuli were assigned a similar mean age, there was an extremely high range of age estimations provided by participants. Age attributions for the creak stimuli ranged from a speaker in his or her late teens to an elderly speaker as old as 80, with the standard deviation of age estimations ranging from 11 to 14 years.

4.3.1.2. Falsetto voice

	Falsetto			
	Faster speed + Lower pitch	**Faster speed + Higher pitch**	**Slower speed + Lower pitch**	**Slower speed + Higher pitch**
Descriptors assigned	*High-pitched (26)* *Calm (15)* *Disguised (11)* *Rapid (8)* *Slow (5)*	*High-pitched (28)* *Calm (21)* *Excited (11)* *Rapid (10)* *Nasal (3)* *Disguised (3)*	*High-pitched (30)* *Disguised (16)* *Calm (12)* *Slow (10)* *Creaky (6)*	*High-pitched (26)* *Calm (22)* *Slow (16)* *Disguised (5)* *Creaky (3)* *Hoarse (3)* *Rapid (3)*
Perceived sex	*Male: **78%*** *Female: **22%***	*Male: **5%*** *Female: **95%***	*Male: **68%*** *Female: **32%***	*Male: **3%*** *Female: **97%***
Perceived age	*Mean age: **30*** *Range:**12-65*** *Std. Dev.: **14***	*Mean age: **28*** *Range: **10-70*** *Std. Dev.: **13***	*Mean age: **36*** *Range: **15-70*** *Std. Dev.: **15***	*Mean age: **31*** *Range: **8-80*** *Std. Dev.: **17***
Total no. of listeners	**40**	**38**	**41**	**38**

Table 4.4 – Descriptors assigned to *falsetto* stimuli by listeners

For the falsetto recordings, the descriptive label *high-pitched* was consistently used across all four conditions, with 70% of listeners associating the falsetto recordings with higher pitch. As was the case for the creak stimuli, the *slow* descriptor was used more frequently for the slower speech rate stimuli than the faster stimuli. The use of the *rapid* descriptor was also more common in the faster samples than the slower samples. These results further illustrate that listeners appear to perceive relative speech rate with some degree of accuracy. The *calm* label appears in the top five traits for each of the four conditions, as was the case for the creak samples. Given that the pitch range of male falsetto overlaps with the pitch range for female talkers (Hollien and Michel, 1968:602),

the presence of the *disguised* label across the conditions could suggest that listeners perceived a male speaker attempting to disguise his voice to sound like a female speaker. As the *disguised* label was predominantly used for the two falsetto examples that were identified more frequently as belonging to male speakers than female speakers, it could suggest that listeners thought that the vocal stimuli represented unsuccessful attempts by male speakers to disguise their sex as female. Further analysis of the perceived sex section of Table 4.4 shows that the pitch manipulation in the higher-pitched falsetto samples appeared to sufficiently fool the listener group into incorrectly identifying the speaker as female. This is evidenced by the near-categorical female identification rates for the two higher-pitched falsetto samples, coupled with the small number of listeners who suspected vocal disguise in these two samples. The mean age estimations across the falsetto samples were lower than those provided for the creaky stimuli. However, the range of age estimations was again widely varied, with some listeners attributing the stimuli to the voice of a child and others attributing the stimuli to the voice of an elderly speaker.

4.3.1.3. Harsh voice

	Harsh			
	Faster speed + Lower pitch	**Faster speed + Higher pitch**	**Slower speed + Lower pitch**	**Slower speed + Higher pitch**
Descriptors assigned	*Angry (29)* *Deep (27)* *Harsh (23)* *Hoarse (14)* *Rapid (10)*	*Angry (27)* *Harsh (23)* *Hoarse (16)* *Creaky (11)* *Harsh (11)*	*Deep (34)* *Angry (21)* *Harsh (19)* *Hoarse (18)* *Creaky (11)*	*Angry (27)* *Harsh (25)* *Creaky (17)* *Deep (16)* *Hoarse (16)*
Perceived sex	*Male: **97%*** *Female: **3%***	*Male: **95%*** *Female: **5%***	*Male: **95%*** *Female: **5%***	*Male: **100%*** *Female: **0%***
Perceived age	*Mean age: **49*** *Range: **30-62*** *Std. Dev.: **9***	*Mean age: **51*** *Range: **30-75*** *Std. Dev.: **11***	*Mean age: **49*** *Range: **30-75*** *Std. Dev.: **13***	*Mean age: **54*** *Range: **30-80*** *Std. Dev.: **13***
Total no. of listeners	**39**	**37**	**40**	**40**

Table 4.5 – Descriptors assigned to *harsh* stimuli by listeners

For the harsh stimuli, the *angry* descriptor was more commonly used than any other label, and was chosen by 67% of listeners across the four conditions. Although this is consistent with Laver's (1994:420) assertion that harsh voice and anger are linked, there is a potential problem with the use of this descriptor as it could imply that the speaker *was* angry as opposed to *sounding as though* they were angry. In this case, an assumption would be made about a speaker's mental or emotional state as opposed to a descriptive judgement about their voice. *Harsh* appears in the top five attributed traits in each of the conditions and was used by 58% of listeners, with *hoarse* and *creaky* also being regularly used. In contrast to the falsetto and creak stimuli, pitch perception appears to have been influenced by the stimuli alteration procedure for the harsh

recordings. Table 4.5 shows that the *deep* label was used more frequently in lower-pitched samples than in higher-pitched samples across the four conditions. One possibility as to why this was the case for the harsh samples but not the falsetto or creak stimuli is that the pitch levels for the harsh stimuli are closer to the normal frequency range for average male speech (see, for example, Hudson et al., 2007) compared with the creak and falsetto stimuli. Listeners seemed to have little trouble in attributing the harsh voice stimuli to a male speaker, and the samples were generally attributed to an older speaker compared with age attributions for other voice qualities. However, the range of age attributions was again wide across the harsh stimuli, with a range as high as 50 years for the slower speed, higher-pitched stimulus.

4.3.1.4. Modal voice

	Modal			
	Faster speed + Lower pitch	**Faster speed + Higher pitch**	**Slower speed + Lower pitch**	**Slower speed + Higher pitch**
Descriptors assigned	*Calm (33)* *Deep (14)* *Slow (8)* *Excited (4)* *Rapid (3)*	*Calm (33)* *Slow (8)* *Nasal (5)* *High pitched (5)* *Rapid (3)*	*Calm (36)* *Slow (24)* *Deep (13)* *Hoarse (4)* *Nasal (2)* *Disguised (2)* *Lisp (2)*	*Calm (33)* *Slow (22)* *High-pitched (6)* *Hoarse (2)* *Harsh (2)* *Nasal (2)*
Perceived sex	*Male: **98%*** *Female:**2%***	*Male: **75%*** *Female: **25%***	*Male: **98%*** *Female: **2%***	*Male: **78%*** *Female: **22%***
Perceived age	*Mean age: **34*** *Range: **19-56*** *Std. Dev.: **7***	*Mean age: **32*** *Range: **17-50*** *Std. Dev.: **10***	*Mean age: **37*** *Range: **25-53*** *Std. Dev.: **8***	*Mean age: **32*** *Range: **14-60*** *Std. Dev.: **11***
Total no. of listeners	**41**	**39**	**40**	**40**

Table 4.6 – Descriptors assigned to *modal* stimuli by listeners

For the modal recordings, the *calm* descriptor was used more frequently than any other label, with 84% of listeners attributing this descriptor. As previously identified for the *angry* label, there is a potential problem with this attribution as it could imply that the speaker *was* calm rather than *sounding as though* they were calm. The read-aloud nature of the samples may have contributed to the frequent use of the *calm* label, and its popularity could also be due to an absence of specific ways to describe a regular modal voice on the bomb threat checklist. The pitch and speed alterations are successfully distinguished by listeners for the modal recordings, with greater use of the *slow* and *deep* descriptors in the slower and lower-pitched conditions respectively. As was the

case for the falsetto stimuli, pitch manipulation in the modal voice stimuli appeared to affect listener sex perception. For the lower-pitched recordings, the classification as male was near-categorical, whereas for modal stimuli with raised pitch, approximately one in every four listeners labelled the voice as belonging to a female speaker. Again, the range of age estimations was wide, ranging from a speaker in their mid to late teens to a speaker aged 50-60. The mean age estimations for the modal samples were also lower than for the harsh and creaky stimuli, with perceptions of speakers in their mid-30s as opposed to speakers in their 40s and 50s.

4.3.1.5. Whispery voice

	Whispery			
	Faster speed + Lower pitch	**Faster speed + Higher pitch**	**Slower speed + Lower pitch**	**Slower speed + Higher pitch**
Descriptors assigned	*Whispery (36)* *Calm (13)* *Deep (10)* *Rapid (10)* *Hoarse (8)*	*Whispery (34)* *Rapid (11)* *Hoarse (8)* *Calm (8)* *Excited (7)*	*Whispery (39)* *Slow (21)* *Calm (16)* *Deep (12)* *Hoarse (9)*	*Whispery (29)* *Slow (17)* *Disguised (15)* *Hoarse (12)* *Creaky (12)* *Calm (12)*
Perceived sex	*Male: **100%*** *Female: **0%***	*Male: **41%*** *Female: **59%***	*Male: **93%*** *Female: **7%***	*Male: **31%*** *Female: **69%***
Perceived age	*Mean age: **37*** *Range: **22-70*** *Std. Dev.: **10***	*Mean age: **40*** *Range: **24-78*** *Std. Dev.: **13***	*Mean age: **36*** *Range: **20-64*** *Std. Dev.: **10***	*Mean age: **44*** *Range: **22-70*** *Std. Dev.: **13***
Total no. of listeners	**40**	**39**	**46**	**41**

Table 4.7 – Descriptors assigned to *whispery* stimuli by listeners

For the whispery voiced stimuli, the perception of *whispery* appeared to be very salient to listeners and was consistently identified across the samples, with an 83% listener attribution rate. Additionally, *whispery* was the most frequently identified trait in each of the four conditions. Both relative speech rate and relative pitch perception were in line with the results for the previously described samples. For the perceived sex of the speaker in the samples where the pitch had been lowered, listeners near-categorically stated that the speaker was male. However, for recordings where the pitch had been raised, listeners were more evenly split when determining the sex of the talker, and favoured attributions of a female speaker as opposed to a male talker. This uncertainty

across the group was particularly evident in the faster speed, higher pitched samples, where 31% of listeners deemed the speaker to be male, and 69% of listeners stated that the speaker was female. As was the case for the other voice qualities, age estimations for the whispery voice stimuli were wide-ranging, from a speaker in their early-to-mid 20s to a speaker who had reached retirement age.

4.3.2. Effect of acoustic variables on listener threat ratings

In addition to describing the voices they heard, listeners were required to evaluate the stimuli by providing ratings for how threatening they thought the speaker in each recording sounded. This section examines the effect of the controlled-for acoustic variables – *voice quality*, *pitch* and *speech rate* – on listener threat attributions. As was the case in Chapter 3, statistical analysis was conducted using R (R Core Team, 2015) using linear mixed effects regression models (hereafter `lmer`) constructed under the `lme4` package in R (Bates et al., 2015). Main effect p-values were calculated through model comparisons, constructed using likelihood ratio tests under the `anova` function in R. Further analysis of within-variable effects was conducted using Holm-Bonferroni corrected Tukey pairwise comparisons, constructed under the `multcomp` package in R (Hothorn et al., 2008). For this model, *voice quality*, *speed*, *pitch* and *perceived speaker sex* were included as fixed effects. Additionally, as two contrasting contexts were used in the experiment, context was included as an interaction term with each of the fixed effects. As pitch manipulations took place within voice quality categories, an interaction between pitch and voice quality was included in the model. Participant was incorporated into the model as a random effect.

4.3.2.1. Voice quality

Five voice qualities were tested in this study: *creak*, *harsh*, *falsetto*, *modal* and *whispery*. The output of the statistical analysis revealed a significant effect for voice quality on listener perceptions of how threatening speakers sounded ($\chi^2(16)=638.5$, $p<0.001$). This effect is illustrated in Figure 4.1, which plots the `lmer` model output for the effect of voice quality on listener threat ratings.

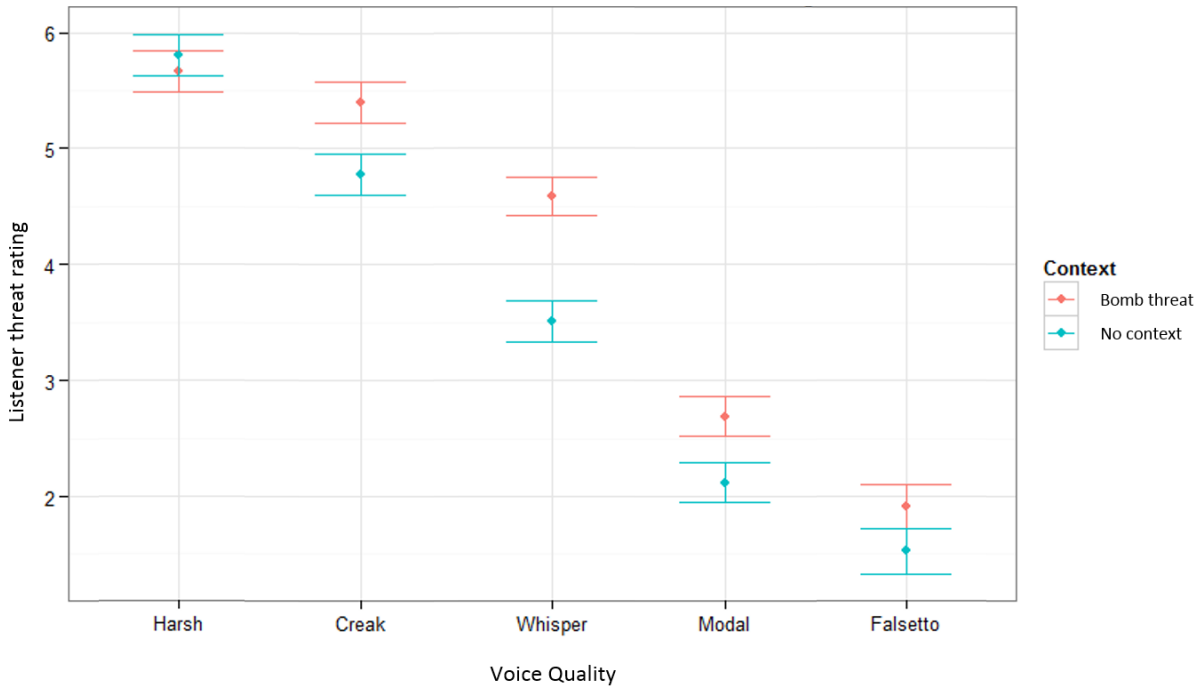


Figure 4.1 – Listener threat ratings for each voice quality

Figure 4.1 illustrates that listener threat attributions were highest (most threatening) for the *harsh* voice quality and lowest (least threatening) for the *falsetto* voice quality. *Modal* voiced stimuli were rated as sounding comparatively less threatening than the *creak*, *whisper* and *harsh* recordings, but more threatening than the *falsetto* recordings. Figure 4.1 also illustrates the effect of the two context groups on listener threat attributions. For all voice qualities aside from the *harsh* recordings, in which high threat ratings were provided by listeners in both experimental context groups, instructing

listeners that they were evaluating bomb threats had an effect on threat attributions. Unsurprisingly, higher threat ratings were assigned by listeners in the bomb threat context. However, Figure 4.1 does illustrate that the effect of context differed between the contrasting voice qualities. Context had the biggest effect for the *whispery* voice stimuli, followed by the *modal* and *creaky* stimuli, the *falsetto* stimuli and finally the *harsh* stimuli. The effect for context was smallest in stimuli at both the higher and lower ends of the scale, and comparatively larger when threat ratings were closer to the middle of the scale.

4.3.2.2. Speech rate

The effect of two contrasting speech rate levels on listener evaluations of how threatening the recordings sounded was examined in this experiment. The two levels tested were *faster* (+10% of overall recording length) and *slower* (-10% of overall recording length). The model output showed that speech rate did not have a significant effect on listener threat ratings ($\chi^2(2)=1.08$, $p=0.58$), with stimuli in the *slower* category not rated significantly differently from stimuli in the *faster* category. This is illustrated in Figure 4.2, which shows the lack of effect for speech rate listener threat ratings, alongside a predictable context effect in line with previous analysis.

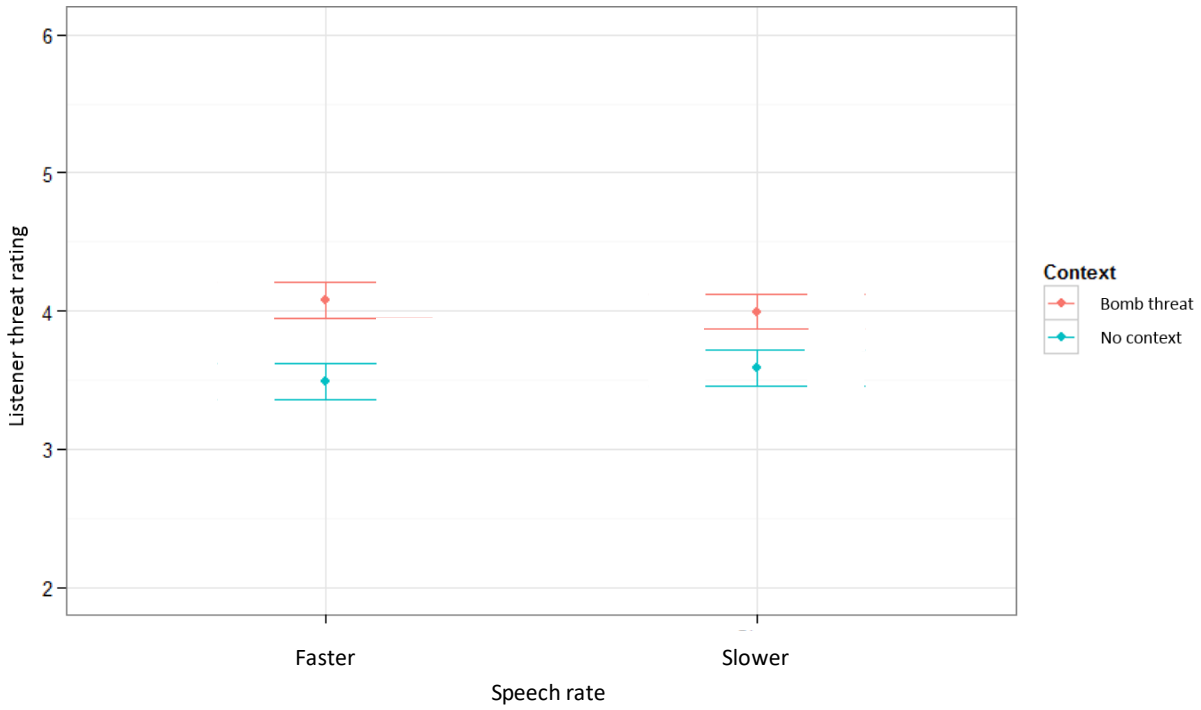


Figure 4.2 – Listener threat ratings for each speech rate category

4.3.2.3. Pitch

As previously outlined in Section 4.2.1, pitch alterations were made relative to the five voice qualities tested in this experiment. The effect for pitch on listener threat ratings was therefore tested in interaction with voice quality in the `lmer` model. The output of this analysis illustrated a significant effect for the interaction between pitch and voice quality *($\chi^2(10)=26.14$, $p=0.003$)*. The effect for pitch in each of the voice quality categories is further illustrated in Figure 4.3, which plots the interaction from the `lmer` model.

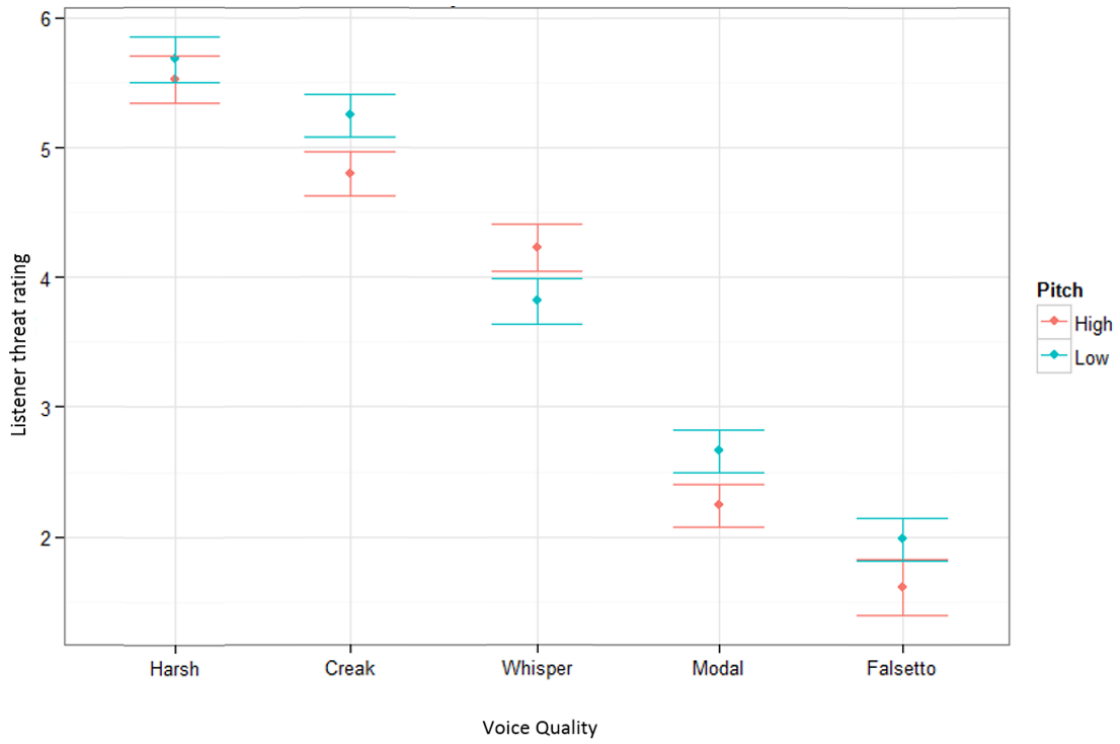


Figure 4.3 – Effect of pitch alterations on listener threat ratings for each voice quality

The effects illustrated in Figure 4.3 show that in each of the voice qualities which contained vocal fold vibration (*creak*, *falsetto*, *harsh*, *modal*), the lower-pitched stimuli were rated as sounding more threatening than the higher-pitched stimuli. This effect is most salient for the *modal* and *creak* voice qualities, and comparatively less clear in the *falsetto* and *harsh* categories. The opposite pattern is seen in the *whispery* voiced stimuli, with the higher-pitched stimuli rated as sounding more threatening than the lower-pitched stimuli.

Given the inherent link between voice quality and F0 (Laver, 1980), with certain voice qualities in the experiment characterised by low F0 (e.g. *creak*) and others characterised by high F0 (e.g. *falsetto*), a further approach for the analysis of the effect of pitch on listener threat ratings is to use the mean F0 values for each acoustic stimulus as a fixed effect. The F0 values for each of the stimuli are displayed in Table 4.8. These values

were calculated using Praat (Boersma and Weenink, 2016), with the maximum pitch level set to 400Hz and the minimum set to 30Hz. All samples were manually checked for aliasing errors before measurements were taken.

Voice quality	Speech rate	Pitch level	F0(Hz)
Creak	Faster/Slower	Higher	**58Hz**
		Lower	**52Hz**
Falsetto	Faster/Slower	Higher	**320Hz**
		Lower	**260Hz**
Harsh	Faster/Slower	Higher	**95Hz**
		Lower	**85Hz**
Modal	Faster/Slower	Higher	**146Hz**
		Lower	**122Hz**
Whispery	Faster/Slower	Higher	**133Hz**
		Lower	**116Hz**

Table 4.8 – Raw F0 values for each of the acoustic stimuli in the experiment

Given the previously-found effect for voice quality on listener threat ratings, an effect of F0 in line with Ohala's (1984) 'frequency code' hypothesis would be argued for this data, owing to the fact that the creak and harsh voice quality recordings were rated as sounding the most threatening, and the falsetto voice quality was rated as sounding the least threatening. However, in order to test for the significance of the F0 values on listener ratings for how threatening the stimuli sounded, a further `lmer` test was conducted, containing the mean F0 values, speech rate and perceived sex as fixed effects, with context included as an interaction term in the model. Participant was also included as a random effect. The model comparison testing on this `lmer` model reported a significant effect *($\chi^2(2)=322.13$, $p<0.001$)* for F0, with lower-pitched stimuli rated as sounding more threatening than higher-pitched stimuli in both context groups.

This effect is shown in Figure 4.4, which plots the output of the effect of F0 on listener threat ratings.

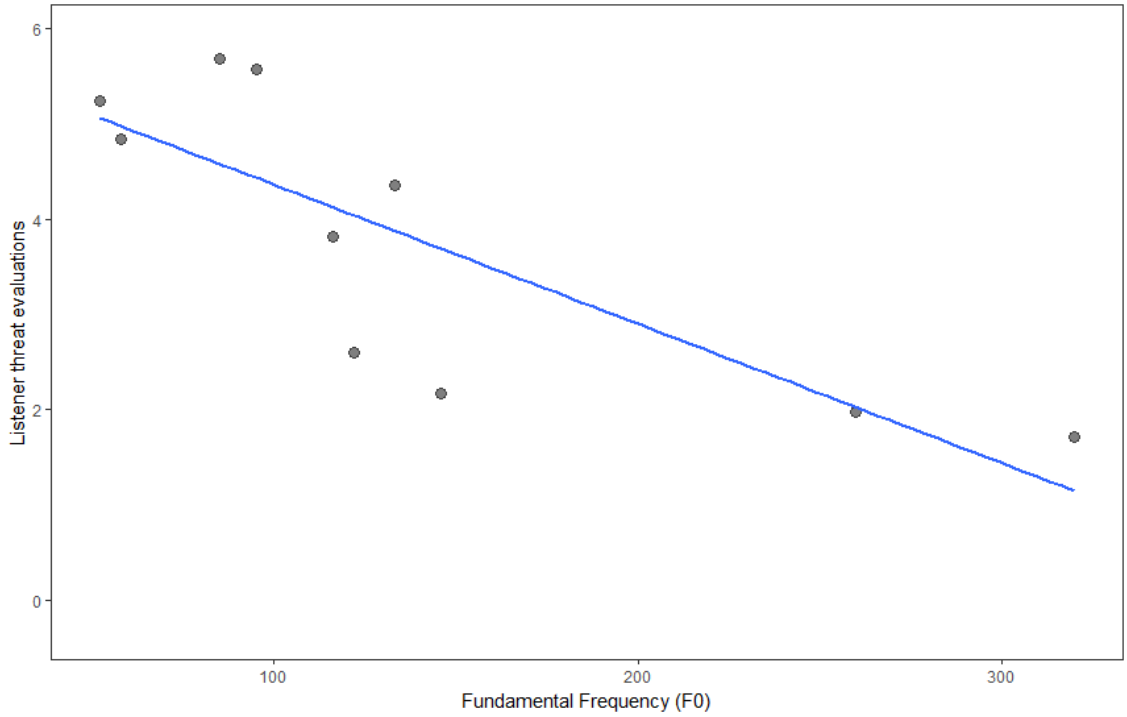


Figure 4.4 – Effect of F0 on listener threat ratings. Points are averaged across listeners for each voice quality sample.

4.3.2.4. Perceived speaker sex

Although all stimuli used in this experiment were produced by a male speaker, the analysis in Section 4.3.1 showed that for some stimuli, listeners thought they were evaluating a female talker. Given this finding, perceived sex was incorporated into the `lmer` model to assess whether those samples perceived to be produced by a male speaker were rated as significantly more or less threatening than those produced by a

perceived female speaker, within the context of the other variables incorporated into the model. Model comparison testing showed that the effect for perceived sex within the `lmer` model was not significant *($\chi^2(2)=3.16$, $p=0.21$)*. The lack of difference between threat ratings assigned to speakers perceived as being male and those perceived as being female is illustrated in Figure 4.5. Figure 4.5 also shows that the difference between the two perceived sexes was particularly minimal in the bomb threat context.

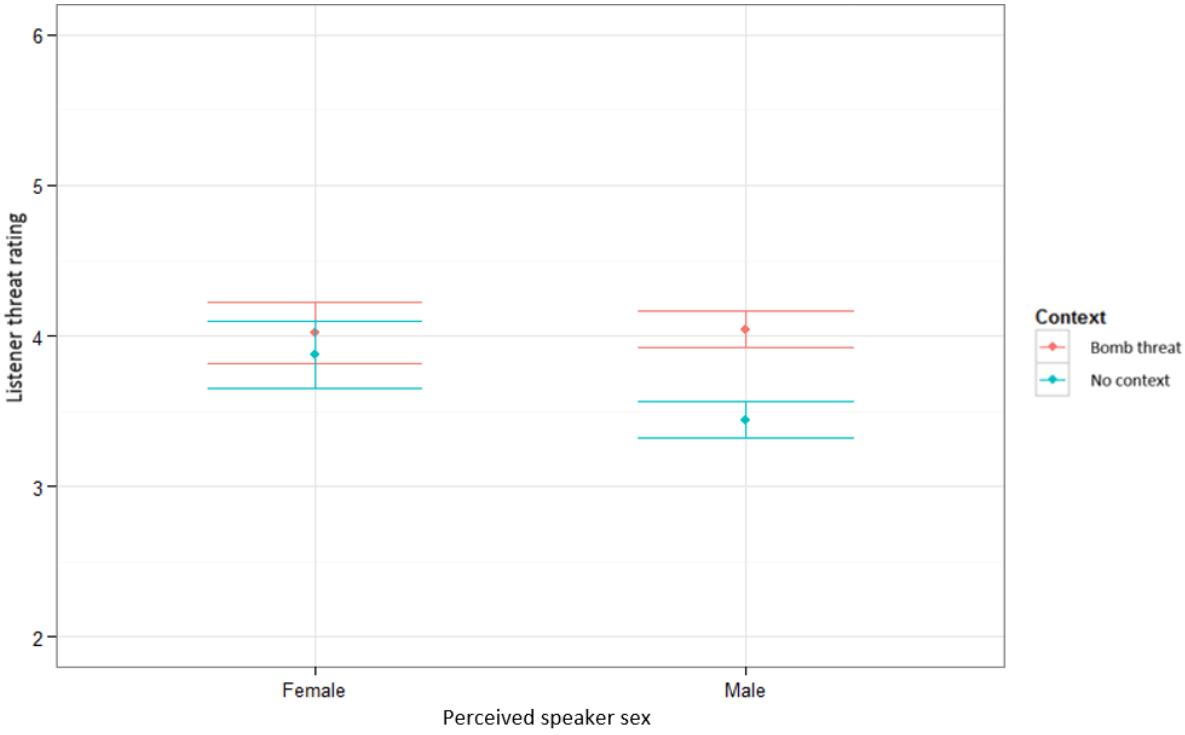


Figure 4.5 – Effect of perceived speaker sex on listener threat ratings

4.3.3 Listener descriptions as fixed effects

In order to draw meaningful conclusions about the findings presented in Section 4.3.2, it is necessary to assume that a perceptual association exists between the acoustic stimuli and subsequent listener threat ratings. For example, the conclusion that the data in

Section 4.3.2 validates the idea that lower-pitched voices are associated with higher threat ratings is based on the assumption that listeners perceive stimuli with lower F0 as being low in pitch, or sounding 'deep'. While this is a plausible assumption, it could be argued that a 'deep' voice for one listener may have a sizeably different mean Hertz value than a 'deep' voice for another listener. Or that terms such as 'deep' and 'high-pitched' have differing acoustic correlates for different listeners. Furthermore, given the link between voice quality and pitch, the results in Section 4.3.2 offer limited scope for deduction of whether voice quality or pitch, or both, were salient to listeners providing threat ratings.

The analysis presented in this section aims to address these issues by analysing the correspondence between both the descriptive attributions and the threat ratings provided by listeners in the experiment. In doing so, it attempts to use listeners' perceptual boundaries as cues for effects, rather than pre-determined fixed values. Each descriptive trait was classified as a categorical variable with two variants; *yes* or *no*. *Yes* indicated that a particular trait was selected, and *no* indicated that a trait was not selected. Threat ratings for the *yes* responses were compared against threat ratings for the *no* responses for each of the traits provided as options for describing the voices presented in the experiment. The accuracy of listeners' descriptions was not considered in this part of the study. If, for example, a listener described a falsetto stimulus as sounding *deep*, then the *deep* classification was included regardless of its rather obvious innaccuracy.

As in Section 4.3.2, statistical analysis was conducted using R (R Core Team, 2015) using `lmer` models constructed under the `lme4` package. An `lmer` model was constructed for each checklist label, incorporating both the label and experimental context as fixed effects. An interaction between context and each perceived trait was

also included in order to account for any differences in traits assigned to the voices by listeners in both context groups. As previously, participant was also included as a random effect. Given the small number of *yes* responses for the *clearing throat*, *laughter*, *crying*, *lisp*, *slurred* and *stutter* descriptors (see Table 4.2), these traits were excluded from the analysis.

Table 4.9 shows the model output for the differences in threat ratings between 'yes' and 'no' responses for the perceived vocal traits, along with the effect of the interaction between the perceived trait and experimental context. As the purpose of this analysis was not to assess whether an overall perceived trait influenced threat ratings, but rather the differences between *yes* and *no* responses for each trait, model comparisons were not used to obtain significance values. Instead, t-values are reported from the model output for the difference between *yes* and *no* responses for each trait. Baayen (2008:248) states that if the t-value in the output of a mixed effects regression model exceeds a value of 2, then significance at an alpha level of 0.05 is achieved.

	Vocal trait effect (t=)	Interaction (with context) effect (t=)
Angry	**11.159**	**2.706**
Deep	**10.85**	*0.24*
Harsh	**10.786**	*1.686*
Creaky	**5.639**	*0.838*
Hoarse	**5.249**	*1.635*
Whispery	**2.630**	**2.689**
Disguised	*0.508*	**4.385**
Rapid	*0.427*	*0.164*
Nasal	*0.007*	*0.772*
Slow	*-0.177*	*1.649*
Excited	*-1.986*	*1.367*
Calm	**-7.098**	**3.859**
High-pitched	**-9.904**	*-0.670*

Table 4.9 – Effect of perceived vocal traits on listener threat ratings (significant results in **bold)**

The results in Table 4.9 indicate several significant effects for the perceived vocal qualities on listener threat ratings, along with contrasting interaction effect patterns. Voices perceived to sound *deep*, *harsh*, *creaky*, and *hoarse* were rated as sounding significantly more threatening than voices for which those descriptive labels were not assigned. Furthermore, there was no significant interaction between the perceived trait and experimental context for these descriptors. These effects are illustrated in the boxplots shown in Figure 4.6, which show that for each of the four perceived vocal traits, threat ratings for voices where the descriptor had been used were higher than for voices where the descriptor was not used. Furthermore, for each of these traits, the difference between voices rated in the bomb threat context and voices rated in the no

context environment was consistent, and in line with the expectation that voices would receive higher threat ratings from listeners in the bomb threat context group.

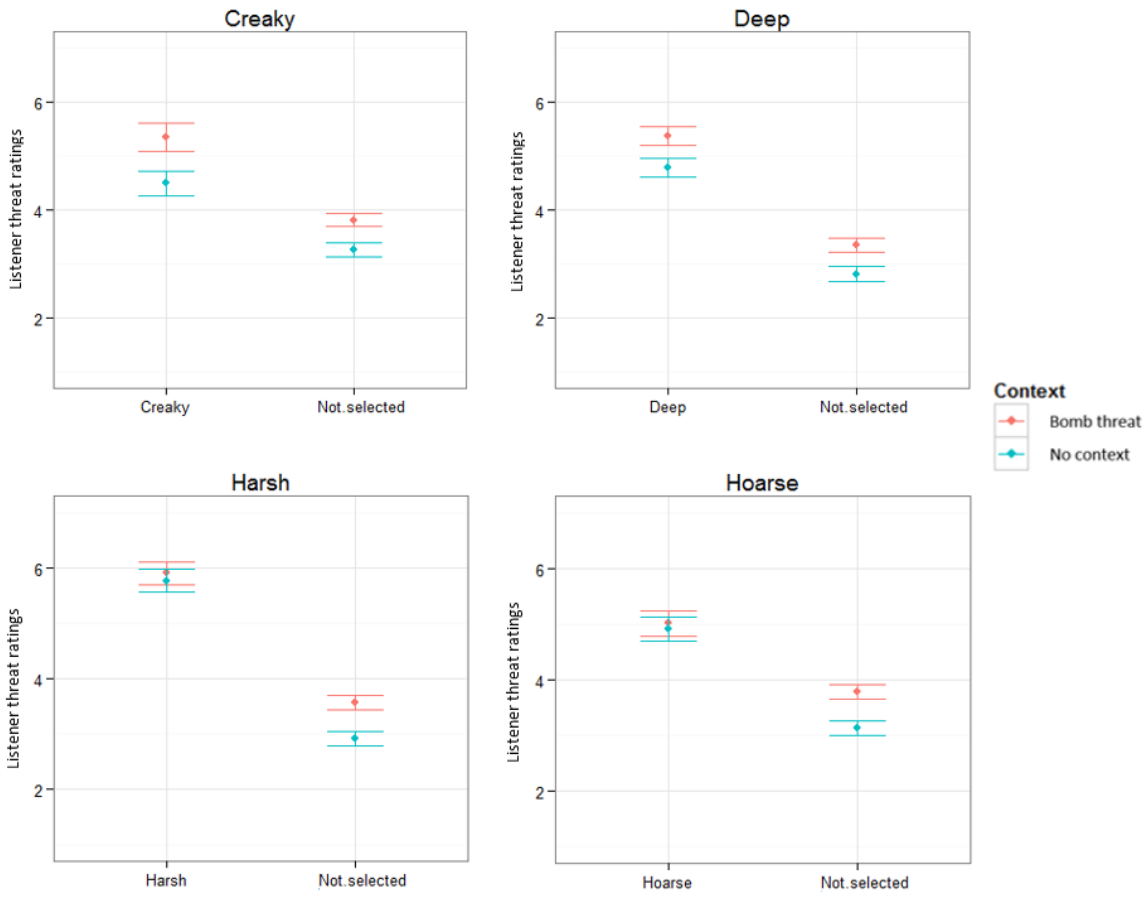


Figure 4.6 – Threat ratings assgined to voices described as sounding creaky, *deep, harsh* and *hoarse* by listeners

The output in Table 4.9 shows that voices described as sounding *angry* and *whispery* were also rated as sounding significantly more threatening than voices where those descriptors were not used by listeners. However, for these descriptors, there was also a significant interaction between the perceived vocal trait and experimental context. Figure 4.7 plots the output of this testing for the *angry* and *whispery* descriptors.

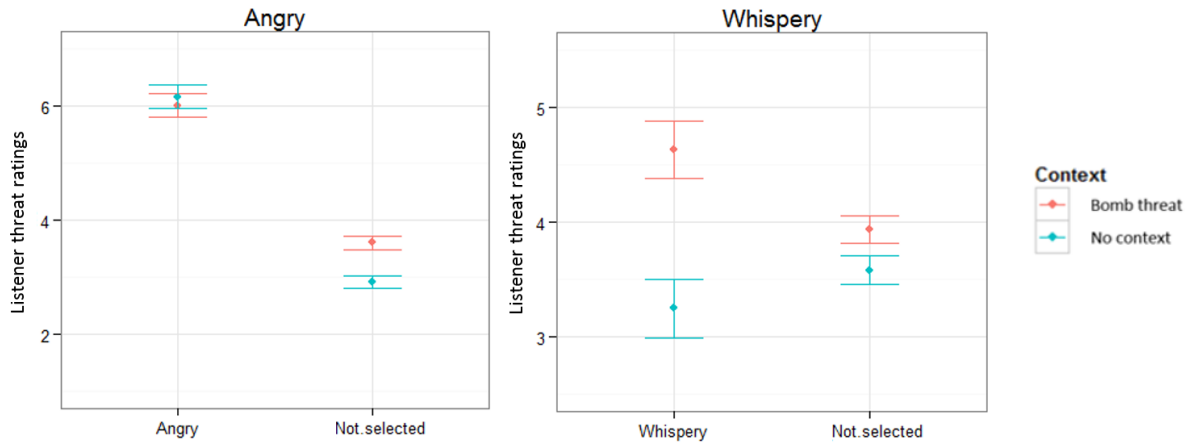


Figure 4.7 – Threat ratings assgined to voices described as sounding *angry* and *whispery* by listeners

Although the interaction terms for both effects were significant, Figure 4.7 shows a different pattern of results for the *angry* and *whispery* descriptors. For the stimuli in which the speaker was perceieved to sound *angry*, listeners assigned higher threat ratings when compared to voices for which the speaker was not described as sounding *angry*. There was a larger difference in assigned threat ratings in the no context group than in the bomb threat context group. There was also a marginally higher overall threat rating assigned by listeners in the no context group for the stimuli described as sounding *angry* compared with those listeners who were in the bomb threat context group. For the stimuli perceived as sounding *whispery*, Figure 4.7 shows that the direction of the effect was different for the two experimental context groups. When listeners were instructed that the utterances they heard were threats, higher threat ratings were assigned to stimuli perceived as sounding *whispery* compared with stimuli not perceived as sounding *whispery*. However, when listeners were not explicitly instructed that the utterances were threats, they asssigned lower threat ratings to utterances described as sounding

whispery compared with those where the *whispery* descriptor was not assigned. Potential motivations and reasons for these results are discussed in Section 4.4.

The results in Table 4.9 also show no significant differences in assigned threat ratings between those voices reported to sound *slow, nasal, rapid* or *excited* and voices for which those labels were not chosen. There were also no significant interactions with context for these vocal traits. The effects for these descriptors are shown in Figure 4.8, which illustrate a clear lack of effect for the stimuli perceived as sounding *nasal*, *rapid* and *slow*. Stimuli in which the speaker was perceived to sound *excited* were assigned lower threat ratings compared with stimuli for which those descriptors were not used. However, this effect was reported by the model as being marginally short of the statistical significance threshold (*t=1.986*).

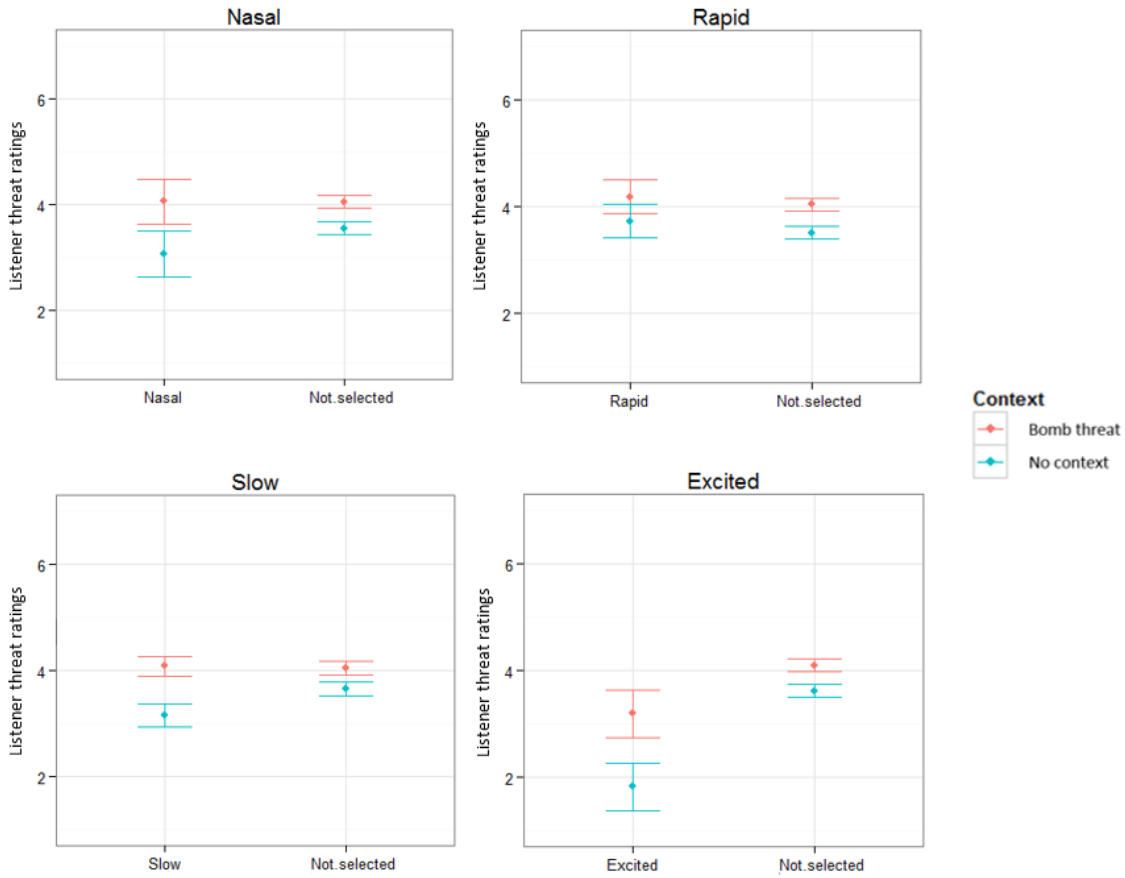


Figure 4.8 – Threat ratings assgined to voices described as sounding *excited*, *slow*, *nasal* and *rapid* by listeners

Table 4.9 additionally shows that voices judged to sound *calm* and *high-pitched* were assigned significantly lower threat ratings than voices to which those labels were not assigned. There was also a significant interaction with experimental context for the stimuli in which the speaker was described as sounding *calm*. These effects are plotted in Figure 4.9. Figure 4.9 also shows a similar effect for stimuli described as being *calm*, although the effect was significantly smaller in the bomb threat context. Potential reasons for this effect are further discussed in Section 4.4.

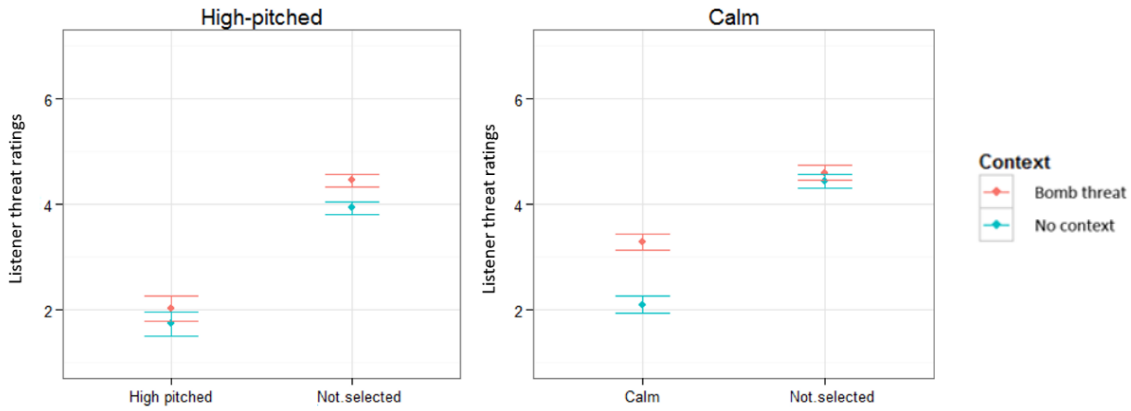


Figure 4.9 - Threat ratings assgined to voices described as sounding *high-pitched* and *calm* by listeners

Finally, Table 4.9 shows no significant difference between those vocal stimuli judged to sound *disguised* and those for which the *disguised* descriptor was not used. However, for this trait, there was a significant interaction with experimental context. This effect is further illustrated in Figure 4.10, which shows that there was a greater effect in the no context environment, with voices judged to sound *disguised* rated as sounding more threatening in the no-context experimental context compared with the bomb threat experimental context.

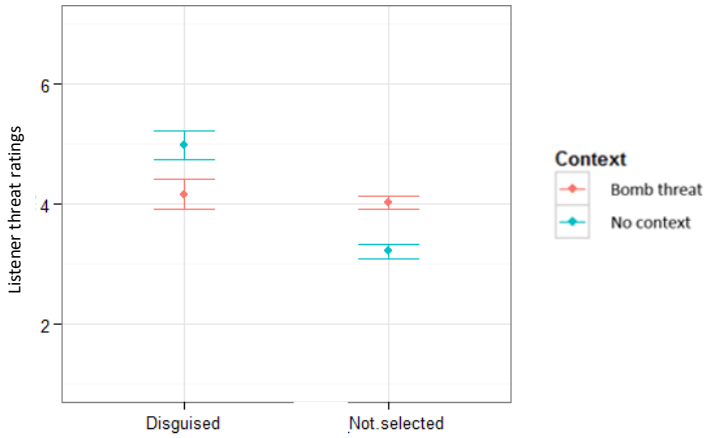


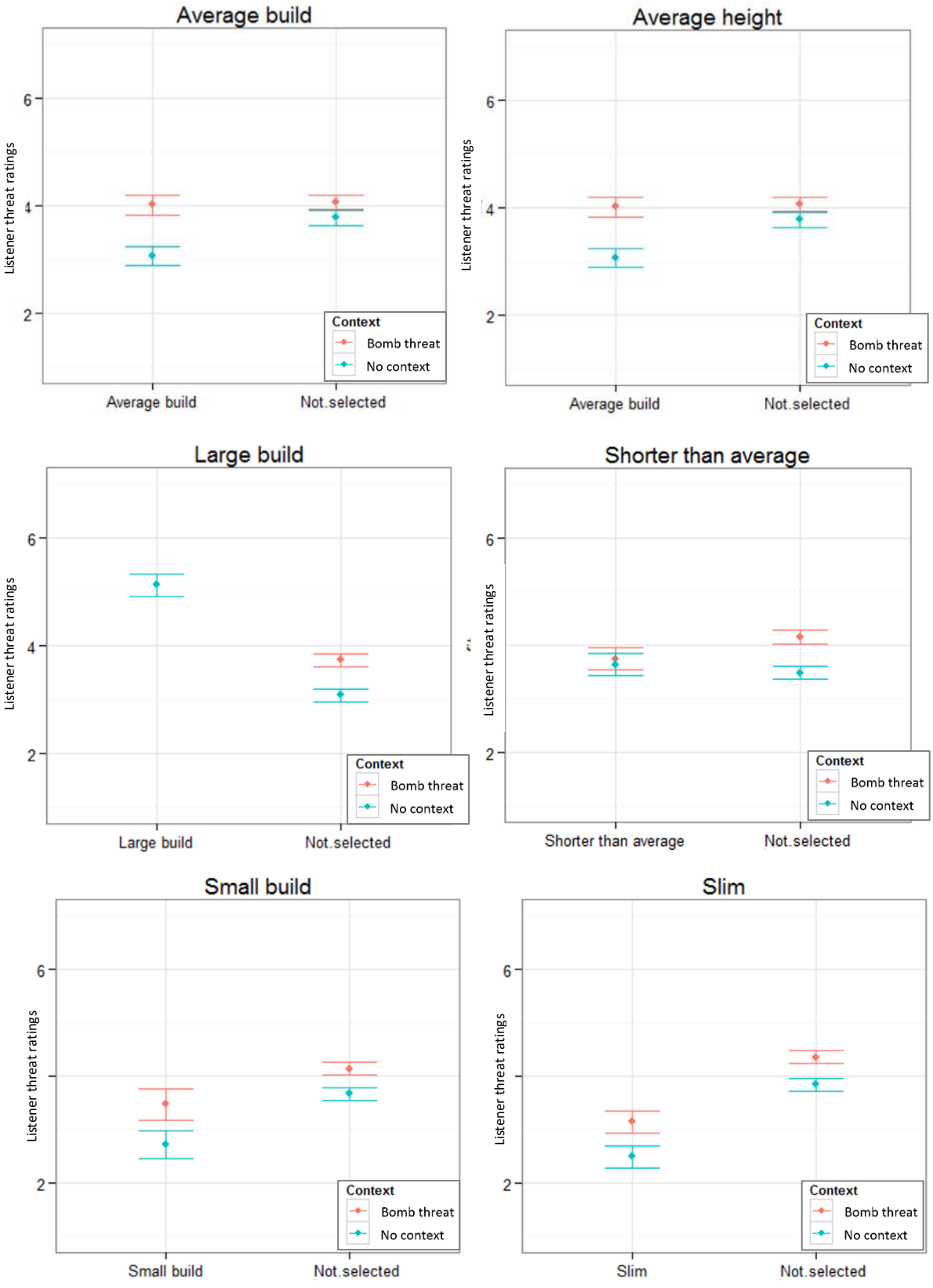
Figure 4.10 - Threat ratings assgined to voices described as sounding *disguised* by listeners

A comparable approach was taken to analyse links between the physical attributions provided by listeners and their perceptions of how threatening speakers sounded. A single `lmer` model was constructed for each physical trait which included the descriptive label, context, and the interaction between context and the perceived physical traits, as fixed effects. As was the case for the vocal attributes, each physical trait was classified as a categorical variable with two variants; 'yes' or 'no'. 'Yes' indicated that a particular trait was selected, and 'no' indicated that a trait was not selected. Participant was included as a random effect. Table 4.10 shows the model output for the differences in threat ratings between 'yes' and 'no' responses for the perceived physical traits.

Perceived physical trait	Physical trait effect (t=)	Interaction effect (t=)	Number of 'yes' responses
Large build	**6.217**	**2.014**	174
Stocky	**4.076**	*1.083*	144
Very tall	*1.373*	*0.159*	51
Taller than average	*1.010*	*0.048*	203
Very short	*0.528*	*0.677*	39
Average build	*0.227*	**2.182**	275
Average height	*0.227*	**2.182**	272
Shorter than average	-1.700	*1.639*	187
Small build	**-2.197**	*0.684*	108
Slim	**-5.332**	*0.416*	188

Table 4.10 - Effect of perceived physical traits on listener threat ratings (significant results in **bold**)

The analysis presented in Table 4.10 suggests that perceptions of speaker build were more closely linked to threat attribution than perceptions of speaker height. Stimuli for which the perceived talker was described as having a *large build* or being *stocky* were assigned significantly higher threat ratings than stimuli for which those descriptive labels were not used. At the opposite end of the scale, speakers described as having a *small build* or being *slim* were judged to sound significantly less threatening than speakers for which those labels were not assigned. The model outputs showed two significant interactions with experimental context for the *average build* and *average height* physical traits. The effects shown in Table 4.10 are further illustrated in Figure 4.11, which plots the model output for each physical trait.


Average build
Listener threat ratings
6
4
2
Average build
Not.selected
Context
Bomb threat
No context
Average height
Listener threat ratings
6
4
2
Average build
Not.selected
Context
Bomb threat
No context
Large build
Listener threat ratings
6
4
2
Large build
Not.selected
Context
Bomb threat
No context
Shorter than average
Listener threat ratings
6
2
Shorter than average
Not.selected
Context
Bomb threat
No context
Small build
Listener threat ratings
6
2
Small build
Not.selected
Context
Bomb threat
No context
Slim
Listener threat ratings
6
2
Slim
Not.selected
Context
Bomb threat
No context

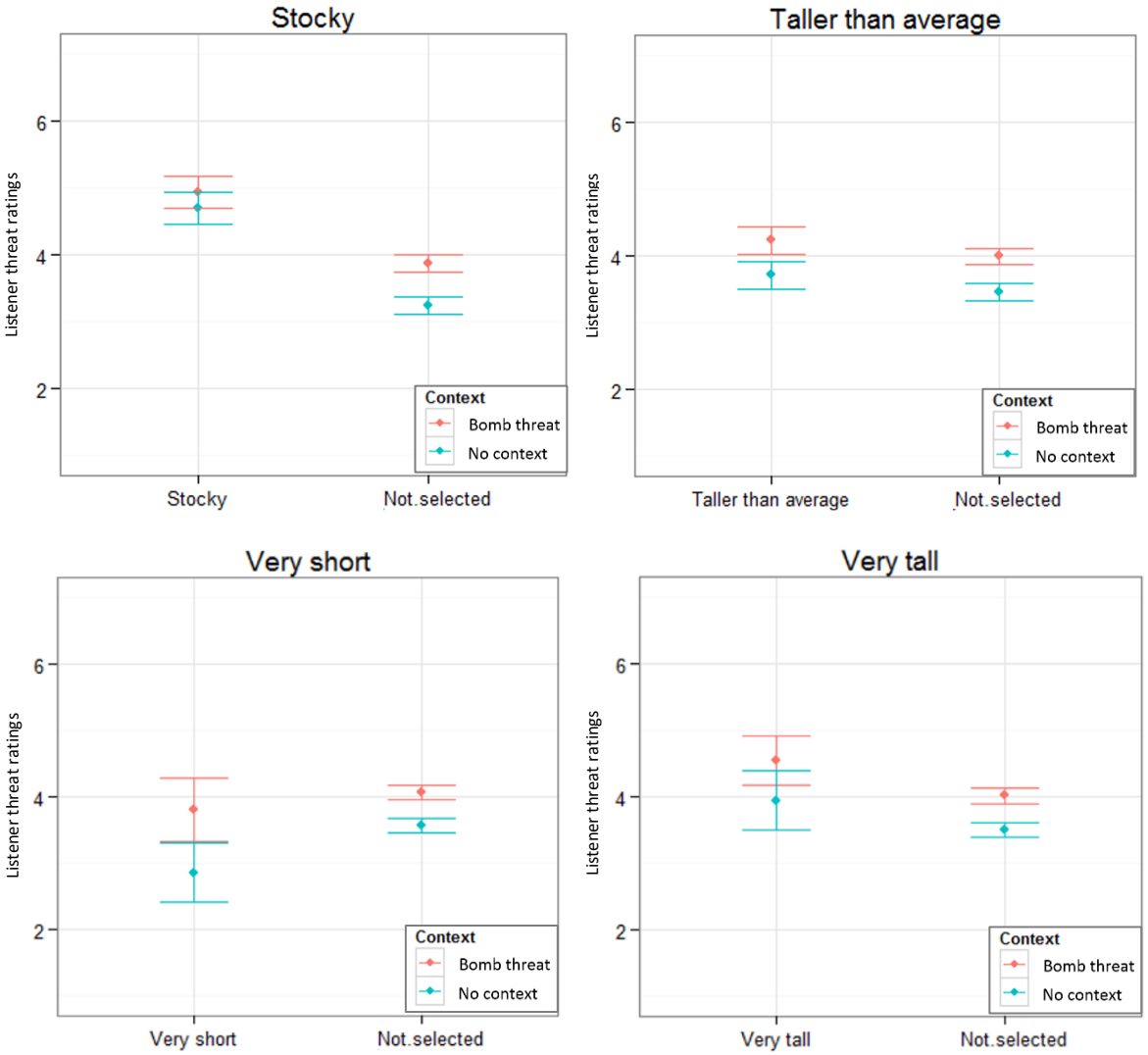


Figure 4.11 - Effect of perceived physical traits on listener threat ratings

One further piece of analysis that was conducted to explore potential 'frequency code' associations in the data was an examination of the perceived physical traits associated with speakers whose voices were described as being either *deep* or *high-pitched*. Given Ohala's (1984) assertion that low-pitched vocalisations are associated with projections of a larger speaker, it would be expected that descriptors associated with projections of larger body size would be more frequently used for voices which were described as being *deep*, and that conversely, descriptors associated with smaller body size

projection would be more commonly used for voices which had been described as being *high pitched*. The results from this analysis are presented in Table 4.11.

	Assigned to voices described as 'deep'		**Assigned to voices described as 'high pitched'**	
	Number (/278)	**%**	**Number (/135)**	**%**
Large build	105	**38**	13	**10**
Taller than average	94	**34**	26	**19**
Average build	92	**33**	34	**25**
Stocky	91	**33**	11	**8**
Average height	84	**30**	42	**31**
Shorter than average	51	**18**	41	**30**
Very tall	24	**9**	6	**4**
Slim	23	**8**	58	**43**
Small build	11	**4**	37	**27**
Very short	12	**4**	13	**10**

Table 4.11 – Physical trait associations for voices described as 'deep' and 'high-pitched'

The figures in Table 4.11 demonstrate links between Ohala's (1984) 'frequency code' hypothesis and the physical and vocal descriptions assigned by listeners in this study. A higher percentage of *large build, stocky* and *taller than average* descriptors were assigned to the voices that were described as sounding *deep* compared with those described as being *high-pitched*. In contrast, a higher percentage of *slim, small build* and *shorter than average* descriptors were assigned to voices that were described as *high-pitched* compared with voices that were described as sounding *deep*. These trends in the results further support the idea that there is a perceptual association between larger body size and low-pitched vocalisations.

4.5. Discussion

The results in this chapter suggest that laryngeal properties of voice can impact on listeners' first-impression formation of unfamiliar speakers with regards to perceived threat. Stimuli associated with lower pitch, both in absolute terms and in relation to voice quality, such as the *creak* and *harsh* samples, were perceived as sounding more threatening than those stimuli associated with higher overall pitch, such as the *falsetto* recordings. The results also show that voice qualities associated with a greater level of laryngeal irregularity were rated by listeners as sounding more threatening than voice qualities associated with laryngeal regularity. This supports the assertions made by previous research about the paralinguistic effects of both pitch and different phonation qualities. Laver (1994:420) argued for an association between harsh voice and high activation emotions such as anger, with Vaissaire (2005:251) stating that glottal irregularity influences affective perceptions of both anger and hostility. Both of these assertions are supported by the results in this study, following the link between perceived anger, threat and aggression shown in the data presented in Chapter 3.

The results also further support research which identifies lower-pitched vocalisations as markers of dominance, anger, threat and the projection of a larger speaker (Ohala, 1984; Feinberg et al., 2005; Gussenhoven, 2002; Chuenwattanapranithi et al., 2009; Xu and Kelly, 2010). The perceptual analysis presented in Section 4.3.3 further emphasises this association. Despite the absence of visual cues, stimuli for which the speaker was described as having a *large build* and being *stocky* were assigned significantly higher threat ratings than stimuli for which those descriptors were not used. Equally, stimuli for which the *small build* and *slim* were used were used were assigned significantly lower threat ratings than stimuli for which those descriptors were not used. Furthermore, the results in Table 4.11 show that descriptors of a speaker being *stocky* or

having a *large build* were also more commonly used for those voices described as sounding *deep* compared with those described as sounding *high pitched.* The reverse also applied, with the *small build* and *slim* descriptors more commonly used when voices were described as sounding *high-pitched*. While somewhat cyclical in nature, these results further strengthen arguments which identify a perceptual association between lower-pitched vocalisations and larger speakers, even in the absence of visual cues.

The results presented in Section 4.3.3 showed that listeners assigned greater levels of threat to voices described as sounding *angry*, *deep*, *harsh*, *creaky*, and *hoarse* compared to voices for which those labels were not assigned. This result would support the view reached from the analysis in Section 4.3.1, which argued that stimuli associated with lower pitch and laryngeal irregularity were perceived by listeners as sounding more threatening than those which were associated with higher pitch and laryngeal regularity. The finding also further highlights the benefits of examining listener perceptions of traits alongside their ratings of the acoustic stimuli.

Several noteworthy interactions were also shown between participants' use of descriptive labels and the experimental context group to which they were assigned. For the majority of vocal descriptors chosen by listeners, trends for both the bomb threat context and no context groups were consistent. For example, listeners in both groups who described voices as sounding *deep* gave those voices higher threat ratings compared with voices which were not described as *deep*. However, for the *whispery* descriptor, listeners in the no context group assigned lower threat ratings to voices described as *whispery* compared with voices not described as *whispery*. However, the opposite pattern was seen for the bomb threat context group. This highlights the

importance of considering context when examining perceived vocal markers of threat. It also suggests that paralinguistic functions of whisper, or perceived whisper, may be more highly context-dependent than other voice qualities or phonation types. This result is also supported by the findings in Section 4.3.1, with Figure 4.1 illustrating a greater effect for context in the whispery stimuli compared with the other voice qualities used in the experiment.

Gobl and Ní Chasaide (2003:204) link whispery voice to perceptions of fear and timidness, which could potentially explain why listeners in the no context group assigned lower threat levels to voices perceived as sounding whispery. Given that listeners in the no context group had no prior reason to assume the utterances were threats, it may have been the case that they felt the speaker was timid, shy and/or afraid, and therefore perceived the speaker to sound less threatening. For listeners who were instructed that the utterances they were evaluating were bomb threats, it can be argued that such associations would be less likely owing to the provision of the contextual information. Laver's (1994:190) association between whispery phonation and secrecy may be more applicable for listeners in the bomb threat context group, perhaps if they felt that the phonation quality was used by the speaker as a method of vocal disguise. Equally, as whispery voice is characterised by a lack of regular vocal fold vibration plus a higher degree of friction through the glottis, listeners in the bomb threat context may have been more willing to evaluate the samples alongside the other qualities which contained a greater degree of laryngeal irregularity, in this case the creak and harsh stimuli.

The results in Section 4.3.3 also highlighted that there were several vocal traits that were linked to increased or decreased threat perception regardless of the context in

which listeners evaluated the utterances. A high degree of overlap was seen between listener threat ratings in the two context group for voices perceived to sound *harsh*, *angry* and *hoarse*, indicating that listeners did not seem to be influenced by context when making these associations. Notably however, although minimal differences existed between the two contextual groups in the assignment of threat ratings to voices described as sounding *angry*, there was a significant difference between threat ratings assigned to voices described as sounding *calm* by listeners in the two context groups. Listeners in the bomb threat context assigned significantly higher threat ratings to voices described as sounding *calm* compared with listeners in the no context group. Given that *calm* could be seen as the antithesis of *angry*, the discrepancy between the effect of context on the two results is noteworthy, and suggests that context had a higher effect on listener ratings at the lower end of the scale than at the higher end. The results of this experiment suggest that the provision of the bomb threat context created a higher floor from which listeners evaluated the utterances they heard, but did not create a lower ceiling for listeners in the no context group, who were still willing to assign high threat ratings in the absence of being explicitly instructed that the stimuli they were evaluating were spoken threats. This result may also reflect the ideas that spoken threats can be delivered in what is perceived to be a 'calm' manner, or by a 'calm' speaker, and still sound threatening.

Predictably, the findings of this study also indicate that instructing half of the listener set to believe that the utterance they heard was a bomb threat had an effect on threat evaluations. This illustrates that the provision of contextual evidence or information can influence perceptual evaluations in lab-based experiments (Dixon and Mahoney, 2004), and highlighted the importance of incorporating context when considering how listeners infer threat to harm from vocal cues. The results presented in this chapter validate the

assertion made by Milburn and Watman (1981:10) that situational context is a contributing element to the system under which threats are both communicated and evaluated.

The second strand of investigation in this study was an examination of how listeners described the voices in the experimental stimuli using an adapted version of the NCTSO bomb threat checklist. The results showed that pitch appeared to be more salient to listeners than voice quality, but the findings were broadly consistent with Watt and Burns' (2012) assertion that linguistically-untrained listeners are capable of describing certain voice qualities with a relative degree of accuracy. The minimal presence of the descriptors *crying*, *clearing throat*, *laughter*, *lisp*, *slurred*, and *stutter* can be viewed as further evidence of listener accuracy with respect to vocal descriptions, as none of these phonetic markers were present in the acoustic stimuli.

However, for the falsetto samples, many listeners mischaracterised the speaker as female rather than male. Given that F0 for a male falsetto voice can extend into, and in some cases beyond, the standard F0 range of a female speaker (Hollien and Michel, 1968), this is not an altogether surprising result. However, it does show that the F0 of a speaker's voice has the ability to fool listeners into thinking that they are hearing a speaker of the opposite sex. This result also illustrates the need for caution in accepting listener evaluations of speaker sex in forensic contexts when there is suspected vocal disguise through the use of falsetto voice. Equally, the split results for the perceived sex of the speaker in the higher-pitched whispered phonation stimuli suggest that equal caution should be advised in accepting listener judgements of speaker sex if that speaker's vocalisations were made using a whispered or whispery phonation type.

Given the finding that there are aspects of the voice that linguistically untrained listeners can describe with some relative accuracy, it can be argued that work should be done to ensure that documents such as the NCTSO elicit relevant information about speakers' voices. However, the results of this study highlight several issues with the bomb threat checklist in its current form. One such issue is that listeners readily used the terms *angry* and *calm* as descriptors. *Calm* was the most applied descriptor in the experiment (314 instances), while 147 *angry* labels were assigned by participants. *Calm* also appeared as the more frequently used descriptor for the modal voice stimuli, and *angry* was the most commonly used descriptor for the harsh voice stimuli. These descriptors have limited speaker identification value because they do not refer to a specific aspect of voice. Additionally, they also open up the possibility of inference from information about a speaker's psychological or emotional state, which could be problematic in legal scenarios. Furthermore, the results in Section 4.3.3 highlighted strong perceptual associations between anger and threat, which could create unwanted biases if listeners are asked to state whether a threatener sounded angry. The same issue could also apply to the *harsh* descriptor, which has a technical linguistic meaning relating to voice quality, but also a wider meaning which could result in a speaker being described as being *harsh*, or with a listener inferring that they had a *harsh* personality. Problematically for the descriptors as used on the NCTSO checklist, it would not be clear as to which type of descriptor was being used by a given checklist user.

Another issue highlighted by the results of this experiment is that the use of categorical, tick-box style descriptors do not accurately capture many of the phonetic variables that the checklist aims to elicit information about. Some of the traits, such as *laughter* or *clearing throat*, lend themselves to categorical yes/no descriptors, but others, such as those determining pitch or aspects of voice quality, are arguably better viewed on a

continuous scale. This was particularly well illustrated for the modal voiced stimuli, for which listeners used fewer descriptors than they did for the more extreme voice quality variants. This may not necessarily be due to the lack of perceptible aspects of voice, but the lack of provision for listeners to describe non-extreme positions using the checklist. Additionally, the lack of provision to provide more detail about certain perceived vocal traits arguably further limits the usefulness of the checklist. This is particularly true for the *disguised* label, which appeared in the top five descriptors for stimuli in the *falsetto*, *modal* and *whispery* voice quality categories. However, owing to a lack of scope for further clarification, it is unclear what type of disguise listeners felt the speaker was adopting.

The current NCTSO bomb threat checklist encourages users to assess, and to provide an estimate of, the age of a given threatener. This task was replicated in this study, with results showing that estimates of speaker age varied considerably both within and between voice quality categories. Given that the speaker was the same person in each recording, these findings highlight a potential weakness with this aspect of the bomb threat checklist. Furthermore, if decisions about pursuing potential threateners were made using information from the checklist, targeting suspects of a particular age could be counterproductive given the wide estimates of age for the speaker in each of the vocal stimuli. It also further highlights the limited potential of lay-witness age identification from voices where voice quality is suspected as a source of vocal disguise.

Watt and Burns (2012) argue that there would be merits in providing a UK-wide standardised document for the elicitation of vocal information from linguistically untrained earwitnesses, along with a set of guidelines for its use. The research presented

in this study would provide evidence to further advocate this approach in relation to the NCTSO bomb threat checklist. The findings support the view that listeners are capable of eliciting some meaningful descriptive information about speakers' voices upon exposure to short recordings. However, more useful and accurate information could be obtained by making alterations to the checklist and providing linguistically-informed advice to its users.

While the results presented in this chapter highlight several key considerations for the overall research questions outlined at the outset of this thesis, it should be acknowledged that as a standalone study, the research in this chapter does not exist without limitations. As Watt and Burns (2012) assert, the stimuli provided by Eckert and Laver (1994) that were used in this study represent examples of extreme voice qualities and therefore would have likely provided stronger perceptual cues than more 'realistic' speech samples. However, the use of such data facilitates the understanding of the extreme bounds of responses. The goal of the research presented in the remainder of this thesis is to assess how well the findings presented in this chapter apply when multiple speakers, talking using their 'regular' voices and in a language familiar to the listeners are considered within a similar evaluative paradigm. This study, while considering more linguistic variables than the research in Chapter 3, did not account for other potentially influencing variables such as intonation (Scherer, 2003), speaker accent (Dixon et al., 2002; Dixon and Mahoney, 2004) or speaker gender (Watt, Kelly and Llamas, 2013). These will be considered in the work presented in subsequent chapters. Overall, the work in the rest of the thesis works towards creating a more holistic model of threat perception within prescribed contexts. The findings of this study can be viewed as an extension to the beginnings of work which aims to critically explore the influence of phonetic variation on listeners' inference of threat from

speakers' voices, along with the practical implications that these inferences have for the analysis of spoken language crimes by lay-listeners and experts alike.

Chapter 5 – Combining production, perception and description

5.1 Introduction

The research presented in this chapter further extends the findings of the research presented in Chapters 3 and 4. The previous chapter highlighted potential links between phonetic properties of voice and listeners' perceptions of how threatening speakers sound. However, while the use of vocal stimuli in languages that were unfamiliar to listeners allowed for a critical assessment of the influence of vocal parameters independently of lexical content, the extent to which this accurately replicates real-world lay-witness voice evaluation tasks is, undoubtedly, questionable. Furthermore, as the utterances used as experimental stimuli in Chapter 4 were designed as reference materials for the illustration of different voice quality types, the extremity of features in the voices presented to listeners may have provided stronger perceptual cues than would otherwise be present in more 'regular' voices. Additionally, while the use of a single speaker in a vocal guise experiment allowed for closer control over the tested prosodic aspects of voice, listeners still only ever heard the voice of a single speaker. While this may have been mitigated to some extent by the notion that listeners may have believed they heard multiple speakers, as is the aim in matched guise experiments such as the one conducted in Chapter 3, widening the scope of the research presented in previous chapters to include multiple speakers rather than verbal guises would arguably help to better test the generalisability of the previously found results.

Acoustic phonetic analysis in this thesis has, so far, been conducted on what could be described as a holistic level, as features have been measured and analysed across entire stretches of speech. The effects of these holistic measurements on listeners' perceptions

of how threatening speakers sounded were then examined. While this process allows for an objective assessment of why one speaker or guise may have been perceived to sound more or less threatening than another speaker or guise, any within-speaker, between-utterance effects that may exist are not accounted for. This approach also does not consider how phonetic variation within certain parts of utterances, such as the realisation of individual words, could influence listeners' judgements of how threatening a speaker sounds.

The research presented in this chapter aims to extend the methodological framework used in Chapters 3 and 4 by applying it to contentful indirect bomb threat utterances produced by multiple British English speakers. The experiment also factors a further phonetic variable, emphasis on particular words, into the evaluation paradigm, and provides an assessment of how the phonetic realisation of emphasis can interact with linguistic markers that have been previously identified as having links to increased threat perception (Napier and Mardigian, 2003; Gales, 2010; Nini, 2017). This chapter firstly considers differences between speakers' productions of indirect threat utterances in what they considered to be a *threatening tone of voice* and a *neutral tone of voice* with respect to emphasis on the word *will*. The extent to which this acts as a perceptual marker of threat for listeners is subsequently assessed.

Additionally, the research in this chapter further expands on the analysis of listeners' voice descriptions presented in Chapter 4 by offering additional insights into how phonetically untrained listeners describe the voices of unfamiliar speakers. This work aims to further assess the usefulness of protocol documents such as the NCTSO bomb threat checklist, which explicitly instructs listeners to describe the voices of bomb threat perpetrators.

5.2. Speakers' productions of a threatening and neutral tone of voice

5.2.1. Methodology

To create the stimuli for this study, 27 speakers (9 male) were firstly instructed to read a neutral passage aloud, and then asked to familiarise themselves with a series of 9 utterances. The neutral passage was a phonetically balanced text entitled 'Fern's Star Turn', with the set of utterances consisting of indirect threats concerning a range of topics. These utterances are detailed in Table 5.1, in the order in which they were produced by the listeners.

Number	Utterance
1	I know where you live
2	I wouldn't do that if I were you
3	Are you sure you want to do that?
4	When I get out of here I'm going to do something about this
5	There's a bomb at York Station. It will go off this afternoon.
6	How's your mum at the moment?
7	Do you know there's a bomb at York Station set to go off this afternoon?
8	It gets really lonely around here at night
9	I'm warning you about a bomb at York Station which will go off this afternoon.

Table 5.1 – List of utterances recorded in advance of the experiment

The aim of requiring speakers to read a neutral passage in advance of producing the utterances in Table 5.1 was for them to become familiar with the recording procedure in advance of being asked to produce the threat stimuli. Once the speakers had familiarised themselves with the indirect threats, they were asked to produce each utterance in what they considered to be a *neutral tone of voice*, and then again in what they considered to be a *threatening tone of voice*. This follows the experimental procedure used by Watt, Kelly and Llamas (2013) in their investigation of spoken threats. No guidance was provided by the researcher on which, if any, linguistic features should be altered by the speaker when they produced the two versions of each utterance, meaning that speakers were free to signal threat or neutrality in any way they felt was appropriate.

As Table 5.1 shows, included within the 9 utterances were the sentences, *"I'm warning you about a bomb at York Station, which will go off this afternoon"* and *"There's a bomb at York Station which will go off this afternoon"*. These utterances were used as target stimuli for the current investigation as the second clause in both utterances is a direct declarative which contains the modal verb *will*. As detailed in Chapter 2, use of the word *will* was highlighted by Gales (2010) as a linguistic feature which people perceive as being a signal of increased threat and intent-to-harm. However, in the Communicated Threat Assessment Reference Corpus (CTARC), the construction was actually more commonly associated with non-realised threats than realised threats. Nonetheless, Gales (2010) reports that the presence of *will* was identified by both threat assessment professionals and scholars as a marker of an increased level of commitment to the threat, rather than as a feature which weakens speaker commitment. Additionally, in a corpus of authentic malicious forensic texts (hereafter MFTs), Nini (2017) found that a higher percentage of prediction modals existed in threatening communications compared with non-threatening communications. Nini (2017:112) argues that prediction

modals such as *will* are utilised by speakers to emphasise certainty surrounding the outcome of the threatened action.

In addition to the use of utterances which contained modal *will*, the stimuli used for this experiment were also designed to closely represent common types of real-world threats. In the MFT corpus, Nini (2017) reports that 59% of threats were indirect, compared with 4% direct and 27% conditional, and that 78% of threat types alluded to a violent act. Nini (2017) also reports that 38% of MFTs were directed towards a third party rather than towards the recipient of the threat, compared with 25% of MFTs that were explicitly and directly targeted towards the recipient. The use of an indirect threat which is not personally targeted towards the recipient, and expresses the threat of a violent act would, therefore, be the most common combination according to Nini's (2017) research. This combination of features reflects the threat type and direction of harm represented by the utterances analysed in this experiment.

Additionally, Napier and Mardigian (2003) further highlight a range of features which are associated with high-level and low-level threats. Features identified as low-level include a lack of detail about times, places and people, along with mitigating language features such as the use of modals such as *may* and *might*. Conversely, higher-level threats are more likely to contain specific details about people, places and times, facts which can be verified, and a threatened action which is both plausible and realistic (Napier and Mardigian, 2003; Gales, 2010). The utterances *"I'm warning you about a bomb at York Station, which will go off this afternoon"* and *"There's a bomb at York Station which will go off this afternoon"* specify both a designated time and a place, and presents a threat which is both specific and realistic, given that only a single bomb and a

single target is mentioned. This is in contrast to the example of a low-level threat provided by Gales (2010) - *"I will blow up every building on campus at the same time"* - where the threat is both unspecific and unrealistic owing to the scale of the threatened action.

Furthermore, the utterances chosen for this study represent examples of indirect threats which are interpretable as other types of speech acts, in this case as either a warning or a statement. Given that the difference between threats and warnings concerns whether or not the speaker has designed his/her utterance to be in the hearer's interest or to the hearer's detriment (Fraser, 1998), *"I'm warning you about a bomb at York Station, which will go off this afternoon"* can be interpreted as being either a threat or a warning by hearers, and as such requires listeners to infer its most likely interpretation. Likewise, it can be argued that a threatener who disguises his or her threat as a statement of fact, as in the utterance *"There's a bomb at York Station, it will go off this afternoon"*, can displace responsibility by removing themselves as an agent of the threatened action.

Measurements were made for *mean F0, F0 range, speech rate* and *mean intensity* in order to capture differences in pitch, tempo and loudness across each of the target utterances. All measurements were made using Praat software. The mean F0 for each of the utterances was calculated using the 'Get Pitch' function. For male voices, the maximum pitch was set to 200Hz and the minimum pitch was set at 75Hz. The maximum pitch value for female speakers was set at 400Hz and the minimum at 100Hz. Errors in the Praat pitch trace were manually corrected before measurements took place. F0 range was calculated as the difference between the maximum and minimum F0

values across each utterance. The speech rate of each stimulus was taken as the average number of syllables per second of speech, while the mean intensity was measured as the average decibel (dB) level across the entirety of each utterance.

In addition to this, measurements of *mean F0, duration* and *mean intensity* were also captured across each individual word within the utterances in order to assess the relevant phonetic cues to prominence. These measurements were taken based on previous assertions about F0, duration and intensity being key acoustic correlates of lexical prominence. Again, all measurements were made using Praat. The sound files were marked up with Praat text grids, with a tier used to separate each word. Duration was measured in milliseconds from the start to the end point of the word, whereas for both F0 and intensity, an average measurement was taken across each token. These measurements were extracted using the ProsodyPro script (Xu, 2013).

5.2.2. Results

In order to test for differences between the measured acoustic phonetic variables in speakers' *threatening* and *non-threatening* productions of the two indirect bomb threat utterances, linear mixed effects models were constructed with each of the phonetic measures as the dependent variable, and both the version of the utterance (*threatening / non-threatening*) and speaker sex (*male / female*) as independent variables. Linear mixed effect modelling was used so that speaker could be included as a random effect, given that the experiment used multiple speakers. P-values were obtained using likelihood ratio model comparison tests constructed using the `anova` function in R. The first step in the analysis was to assess the differences between utterance-level phonetic features used by speakers in their *threatening* and *neutral* tone of voice productions.

Table 5.2 shows the results of the model comparisons testing for significant difference between speakers' *threatening* and *neutral* productions of the two indirect bomb threat utterances.

	χ^2	Df	p
Mean F0	6.4335	1	**0.01**
F0 range	6.77	1	**0.009**
Mean Intensity	3.14	1	0.08
Speech rate	1.17	1	0.19

Table 5.2 – Output of lmer testing for phonetic differences between *threatening* and *neutral* utterances in production data (significant effects displayed in **bold**)

The output in Table 5.2 shows two significant differences in *mean F0* and *F0 range* between listeners' productions of utterances produced with a *neutral tone of voice* and utterances produced in a *threatening tone of voice*, with no significant effects for *mean intensity* or *speech rate*. These effects are further illustrated in Figures 5.1 and 5.2, which show trends in the data for utterances produced in a threatening tone of voice to have a higher F0 range (Figure 5.1) and higher mean F0 (Figure 5.2) compared with those produced in a neutral tone of voice.

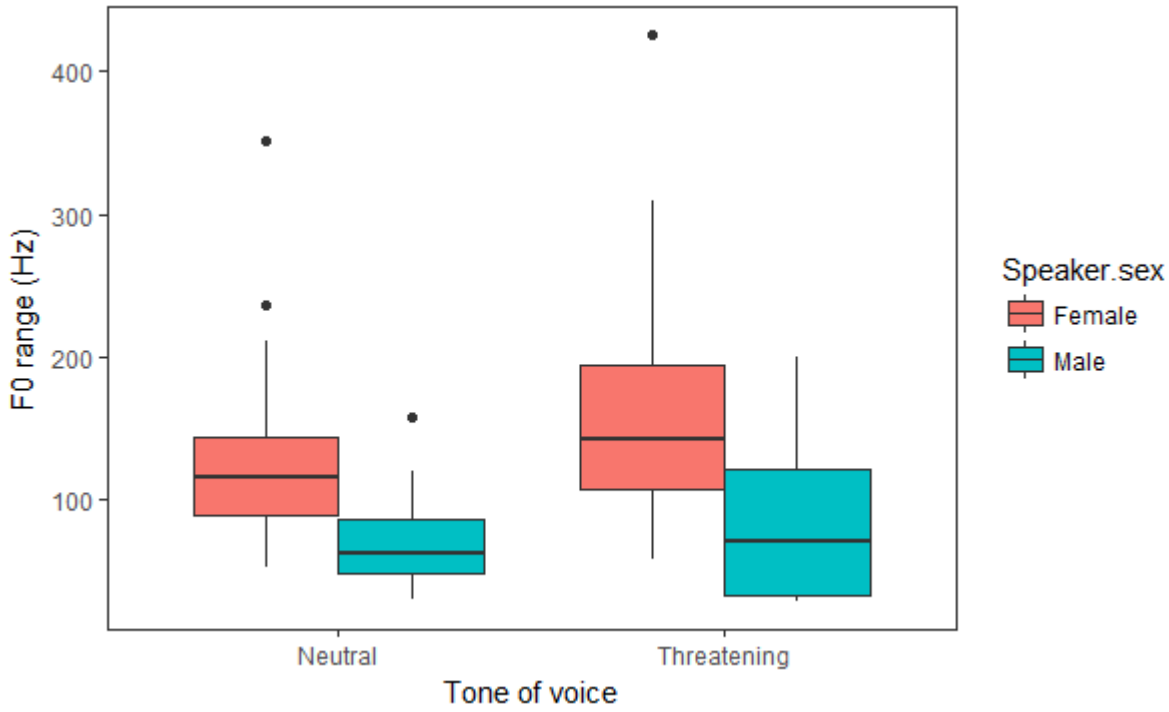


Figure 5.1 – Differences in F0 range between *threatening* and *neutral tone of voice* productions of the indirect threat utterances

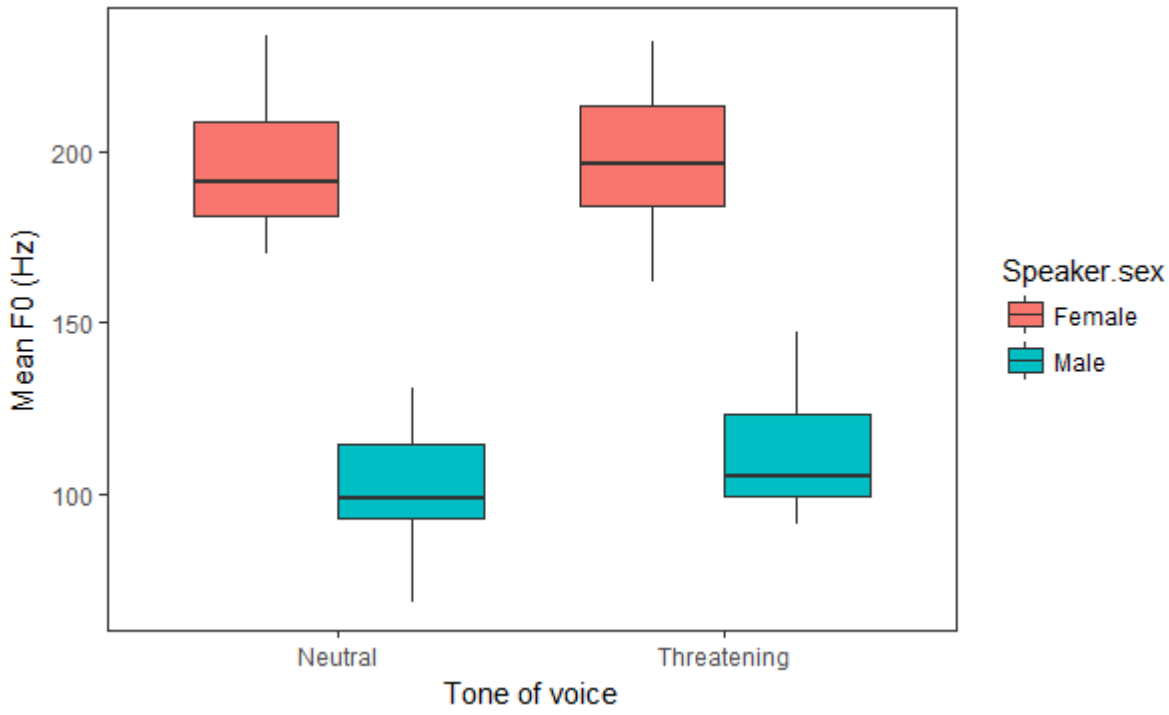


Figure 5.2 – Differences in mean F0 between *threatening* and *neutral tone of voice* productions of the indirect threat utterances

The second part of this analysis aimed to ascertain whether speakers would produce utterances in a *threatening* tone of voice with greater emphasis on the word *will*, given that previous research has linked the use of this token with increased levels of threat (Napier and Mardigian, 2003; Gales, 2010; Nini, 2017). For this analysis, separate linear mixed effects regression models were created containing the mean F0 (Hz), mean intensity (dB) and duration (ms) for the *will* tokens as the dependent variable in each, with the tone of voice (threatening/neutral) and speaker sex (male/female) forming the independent variables in all three models. As before, speaker was included as a random effect, and significance values were calculated using likelihood ratio model comparison tests. The output of this analysis is shown in Table 5.3, below.

	χ^2	Df	p
Mean F0	10.177	1	**0.001**
Mean Intensity	6.5708	1	**0.01**
Duration	18.841	1	**<0.001**

Table 5.3 – Output of lmer testing for phonetic differences between *threatening* and *neutral* realisations of *will* across the sample of speakers (significant effects displayed in **bold**)

The results in Table 5.3 illustrate a significant difference in mean F0, mean intensity and duration of the *will* tokens between speakers' *threatening* and *neutral* productions of the two indirect bomb threat stimuli. Figures 5.3, 5.4 and 5.5 further illustrate these differences, and show the trend in the data for a greater degree of phonetic emphasis to be placed on the word *will* in the *threatening* tone of voice productions with respect to the three measured variables.

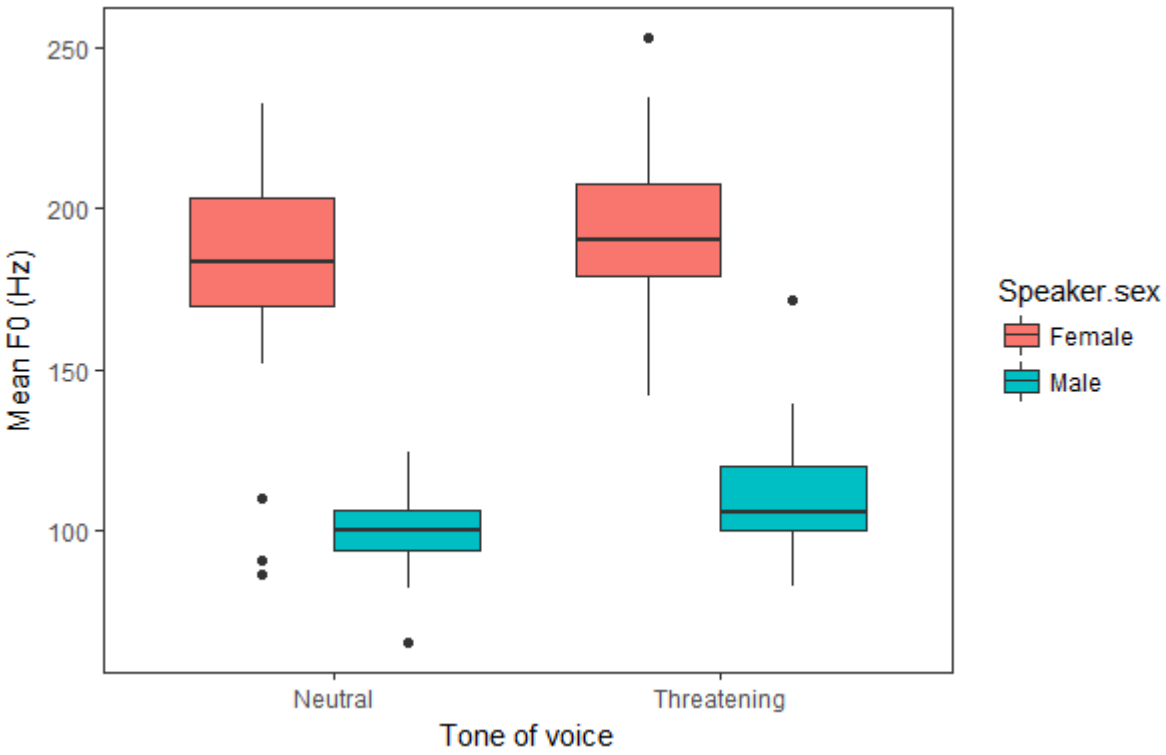


Figure 5.3 – Differences in mean F0 between *threatening* and *neutral tone of voice* productions of *will*

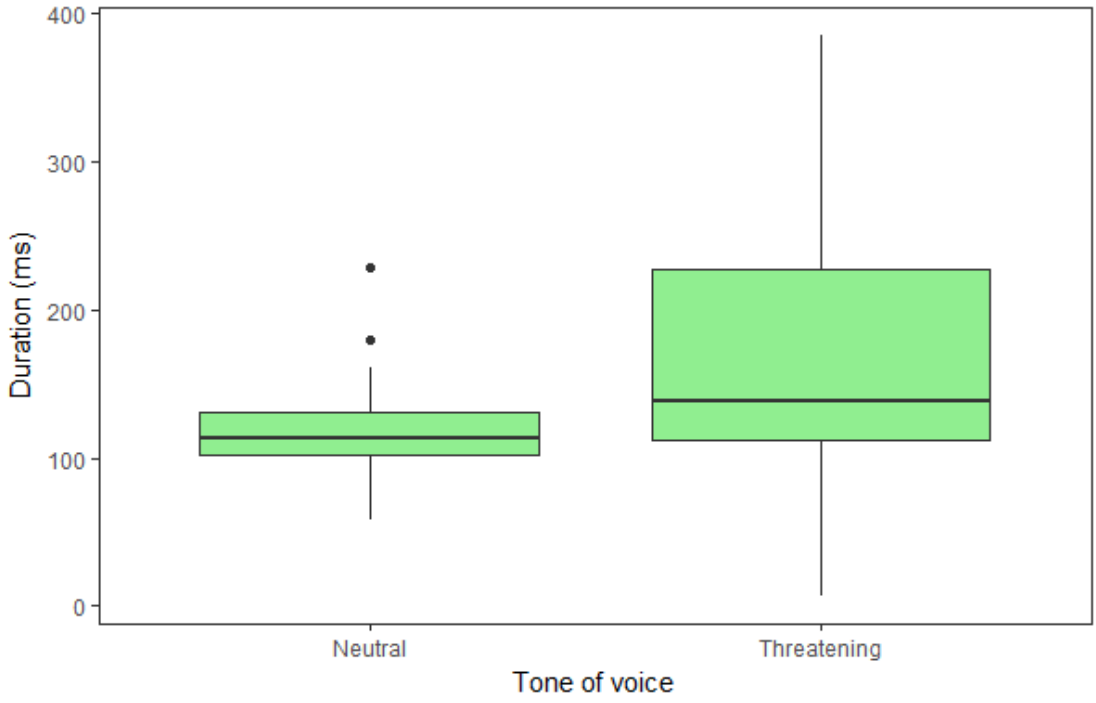


Figure 5.4 – Differences in duration between *threatening* and *neutral tone of voice* productions of *will*

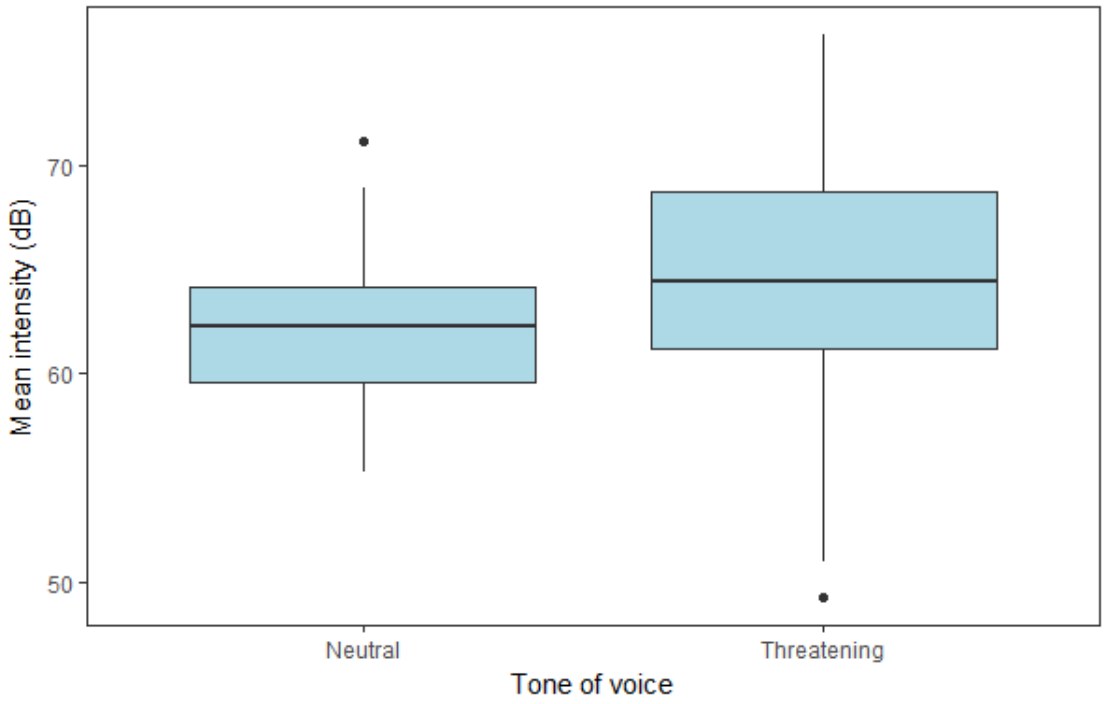


Figure 5.5 – Differences in mean intensity between *threatening* and *neutral tone of voice* productions of *will*

To further assess whether the previously described effects were confined to the word *will* or replicated for other words across the data, further testing was done to examine the differences in the phonetic prominence of other words between the *threatening* and *neutral* tone of voice productions of the two utterances used in the experiment. In this analysis, only the words that were present in both utterances were analysed, to ensure comparable testing with the *will* token. These words were *bomb*, *York*, *station*, *afternoon*, *this*, *go* and *off*. As was the case for *will*, mean F0 (Hz), mean intensity (dB) and duration (ms) measurements were taken for each word in each utterance, and these formed the dependent variable in three linear mixed effects regression models. Tone of voice (threatening/neutral) and speaker sex (male/female) were included as the independent variables in all three models. Directly replicating the previous analysis,

speaker was included as a random effect, and significance values were calculated using likelihood ratio model comparison tests. The results of this testing are shown in Table 5.4, below.

Word	**Measure**	**χ^2**	**Df**	**p**
Bomb	F0	0.0482	1	0.83
	Duration	2.3747	1	0.12
	Intensity	**4.652**	**1**	**0.03**
York	F0	0.0661	1	0.80
	Duration	0.0166	1	0.90
	Intensity	0.1933	1	0.66
Station	F0	0.1243	1	0.72
	Duration	0.1826	1	0.67
	Intensity	0.0794	1	0.78
Afternoon	F0	0.4874	1	0.49
	Duration	**10.224**	**1**	**0.001**
	Intensity	0.0444	1	0.83
This	F0	0.1337	1	0.71
	Duration	2.841	1	0.09
	Intensity	0.2866	1	0.59
Go	F0	0.1306	1	0.72
	Duration	2.7022	1	0.10
	Intensity	2.1014	1	0.15
Off	F0	0.6104	1	0.43
	Duration	3.1174	1	0.07
	Intensity	1.5863	1	0.21

Table 5.4 – Output of lmer testing for phonetic differences between *threatening* and *neutral* realisations of different words across the sample of speakers for the *"I'm warning you about a bomb at York Station which will go off this afternoon"* and

"There's a bomb at York Station. It will go off this afternoon" utterances (significant effects displayed in **bold**)

The results in Table 5.4 show that, unlike the results in Table 5.3, there were no consistent patterns which would suggest that any other words in the two tested utterances tended to be realised by speakers with a greater degree of emphasis. The output of the testing shown in Table 5.4 shows isolated significant effects for the duration of the word *afternoon* and the intensity of the word *bomb*, but these were not consistent with significantly different realisations in F0 and intensity for *afternoon*, and F0 and intensity for *bomb*. This is unlike the results for the *will* tokens, which showed a significant difference between *threatening* and *neutral tone of voice* productions across the sample of speakers for all three of F0, intensity and duration.

5.3. Listeners' perceptions of a threatening and neutral tone of voice

5.3.1. Methodology

Following the production analysis, a perceptual experiment was conducted to further investigate a potential perceptual association between phonetic emphasis on the modal verb *will* and listener inference of threat from speakers' voices. Utterances from six speakers were chosen from the larger data collection for use in the perceptual experiment. These six talkers were equally split in accordance with speaker sex (3 male, 3 female). The utterance *"I'm warning you about a bomb at York Station, which will go off this afternoon"* was used. The six speakers were selected because they all placed greater emphasis on the *will* token in their *threatening tone of voice* production of the utterance compared with their *neutral tone of voice* production. Although *will* has been

identified as a marker of increased commitment to a threat (Gales, 2010; Nini, 2017), and the production analysis in this chapter showed significant phonetic differences between realisations of *will* in the *threatening tone of voice* and *neutral tone of voice* productions, no work has yet examined whether the degree to which *will* is emphasised in spoken threats causes listeners to infer greater threat to harm in speakers' voices. The choice of speakers for this experiment allowed for an assessment of whether the interaction between a previously-identified lexical marker of increased threat and its phonetic realisation influenced listener attributions of how threatening speakers sounded. An example of the difference in emphasis across the *will* tokens in one speaker's *threatening* and *neutral* productions is illustrated in Figures 5.6 and 5.7 below. These figures illustrate that for Speaker 2 (male speaker), the *will* token in the *threatening tone of voice* production (Figure 5.6) was produced with raised F0, higher intensity and a longer duration than the *will* token in the *neutral tone of voice* realisation (Figure 5.7).

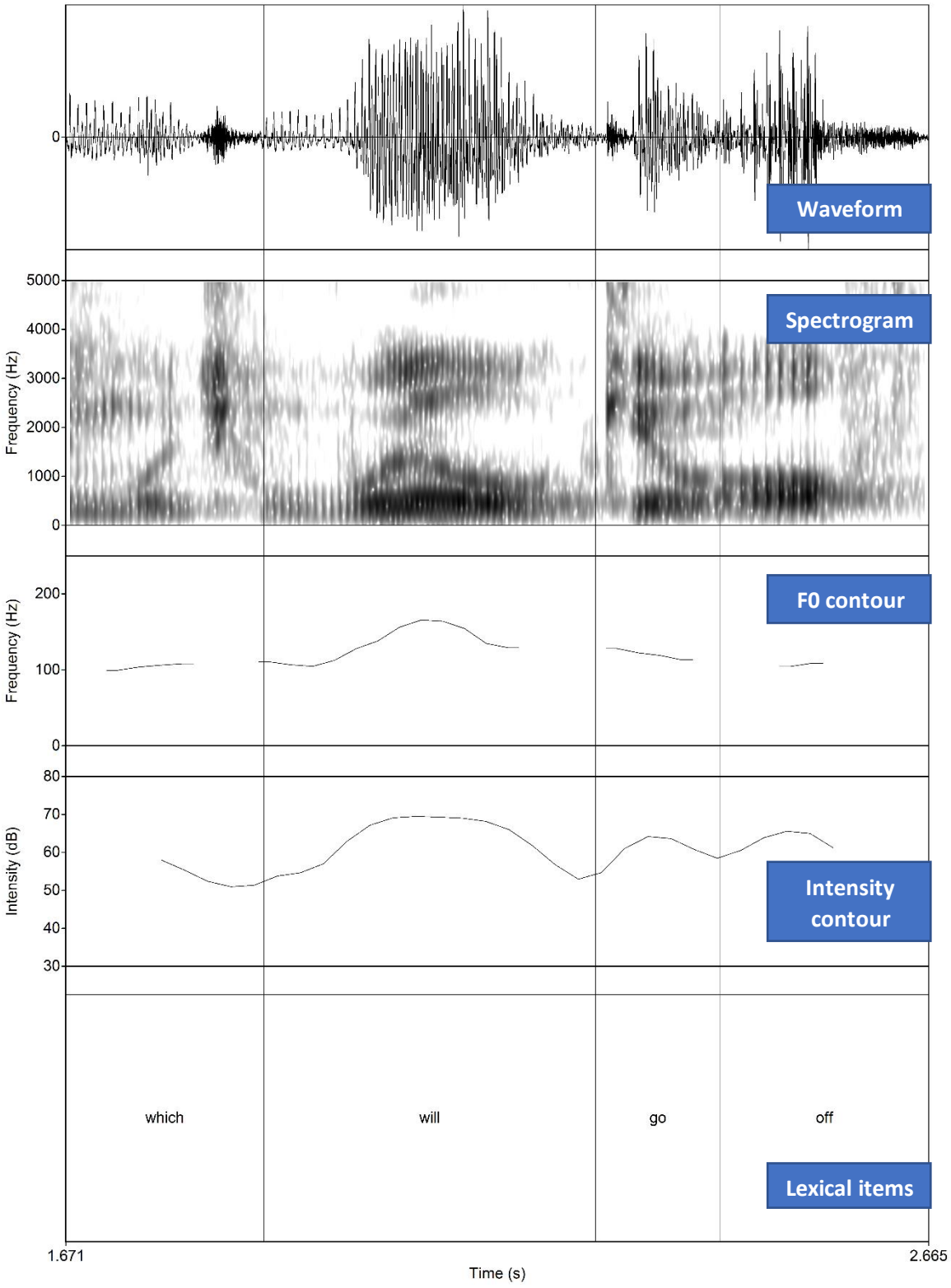


Figure 5.6 – *will* produced in the *threatening tone of voice* utterance by Speaker 2

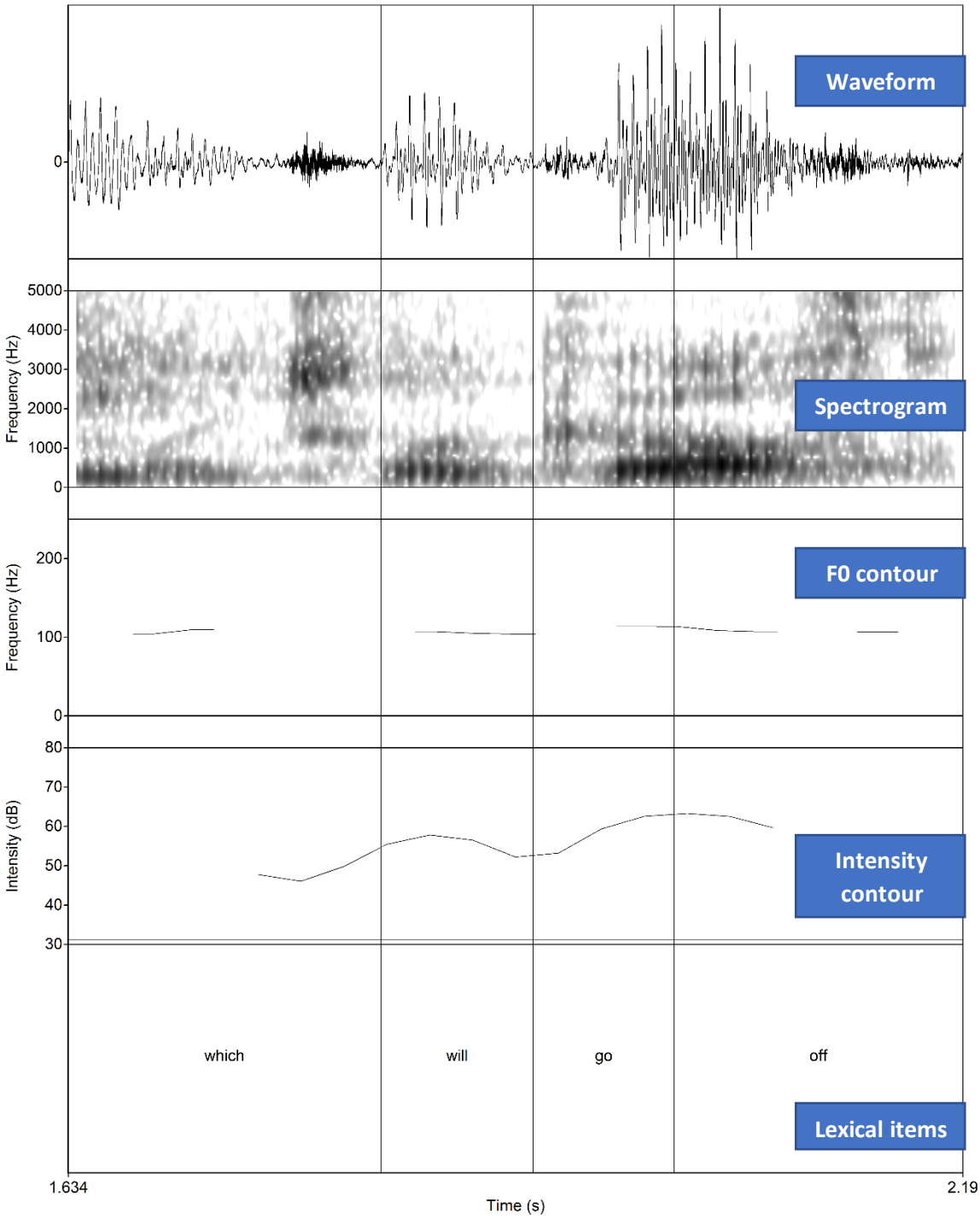


Figure 5.7 – *will* produced in the *neutral tone of voice* utterance by Speaker 2

A total of 40 participants (mean age = 20.1, age range = 18-30) provided informed consent to take part in the perception experiment. They were instructed to listen to the auditory stimuli outlined above and then answer a series of questions about the utterances they had heard. All participants were recruited from the student population at the University of York and received either payment or course credit in exchange for their participation. No participant reported any hearing impairments and all participants had normal or corrected-to-normal vision. All participants were tested in the Department of Psychology at the University of York, and were either native British English speakers (33/40) or had native-like competency in English (7/40).

Participants were instructed that they would be exposed to a series of bomb threats that had been telephoned into emergency service operators. This context was designed to provide a more forensically realistic experimental context for participants, and to mirror the context used in the experiment presented in Chapter 4. Participants were also told that in such contexts, listeners can be asked to provide information about an unknown speaker's voice. Participants were fully informed at the end of the experiment that the recordings were not real bomb threats, and were given the option to withdraw their data following the disclosure of this information. However, all listeners consented for their data to be used upon completion of the experiment. For each vocal stimulus, listeners were instructed to listen to the recording and to answer a series of questions relating to the voice they heard. As was the case for the study presented in Chapter 4, the experiment was designed to elicit information in a comparable way to the current bomb threat checklist document provided by the UK NCTSO. Listeners were asked to tick any descriptors that they felt applied to each speaker's voice from a list which included *calm, crying, clearing throat, angry, nasal, slurred, excited, stutter, disguised, slow, lisp, rapid, deep, familiar, laughter, hoarse* and *other (please state)*. Following work

conducted by Griffiths (2012) on the role of perceived familiarity in forensic speaker description tasks, participants were also asked to state if any of the voices they heard sounded familiar, and if so to say who the speaker sounded like. Listeners were instructed to tick all the boxes that they felt applied to each speaker's voice, with no upper or lower limit placed on how many descriptors could be chosen. In a comparable way to the experiment in Chapter 4, listeners were also asked to rate each voice for how *intelligent*, *threatening* and *friendly* they thought the speaker sounded using seven-point Likert-type scales. Listeners were unfamiliar with the speakers they heard, had no prior formal phonetic training, and were not provided with such training in advance of the task.

5.3.2. Results

In this study, listeners were instructed to provide ratings of how threatening they thought the speaker in each recording sounded. As the recordings presented to listeners were taken from two contrasting 'tone of voice' groups, this analysis aimed to ascertain whether listeners perceived differences between those utterances produced in a *threatening tone of voice* by speakers, and those which were produced in what the speaker considered to be a *neutral tone of voice*. To do this, statistical analysis was conducted using R (R Core Team, 2015) using random-intercept linear mixed effects regression models (hereafter `lmer`) constructed using the lme4 package (Bates et al., 2015). Main effect p-values were calculated via model comparisons, using likelihood ratio tests under the `anova` function in R. In order to test the effect of 'tone of voice' on listener threat ratings, an `lmer` model was constructed with listener threat ratings as the dependent variable, tone of voice (*threatening / neutral*) and speaker sex as independent variables, and both participant and speaker included as random effects.

These random effects were included as the experiment contained multiple speakers and multiple listeners.

The analysis in this section was conducted under the hypothesis that those utterances which had been produced in a *threatening tone of voice* by speakers would receive higher threat ratings from listeners than utterances produced in a *neutral tone of voice*. The results validated this hypothesis, showing a statistically significant difference between listener threat ratings for the *threatening* and *neutral tone of voice* utterances *($\chi^2(1)=29.72$, $p<0.001$)*. As expected, utterances spoken in a *threatening tone of voice* were rated as sounding more threatening by listeners. This effect is further illustrated in Figure 5.8, which plots listener threat ratings for the two 'tone of voice' groups. The figure shows that despite the words being the same in every utterance, the 'tone of voice' adopted by speakers significantly influenced listener threat evaluations in the expected direction.

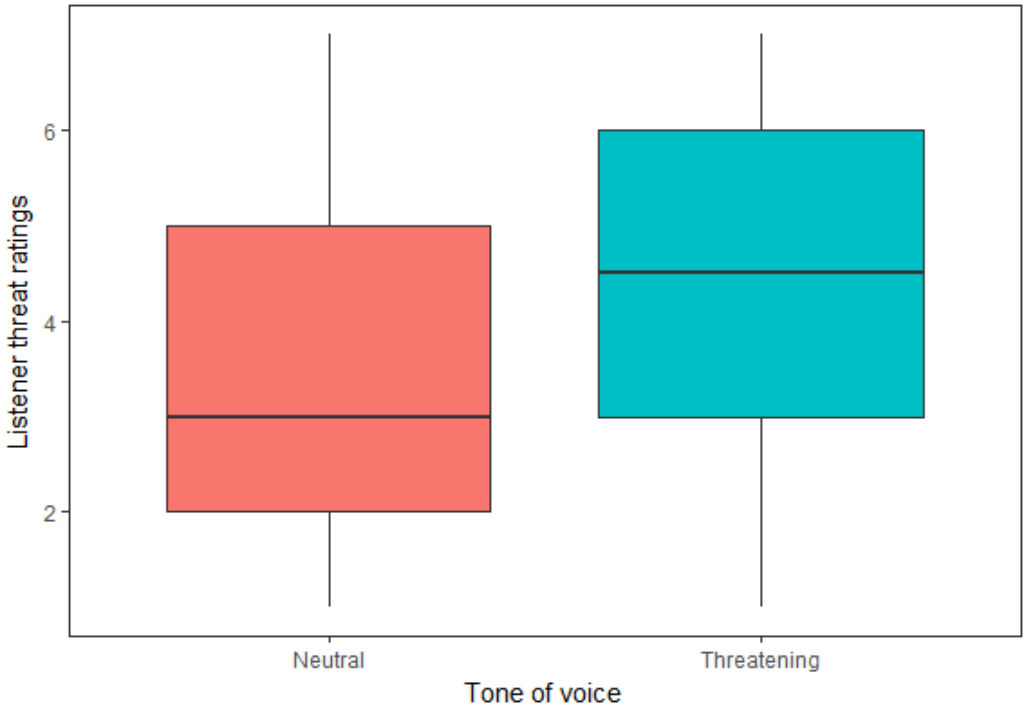


Figure 5.8 – Effect of 'tone of voice' on listener threat ratings. The plot displays raw data scores.

5.4. Listeners' descriptions of speakers' voices

5.4.1. Relationship between listeners' threat ratings and voice descriptions

Following the research presented in Chapter 4, this section presents an analysis of the relationship between listeners' descriptions of speakers' voices and the threat evaluations assigned to those voices. This facilitated further analysis of the factors which caused listeners to perceive greater or lesser levels of threat in speakers' voices. As was the case in Chapter 4, each descriptor was classified as a single variable with two variants; *yes* and *no*. Threat ratings for the *yes* responses were compared against threat ratings for the *no* responses for each descriptor. Table 5.5 lists the number of *yes* labels assigned by listeners for each of the checklist traits.

Vocal trait	No. of 'Yes' responses	Vocal trait	No. of 'Yes' responses
Angry	60	Hoarse	15
Calm	275	Laughter	3
Clearing throat	20	Lisp	7
Creaky	22	Nasal	20
Crying	3	Rapid	145
Deep	115	Slow	144
Disguised	57	Slurred	32
Excited	53	Stutter	5
Harsh	74	Whispery	10
High-pitched	64		

Table 5.5 – Number of 'yes' responses for each descriptive trait

Given the wide variety in the number of *yes* responses assigned to the different descriptors, a minimum threshold was set to ensure that descriptors which were not frequently selected were not included in the analysis. This threshold was set at 10% of the total number of times a given trait could be selected in the experiment (480), meaning that there was a minimum requirement of 48 *yes* responses for a given descriptor to be included in the analysis. This resulted in *clearing throat*, *creaky*, *crying*, *hoarse*, *laughter*, *lisp*, *nasal*, *slurred*, *stutter* and *whispery* being excluded on the grounds of not having a sufficient number of responses.

Following the analysis in Chapter 4, Section 4.3.3, t-values are reported as a statistical measure of the difference in listener threat ratings between *yes* and *no* responses for each descriptor. These were produced using the `summary()` function in R. This process follows Baayen (2008:248), who states that if the t-value in the output of a mixed effects regression model exceeds a value of 2, comparable significance at an

alpha level of 0.05 is achieved. These t-values are reported in Table 5.6. Positive values indicate that threat ratings were higher when the descriptor was selected than when it was not selected, whereas negative values indicate that threat ratings were lower when the descriptor was not selected than when it was selected.

Perceived vocal trait	Vocal trait effect (t=)
Harsh	**8.057**
Angry	**7.296**
Deep	*1.735*
Excited	*1.715*
Disguised	*1.034*
Rapid	*-1.067*
Slow	*-1.393*
High-pitched	*-1.436*
Calm	**-3.473**

Table 5.6 – Effect of perceived vocal traits on listener threat ratings (**bold** indicates a significant result)

Table 5.6 shows that voices described as sounding *harsh* and *angry* were assigned significantly higher threat ratings than voices for which those descriptors were not assigned. Conversely, voices described as sounding *calm* were assigned significantly lower threat ratings than voices for which the *calm* descriptor was not attributed. This mirrors the results in Chapter 4, with the effects illustrated in Figure 5.9, below.

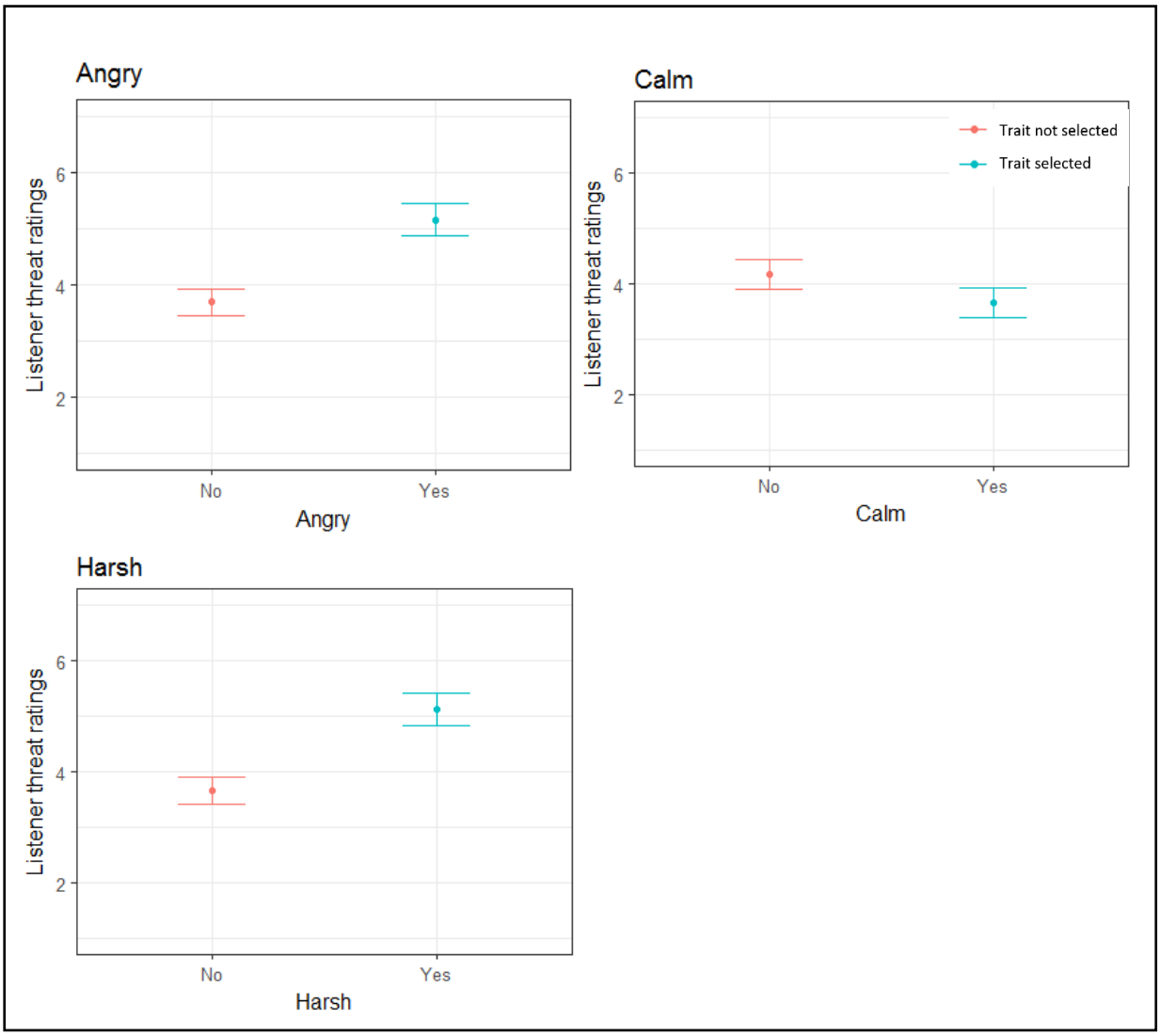


Figure 5.9 – Threat ratings assigned to voices described as sounding *angry*, *calm* and *harsh*

Table 5.4 also shows that there were no significant differences between threat ratings assigned to voices described as sounding *excited*, *disguised*, *rapid* and *slow* and voices for which those descriptors were not assigned. These findings are further illustrated in Figure 5.10, which plots the differences between threat ratings assigned to the *yes* and *no* groups for the *excited*, *disguised*, *rapid* and *slow* descriptors.

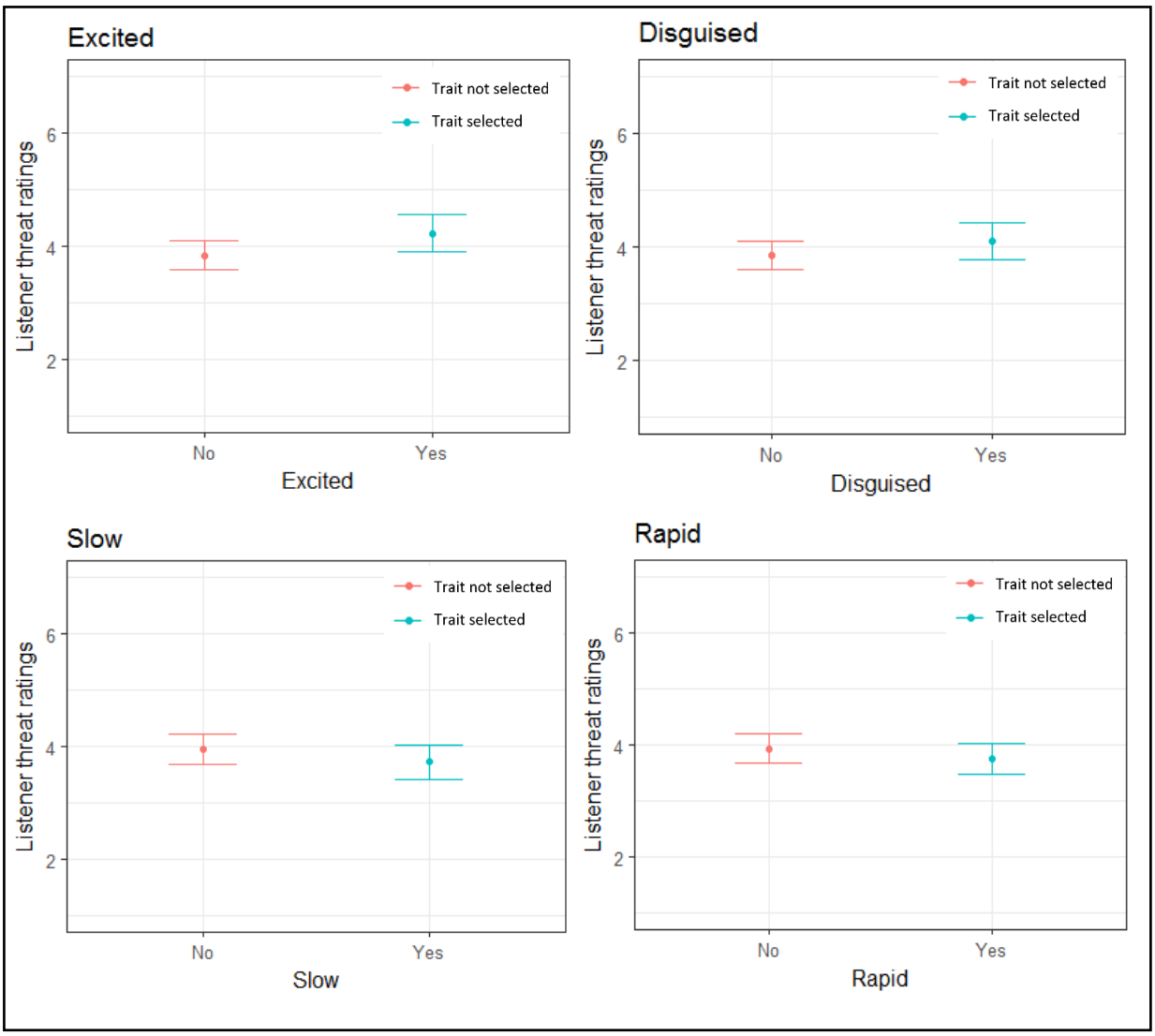


Figure 5.10 - Threat ratings assigned to voices described as sounding *excited*, *disguised*, *slow* and *rapid*

There is, however, divergence between the results displayed in Chapter 4, Section 4.3.3 and the results of this experiment with respect to the threat ratings assigned to voices described as sounding *deep* and *high-pitched*. These descriptors produced significant effects for the data in Chapter 4, but there were no significant differences in this experiment between listener threat ratings for voices described as sounding *deep* and *high-pitched* compared with those voices for which the *deep* and *high-pitched* descriptors were not assigned. This result is most likely due to the substantially greater amount of variation in mean F0 values in the data used in Chapter 4 compared with the

data in this chapter. However, the direction of the non-significant effects for voices described as *high-pitched* and *deep* are in the expected direction, given the results in Chapter 4. Figure 5.11 illustrates the differences between threat ratings for voices described as sounding *deep* and those for which the *deep* descriptor was not used, alongside threat ratings for voices described as sounding *high-pitched* and voices for which the *high-pitched* descriptor was not selected.

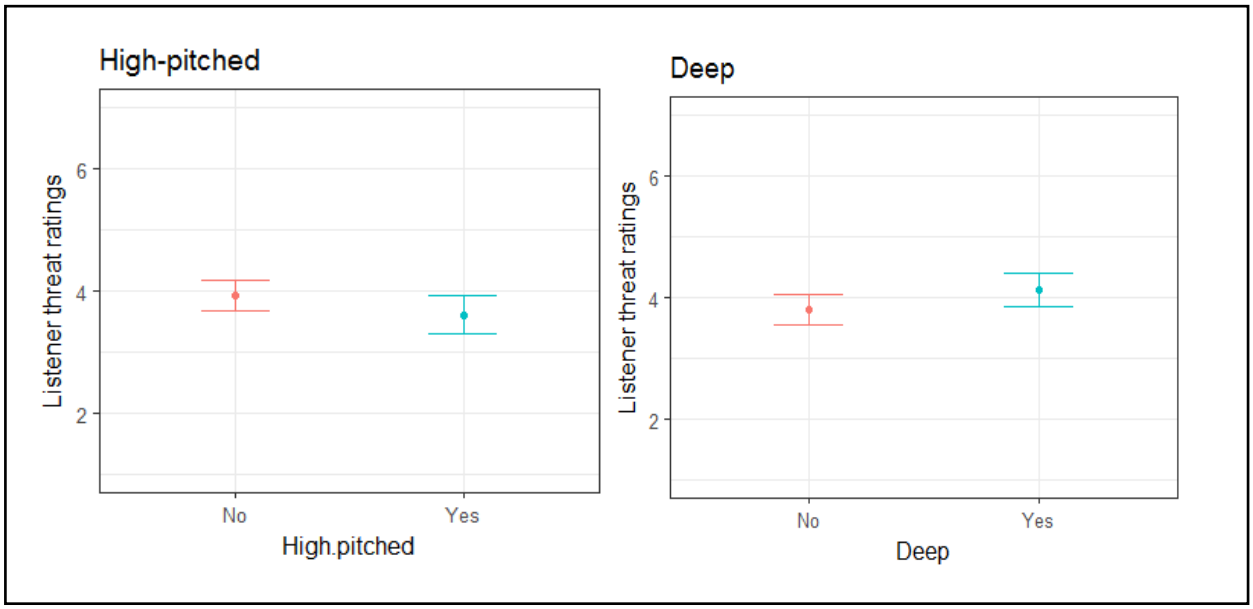


Figure 5.11 – Threat ratings assigned to voices described as sounding *deep* and *high-pitched*

5.4.2 Accuracy of listener descriptions

A further consideration for the analysis presented in this study is the accuracy with which linguistically untrained lay-listeners were able to describe particular aspects of speakers' voices. This was considered in Chapter 4 with respect to voice quality and pitch. However, as voice quality descriptors were infrequently used in this experiment, only the accuracy of pitch and speech rate attributions will be considered in this section.

5.4.2.1 Pitch perception

In this experiment, listeners were able to evaluate that a speaker's voice sounded either *deep* or *high-pitched*. They also had the option to select neither of these vocal traits when providing descriptions. Analysis of listeners' pitch attributions showed that, unsurprisingly, no listener labelled a voice as being both *high-pitched* and *deep,* reinforcing the terms' status as opposites. Table 5.7 details the number of *deep* and *high-pitched* descriptors assigned to each voice by listeners, along with the number of listeners who selected neither option.

Speaker	Speaker sex	Tone of voice	Mean F0 (Hz)	Number of descriptors assigned		
				Deep	*High-pitched*	*Neither*
3	Male	*Neutral*	96	**20**	**1**	**19**
5	Male	*Neutral*	100	**24**	**1**	**15**
3	Male	*Threatening*	103	**13**	**4**	**23**
5	Male	*Threatening*	106	**20**	**0**	**20**
2	Male	*Neutral*	125	**11**	**3**	**26**
2	Male	*Threatening*	127	**12**	**3**	**25**
1	Female	*Neutral*	176	**2**	**13**	**25**
1	Female	*Threatening*	176	**3**	**5**	**32**
4	Female	*Neutral*	180	**5**	**8**	**27**
4	Female	*Threatening*	180	**4**	**6**	**30**
6	Female	*Threatening*	184	**0**	**11**	**29**
6	Female	*Neutral*	192	**1**	**9**	**30**

Table 5.7 – Numbers of pitch-related descriptors applied to each voice (listed in order of lowest mean F0 to highest mean F0 – deep in blue, high-pitched in red and neither deep nor high pitched in green)

The values in Table 5.7 indicate several tendencies in the way that pitch descriptions were used by listeners when describing speakers' voices. Firstly, the *deep* descriptor was more commonly applied to male voices than it was to female voices. Listeners showed reluctance to use the *deep* descriptor when asked to describe a female voice, with higher usage for all six male voices than for the six female voices. The reverse also

applied with respect to the *high-pitched* descriptor. This was rarely used to describe male voices and was more readily assigned to the voices of female speakers. These results could suggest that within the experiment, listeners perceived pitch holistically rather than within biological sex categories. Rather than viewing the two sexes as distinct categories with different perceptual boundaries for pitch, listeners appeared to apply one set of perceptual boundaries to both sexes. This view would be supported by the fact that all the female voices in the experiment had mean F0 values below the reported average F0 for female speakers (approximately 200-220Hz according to Simpson, 2009), yet the *deep* descriptor was used minimally to describe them.

A further observation from the data in Table 5.7 is that listeners did, to some extent, perceive pitch for the male speakers in line with the measured Hz values. Considering the values in Table 5.5 alongside reported male-speaker F0 averages of around 120Hz (Laver, 1994; Ogden, 2009) and the reported SSBE F0 population distribution of 90-140Hz (Hudson et al., 2007), it can be seen that for those voices with mean F0 values lower than 120Hz (those of Speaker 3 and Speaker 5), the *deep* descriptor was more commonly used than it was for those recordings which had a mean F0 value of above 120Hz (Speaker 2).

The values in Table 5.7 show that throughout the experiment, many voices were described as sounding neither *deep* nor *high-pitched*. For 9 of the 12 voices, the attribution of *neither deep nor high-pitched* was the most commonly used, with only two voices (Speaker 3 *neutral tone of voice* and Speaker 5 *neutral tone of voice*) being described as *deep* more frequently than *neither deep nor high-pitched*. It can be argued that given the more regular nature of the voices used in this experiment as compared

with those in Chapter 4, the *neither* attribution is not an altogether inaccurate choice. It does, however, highlight the potential limitations of a checklist such as the one provided by the NCTSO, which is designed to elicit as much meaningful information about voices as possible from the earwitness to a given bomb threat, yet only presents extreme ends of perceptual scales.

Finally, Table 5.7 also illustrates examples where opposite terms were used by different listeners to describe the same voice. For example, Speaker 2's *neutral tone of voice* utterance was described as sounding *deep* by 11 listeners and *high-pitched* by 3 listeners. Likewise, Speaker 4's *neutral tone of voice* utterance was described as sounding *deep* by 5 listeners, with 8 stating that the same voice sounded *high-pitched*. This illustrates the differing perceptual boundaries that listeners have and the complexities in eliciting meaningful, consistent descriptions about a speaker's voice from a sample of listeners. Equally, the result highlights that just because one checklist user described a voice as having a specific vocal trait, there is no guarantee that another listener would opt to describe that same voice with the same trait.

5.4.2.2 Speech rate

A similar process to the procedure followed in the analysis in Section 5.3.4.1 was undertaken in order to analyse how accurately listeners could perceive differences in speech rate across the experimental stimuli. Within the experiment, listeners were given options to say that a voice sounded either *rapid* or *slow*, along with the option to select neither of those descriptors. Table 5.8 illustrates the number of *rapid* and *slow* descriptors assigned to each voice by listeners, along with the number of listeners who selected neither option. Although the results show that *rapid* and *slow* were treated as

opposites by the majority of listeners, both descriptors were used simultaneously on two occasions by different listeners within the experiment.

Speaker	Tone of voice	Speech rate	Number of descriptors assigned		
			Slow	Rapid	Neither
3	*Neutral*	4.9	**32**	**0**	**8**
5	*Threatening*	5.1	**14**	**2**	**24**
6	*Neutral*	5.4	**18**	**4**	**18**
4	*Neutral*	5.5	**7**	**14**	**19**
6	*Threatening*	5.5	**8**	**10**	**22**
3	*Threatening*	5.6	**14**	**11**	**15**
5	*Neutral*	5.6	**23**	**3**	**14**
2	*Threatening*	6	**3**	**22**	**15**
4	*Threatening*	6	**3**	**23**	**14**
1	*Threatening*	6.1	**8**	**18**	**14**
1	*Neutral*	6.4	**10**	**14**	**17**
2	*Neutral*	7.4	**4**	**24**	**13**

Table 5.8 – Numbers of speech rate descriptors applied to each voice (listed in order of showest speech rate to fastest speech rate – slow in blue, rapid in red and neither slow nor rapid in green)

Considering the values displayed in Table 5.8 against reported norms (see, for example, Goldman-Eisler, 1968; Gold, 2014), it can be argued that listeners perceive speech rate with a higher degree of accuracy at the more extreme ends of the scale. As detailed in Chapter 2, Section 2.5.3, typical speech rate norms have been reported as falling between 4.4 and 7.3 syllables per second (Goldman-Eisler, 1968; Gold, 2014). The range across the sample in this experiment is consistent with these reported norms, with

the slowest speech rate measured at 4.9 syllables per second and the fastest at 7.4 syllables per second. Examining the slowest stimulus in the sample in terms of the number of syllables per second (Speaker 3 *neutral tone of voice*), it can be seen that 32 listeners described this speaker's voice as sounding *slow*, with no listeners using the *rapid* descriptor. Conversely, for the fastest stimulus in terms of the number of syllables produced per second of speech (Speaker 2 *neutral tone of voice*), 24 participants described the voice as *rapid*, with only four using the *slow* descriptor. However, there were voices within the experiment for which measured speech rate and the assignment of the *slow* and *rapid* descriptors did not align. For example, Speaker 4 and Speaker 5's *neutral tone of voice* utterances differed in speech rate by just 0.1 syllables per second, yet over three times as many participants described Speaker 5's *neutral tone of voice* utterance as sounding *slow* as described Speaker 4's neutral utterance as sounding *slow*. This would suggest that other factors, possibly aspects such as pausing and perceived articulatory clarity, could influence listener perceptions of speaking tempo alongside the number of syllables per second of speech.

Additionally, the results in Table 5.8 also illustrate that for some voices, a similar number of listeners used the *slow* descriptor as used the *rapid* descriptor to describe the same voice. For example, Speaker 6's *threatening tone of voice* utterance was described as sounding *slow* by 8 listeners, but *rapid* by 10 listeners. Likewise, Speaker 3's *threatening tone of voice* utterance was described as *slow* by 14 listeners, with 11 listeners stating that the same recording sounded *rapid*. This again highlights issues with listeners' differing perceptual boundaries, and shows how differently the same voice can be evaluated by multiple listeners. The results in Table 5.8, while showing that listeners can perceive some voices accurately with respect to speech rate, also illustrates the challenge that remains in improving the consistency of the judgements

provided by linguistically-untrained listeners with respect to the perception and description of certain aspects of voice.

5.4.3. Unconstrained descriptions of speakers' voices

In addition to the checkbox options, users of the NCTSO bomb threat evaluation document are given the opportunity to provide additional unconstrained descriptions of a speaker's voice. This aspect of the checklist was incorporated into the design of the experiment in this chapter. Participants were given the option to specify additional aspects of a speaker's voice to those options on the checklist without any instruction or additional guidance provided by the researcher. From the 480 evaluations provided by listeners in the experiment, 77 (16%) unconstrained descriptions were listed. These are detailed in Table 5.9, listed alphabetically and split in accordance with the threat rating assigned to the voice for which the descriptor was provided. A number in parentheses alongside a descriptor signifies the number of times that that particular descriptor was used. As an example, four listeners described a voice as sounding 'serious' and assigned a threat rating of 2 to the same voice.

Listener threat rating	Free descriptions provided by listeners
1	Fake, Female voice, Honest, Innocent, Matter-of-fact, Professional, Serious, Slightly sexual/teasing, Sounds like reading from a script.
2	Anxious, Bored, Clear, Frightened, Hurried, Matter-of-fact, Not thought out, Rehearsed, Relaxed, Rushed, Scared, Scottish, Self-conscious, Serious (4), Slightly nervous (3), Slightly sexual/teasing, Sweet, Worried.
3	Authoritative, Certain, Clear, Concerned, Concise, Confident (4), Female voice, Forceful, Matter-of-fact.
4	Assertive, Clear (3), Controlled, Cultured, Forceful, Matter-of-fact, Measured, Monotone.
5	Arrogant, Caller has been watching too many action films!, Certain, Confident, Convincing, Monotone, Nervous, Posh (2), Scottish twang, Sly, Trustworthy.
6	Blunt, Business-like, Drawl, Echoey, Emphatic, Muffled, Playful, Serious (2), Smug, Urgent.
7	Assertive, Threatening (4)

Table 5.9 – Free descriptions of voices provided by listeners

Unsurprisingly, the unconstrained descriptions generally indicated that the participants in this experiment did not use the same language to describe voices as would be

expected from trained linguists or phoneticians. This could be partly due to the task having already elicited vocal descriptions in the form of the checklist traits, alongside a reflection of Shuy's (1993) assertion that non-linguist earwitnesses lack the vocabulary to describe other speakers' language. However, following Griffiths' (2012:260) argument, which states that "identifying and engaging with non-linguists' own perceptual scales should inform practice amongst forensic linguists and legal professionals", it can be argued that eliciting such descriptions and examining how speakers describe voices could nevertheless prove useful in gaining a greater understanding of how phonetically untrained listeners perceive speakers' voices.

The unconstrained descriptions provided by listeners revealed several themes. Certain descriptors relate to either a linguistic aspect of the talker's speech or the acoustic quality of the recording, despite the previously-highlighted issue that lay-listeners do not use comparable terms to those which would be expected of a linguist or phonetician. Examples include the perception that the speech in question sounded *muffled*, that the sample was *echoey*, that the talker's speech was *monotone*, that the voice was a *female voice*, and certain, albeit at times inaccurate, descriptions of speaker accent such as *Scottish twang*. There are also examples of descriptors which, although not directly linguistically related, can be traced to specific aspects of speakers' voices. For example, the description of a caller sounding *posh* could relate to the perception of a SSBE or RP accent, and the comment that the caller *"sounds like they are reading from a script*" could suggest a slower speech rate and pausing patterns which are more consistent with the perception of read speech than spontaneous speech. *Clear* could relate to articulatory precision, while *forceful* could correspond to increased loudness, and therefore higher F0 and intensity.

However, there are a number of potentially problematic descriptors within the data in Table 5.9 which relate to aspects of the speaker's personality or emotional state. Several of these descriptors also appear to relate to the listener's level of perceived threat. For example, descriptors such as *fake*, *honest*, and *innocent* are used to describe voices which were assigned a low threat rating of 1, whereas voices which were perceived as conveying the highest level of threat (threat rating of 7) were described using terms such as *threatening* and *assertive*. Although not a uniform pattern, this tendency further illustrates the potential lack of separation between listeners' evaluations of a given speaker and descriptions of that same speaker's voice. Furthermore, descriptors such as *innocent, serious, sweet, confident, certain, convincing, blunt, playful* and *smug* are difficult to link to specific linguistic or phonetic qualities, yet were provided by listeners via the format of a checklist which is designed to elicit meaningful descriptions of a talker's voice.

5.4.4. Listener familiarity judgements

Following the work of Griffiths (2012), whose research on forensic speaker description tested the idea of having listeners provide 'soundalikes' for voices they were asked to describe, this study asked listeners to answer the question, *"Does this voice sound familiar (e.g. like a family member, TV personality, actor, sportsperson)? If so, who does it sound like?"*. From a total of 480 voice evaluations, only 18 responses to this question were provided. Within this small number of responses, three broad categories emerged; public figure, personal connection and character/occupational stereotype. Table 5.10 displays these 18 responses, split in accordance with the category with which they were subsequently assigned to. However, the question seemed to yield little in the way of helpful descriptors, and the small number of responses somewhat limits further meaningful analysis of this data. In order to help listeners answer this question, the

instructions gave hints as to the kinds of people that could be described. These were *family member, TV personality, actor,* and *sportsperson*. Notably, the descriptions provided by the listeners in this experiment seem to have been heavily influenced by the wording of the question. When a specific person was named, they were always either an actor (e.g. Daniel Craig), a TV personality (e.g. Philip Schofield) or a sportsperson (e.g. Jessica Ennis-Hill). It would therefore appear as though listeners were basing their descriptions quite heavily on the hints given within the question.

Public figure	**Personal connection**	**Character/Occupational stereotype**
Thomas Brodie-Sangster	*Family member*	*Reporter*
Ben Whishaw	*My lecturer*	*Policeman*
John Oliver	*My friend*	*Actor*
Daniel Craig	*Family member*	*Newsreader*
Jessica Ennis-Hill		*Film spy character*
Iwan Rhoen		*News reporter*
Philip Schofield		*Ramsey Bolton*

Table 5.10 – Listener responses to question, "*Does this voice sound familiar (e.g. like a family member, TV personality, actor, sportsperson? If so, who does it sound like?*

5.5. Discussion

The goal of the current experiment was to build on the previous work presented in Chapters 3 and 4 by further analysing how listeners perceive threat in the voices of multiple unfamiliar speakers. It also aimed to combine perception and production analysis by ascertaining whether a link existed between speakers' *neutral tone of voice* and *threatening tone of voice* productions of an indirect bomb threat utterance with respect to the level of emphasis on the modal verb *will*. The analysis in this chapter further contributes to the work of Napier and Mardigian (2003), Gales (2010) and Nini (2017), who all highlighted how the use of modal *will* can act as a marker of increased threat within a potentially threatening utterance. The production analysis presented in Section 5.2 showed a trend for listeners to produce *will* tokens with greater phonetic emphasis in the *threatening tone of voice* utterances than the *neutral tone of voice* utterances with respect to the acoustic phonetic variables of F0, intensity and duration. The analysis in Section 5.2 showed significant differences between *threatening* and *neutral* tone of voice productions of the utterances *"There's a bomb at York Station. It will go off this afternoon"* and *"I'm warning you about a bomb at York Station which will go off this afternoon"* with respect to mean F0, duration and mean intensity across the *will* token. It can therefore be argued that when attempting to make their utterances sound more threatening, there was a significant trend in the data for speakers to place greater phonetic prominence on the word *will*. This trend was not replicated on other words which were shared between the two utterances used as experimental stimuli, which further highlights the relative uniqueness of the *will* tokens in these realisations.

With respect to the perception experiment, the results obtained from the study presented in this chapter help to further illustrate that even when contentful indirect threat

utterances are presented to listeners, perceived characteristics of the speakers' voices can still play a role in shaping the inference of how threatening speakers sounded. The significant effect for 'tone of voice' in the perception experiment further develops the results of the experiments presented in Chapters 3 and 4, and adds further weight to Milburn and Watman's (1981) conclusion that the way in which a potentially threatening utterance is produced can be just as, and possibly more, important than the words spoken. Had the speakers' 'tone of voice' not influenced listeners' judgements, then no difference in threat ratings would have been present between the two 'tone of voice' utterance groups within the experiment, given that the wording of the utterances was the same in both 'tone of voice' groups. Furthermore, as the stimuli presented to listeners were designed to present differences in lexical prominence on the *will* tokens between speakers' *neutral* and *threatening* productions, it can be argued that lexical prominence on the word *will* also influenced listeners' perceptions of threat within the experiment, in addition to speakers' productions. This result highlights a potential synergy between the production and perception of threat within the sample of speakers and listeners in the experiment.

Furthermore, the work presented in Section 5.4.1 was designed to further develop research exploring the relationship between listeners' own perceptual boundaries for aspects of voice and the perception of threat in the voices of unfamiliar speakers. The results showed that threat ratings significantly differed between fewer of the *yes*/*no* voice descriptor categories compared with the experiment in Chapter 4. However, the previously-described problems with the use of *angry* and *calm* as descriptors of aspects of threateners' voices (Chapter 4, Section 4.4) were again highlighted by the results in this chapter. Voices described as sounding *angry* were assigned significantly higher threat ratings than voices for which the *angry* descriptor was not used, and the opposite

pattern was shown for the *calm* descriptor. Collectively, these results further reveal potentially unwanted biases in listeners' judgements of threatening utterances produced by anonymous or unfamiliar speakers.

The analysis in Chapter 4 raised a potential ambiguity regarding the term *harsh*, which has a specific meaning relating to voice quality, but also a more general, non-technical meaning attached to the characteristics of a speaker's personality. Given that the experiment presented in Chapter 4 used stimuli that were produced in a phonetically *harsh* voice quality, it was unclear which of the two interpretations listeners were using when providing descriptions and evaluations. However, the results in this chapter would suggest that phonetically-untrained listeners are more likely to draw on the non-technical definition of the term *harsh* when describing and evaluating aspects of voice. While other voice quality descriptors such as *whispery*, *creaky* and *hoarse* were used substantially more infrequently in this experiment when compared with the experiment presented in Chapter 5, *harsh* still remained a frequently-used term which aligned with *angry* both in terms of frequency of use and the association between perceived harshness and increased threat inference.

A further strand of analysis in this chapter concerned the degree to which listeners could accurately perceive pitch and speech rate, compared with quantitative measurements of the two variables. This analysis highlighted certain complexities surrounding listeners' perceptual boundaries for specific aspects of voice. It showed that while some voices appeared to be identified accurately with respect to pitch and speech rate, others were not, with different listeners sometimes using opposite terms, such as *deep* and *high-pitched*, to describe the same voice. The experiment in this chapter further highlights the

need for more research to be carried out in order to improve methods for eliciting more consistent descriptions of phonetic aspects of voice from phonetically untrained listeners.

The analysis in this chapter further progresses the analysis and research presented in the previous two chapters of this thesis by obtaining listeners' evaluations of multiple speakers rather than either one speaker or multiple guises produced by a single speaker. It also integrated analysis of speakers' productions alongside listeners' perceptions of a *neutral tone of voice* and a *threatening tone of voice*, which serves as a development of the work presented in Chapters 3 and 4. However, the focus of this work has predominantly centred around one variable – the perception and production of emphasis on the word *will*. The goal of the work in the following chapter aims to further develop the findings of this experiment and those in the previous chapter by integrating more phonetic and linguistic variables into an experimental framework which contains multiple speakers and non-manipulated stimuli.

Chapter 6 – Towards a more integrated assessment of threat perception

6.1 Introduction

The research presented in this chapter builds on both the previous methodologies and findings from Chapters 3, 4 and 5. In doing so, the purpose of the research in this chapter is to provide a more integrated assessment of the relationship between phonetic and linguistic aspects of speakers' utterances and listeners' evaluations of threat and intent to harm, alongside more general characteristics of the voices they are exposed to. To summarise the study, a group of listeners were played audio recordings of speakers producing simulated bomb threats. The study investigated the influence of a range of linguistic and phonetic variables on listeners' evaluations of how threatening these speakers sounded and how much intent to harm was conveyed through their speech. Of course, there was no actual intent to harm conveyed in any of the utterances, so the focus of the experiment was to assess listeners' perceptions of speaker intent to harm. The goal of the analysis was to gain further understanding of the motivations behind listeners' decisions about what makes a speaker sound more or less threatening, and to assess the roles that different aspects of voice play in shaping those decisions. This was done using an integrated design testing regular vocal variation rather than manipulating particular phonetic aspects of voice and testing the effect of those manipulations on listeners' evaluative judgements.

Additionally, following the work of Griffiths (2012), Watt and Burns (2012), Dixon, Foulkes and LaShell (2013), Smith et al. (2018) and the research presented in Chapters 3, 4 and 5 of this thesis, the study discussed in this chapter also further assesses how accurate listeners are at making certain judgements about specific aspects of speakers' voices. The research presented focuses on two particular aspects of voice, *average pitch*

and *regional accent*, as well as unconstrained vocal descriptions. This was conducted in order to examine whether listeners' judgements of the high-pitchedness of a speaker's voice would align with acoustic measurements of average Fundamental Frequency (F0), the key acoustic correlate of vocal pitch (Laver, 1994). Furthermore, the work also builds on Griffiths' (2012) work by eliciting and examining listeners' descriptions of the regional accents of speakers of three different varieties of English. These two aspects of voice were chosen owing to the fact that neither parameter relies on lay-listeners' ability to interpret specialised linguistic or phonetic terminology, and on account of both types of voice characteristics being captured by earwitness evidence documents such as the NCTSO bomb threat evaluation checklist. One overarching goal of the work presented in this thesis is to provide empirically-driven recommendations for improving the way in which real-world spoken threats are handled and evaluated by those working at different stages of the legal process, from earwitnesses and juror to police officers and lawyers. The work in this chapter further assesses the suitability of documents such as the NCTSO bomb threat checklist to obtain reliable evidence about speakers' voices for use in investigative or evidential purposes.

Finally, the work in this chapter further extends the research presented in Chapter 5 by exploring the relationship between listeners' evaluations of threat and intent to harm, and the descriptions that listeners provide of speakers' voices. In this experiment, listeners were asked to assess aspects of speakers' body size in order to further test the hypothesis that lower perceived pitch would link to listeners' perceptions of a larger speaker, as was the case for the data in Chapters 3 and 4. The experiment also elicited unconstrained descriptions of speakers' voices from listeners, and used the EmoLex sentiment lexicon (Mohammad and Turney, 2013a; 2013b) to examine whether the emotional content of the voice descriptions provided by listeners was in some way

shaped by the fact that the verbal content of the utterances were designed to represent indirect bomb threats.

6.2 Methodology

6.2.1 Stimuli

The experimental stimuli and experimental design in this study was similar to the design of the study presented in Chapter 5. The experimental stimuli were comprised of 48 voice recordings produced by 12 student volunteers (6 male, age range = 18-30). Speakers provided informed consent to be recorded producing the utterances *"There's a bomb at York Station. It will go off this afternoon"* and *"I'm warning you about a bomb at York Station, which will go off this afternoon"*. As previously explained in Chapter 5, the stimuli were constructed using commonly-found features in real-world threats (Napier and Mardigian, 2003; Gales, 2010; Nini, 2017). These included the use of indirect threat utterances (in this case utterances which had alternative interpretations as either a warning or a statement of fact), talk of a violent act (in this case the detonation of a bomb) and utterances in which the violent act was directed towards a third party rather than the direct recipient of the utterance. Speakers were instructed to produce each utterance twice, once with additional emphasis on the word *will* and once with additional emphasis on the word *this*. The aim of this difference was to further assess whether placing utterance-level emphasis on the modal verb *will* would prompt listeners to deem the utterance to sound more threatening, given previous research linking the use of *will* in a threat with increased certainty and commitment towards the threatened act (Gales, 2010; Nini, 2017). This also further develops the research in Chapter 5 which highlighted emphasis on *will* as a potential marker of perceived threat.

Experimental stimuli were based on indirect threat utterances that could also have been interpreted as other types of speech acts, in this case warnings, promises and statements of fact. This was done on account of the notion that indirect threats are more problematic for legal professionals on account of their lexical and pragmatic ambiguity, and following Gingiss' (1986:155) assertion that "the problem of indirect threats is one that the courts must deal with". In order to assert some control over context, the experiment again was based around a real-world scenario in which there is the potential for the evaluation of how threatening unfamiliar speakers sounded, namely the evaluation of emergency service calls involving indirect bomb threats. This provided a framework which was general enough to draw meaningful conclusions about the perception of threat and intent to harm from vocal and facial stimuli, whist also retaining some contextual control over the experiments.

Recordings were conducted in a quiet recording environment using a Zoom H4N handheld recorder with the microphone placed on a table approximately 30cm from each speaker. Among the group of speakers, four were self-identified speakers of Standard Southern British English (SSBE), four were self-identified speakers of Northern Irish English, and four were self-identified L2 speakers of English having 'Middle Eastern' languages as an L1 (three Arabic speakers, one Persian speaker).[2] Each accent group contained an equal number of male and female speakers. These accents were picked so as to enable a comparison between a standard variety of English, a non-standard variety of English, and a foreign-accented English. It was anticipated that the SSBE accent would be rated more neutrally compared with the Northern Irish and Middle Eastern accents with respect to the inference of threat and intent to harm,

[2] Arabic and Persian are of course languages with highly distinct phonologies, but the view is taken that in the present context the differences in the way these participants speak English are not large enough to create significant disparities in terms of the listeners' evaluations of the speakers' accents.

given the potential for cultural associations and stereotyping of speakers with both Northern Irish and Middle Eastern-sounding accents in relation to bomb attacks. Rather than adopting a matched guise design, as has been used in other research on accent evaluation in legally-relevant research (see, for example, Dixon, Mahoney and Cocks, 2002; Dixon and Mahoney, 2004), authentic speakers of each of the tested accents were used in an attempt to more accurately simulate real-world voice evaluation situations. Once recorded, the acoustic stimuli were band-pass filtered between 300 and 3400Hz to replicate the landline telephone channel (Künzel, 2001; Nolan et al., 2013). A 0.5-second period of silence was added to the end of each utterance, and this was followed by a one-second long 175Hz tone which was designed so as to resemble the hangup tone signalling the termination of a call.

The study investigated the relative influence of a range of linguistic and phonetic parameters on listeners' threat and intent to harm evaluations. These included median F0 as an average measure of how high-pitched a speaker's voice was; F0 range, as a measure of how much intonational variation was present in each utterance; speaker accent (SSBE, Northern Irish, Middle Eastern); emphasis pattern (emphasis on 'will' / emphasis on 'this'); utterance ("I'm warning you about a bomb…" / "There's a bomb…"). For male speakers, the pitch range in Praat (Boersma & Weenink, 2016) was set between 75Hz and 300Hz, whereas for female speakers the range was set at 100-500Hz. The median F0 measurements for each voice were extracted using the ProsodyPro script (Xu, 2013) in Praat (Boersma and Weenink, 2016), with pitch trace errors being manually corrected before the script was used.

In addition to examining the influence of measured phonetic variables, the study also assessed the influence of *perceived pitch* and *perceived speed* on listeners' threat and

intent evaluations. To obtain these measures, participants were asked to assess how high-pitched and fast each speaker's voice sounded on a scale from 0 (*very low-pitched / very slow*) to 100 (*very high-pitched / very fast*). This reflects the procedures that exist for eliciting information from earwitnesses in forensic settings, which often ask for information about an offender's voice. These procedures often make use of documents specifically relating to the evaluations of spoken threat utterances, such as the UK National Counter Terrorism Security Office bomb threat checklist (National Counter Terrorism Security Office, 2016).

6.2.3 Participants and procedure

A total of 85 participants (mean age = 20, age range = 18-55) took part in a perception experiment, during which they were tasked with evaluating the experimental stimuli. All participants were students at the University of York and received payment or course credits for their participation. Informed consent was provided prior to participation. All participants were tested in the Department of Psychology at the University of York, and were native British English speakers. Participants were provided with closed-cup headphones in a quiet environment and were instructed to listen to each voice and to answer a series of questions about the speaker they heard. In order to provide a forensically relevant context to the experiment, participants were told that the recordings they would hear were from calls made to emergency service operators. However, it is acknowledged that there is the potential for the repetitive nature of the experiment to have caused some listeners to believe that the stimuli were simulated. Participants were asked to assess how threatening each speaker sounded, and how much intent to harm was conveyed by each talker's speech using a 0 (*not-at-all threatening/no intent to harm*) to 100 (*extremely threatening/certain intent to harm*) scale. So as to

avoid excessive repetition, individual listeners were presented with a subset of the complete set of voices to evaluate. The effect of order was further minimised by using a computer-generated randomised order for each listener, meaning that no two listeners heard the same subset of voices in the same order. The mean number of voices evaluated per listener was 11, and the mean number of times each utterance was evaluated was 20. Participants were also free to listen to each recording as many times as they wished to, and the evaluations took place immediately after exposure to each stimulus so as to avoid memory to be a factor in the evaluation process.

Additionally, listeners were instructed to say how high-pitched they thought the voice of each of the speakers they heard was using a scale ranging from 0-100, where 0 represented 'very low-pitched' and 100 represented 'very high-pitched'. Listeners were also instructed to say what accent they thought each speaker had. This was done using an open-answer format, in response to the question *"What accent do you think this speaker has? Leave the box blank if you are unsure"*. In order to assess the relationship between evaluations of perceived pitch and perceived body size, listeners were instructed to evaluate speakers' body size based on information contained in speakers' voices. As was the case in Chapter 5, these evaluations were made by selecting from a list which included *large build, small build, average build, slim, stocky, very tall, taller than average, average height, shorter than average* and *very short*. Additionally, listeners were given an open-ended response box and told to comment on any other notable aspects of the speakers' voices. These descriptions were unconstrained and listeners were free to say anything they wished to about the voice of each speaker they heard. Information was also collected about how similar listeners thought their own accents were to a range of different UK accents. This list included *Oxford/Cambridge* (designed to reflect SSBE), *Newcastle, Yorkshire, Manchester, Liverpool, Belfast* and

Glasgow. This information was collected in order to assess perceived similarity between listeners' accents and two of the target varieties in the experiment (SSBE and Northern Irish English). The other accents were included as distractors so as not to focus listeners' attention entirely on the target varieties. Finally, given that the research was not concerned with listeners' abilities to remember speakers from aspects of the speakers' voices, the evaluation of each voice sample took place immediately after exposure to that sample.

6.3 Results

6.3.1 Relationship between listener evaluations of threat and intent to harm

Following the finding in Watt, Kelly and Llamas (2013) that listener ratings of how threatening speakers sounded were closely correlated with the inferred level of intent to harm, it was expected that a comparable result would be evident in these data. Figure 6.1 plots the correlation between listener ratings for how threatening speakers sounded with the perceived level of intent to harm. As expected, the ratings were strongly correlated (Pearson's $r = 0.77$, df = 973, $p<0.001$), although the r^2 value of 0.59 shows that the two variables were not as closely correlated as might have been expected, with a general trend for the intent to harm judgement to be higher on the rating scale than the comparable threat judgement.

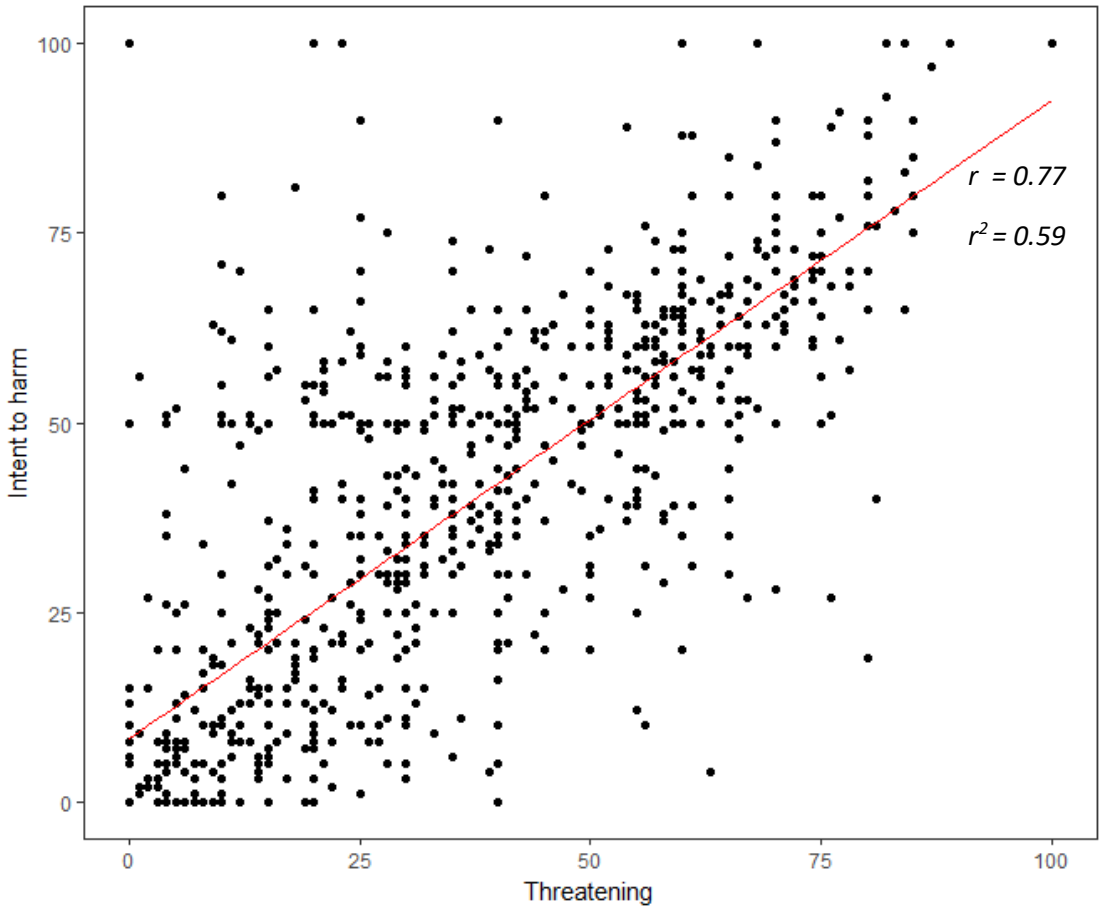


Figure 6.1 - Relationship between listeners' evaluations of threat and intent to harm

Watt, Kelly and Llamas (2013) found that when listeners rated productions of *"I know where you live"* for both threat and intent to harm, 73.9% of threat ratings were equal to, or higher than, ratings for intent to harm. In this experiment, however, 37% (n=369) of threat ratings were higher than the comparable intent to harm rating, 44% (n=435) of intent to harm ratings were higher than the comparable threat rating, and 19% (n=184) of threat and intent ratings for the same stimulus were equal to each other. This could reflect the more indirect nature of an utterance such as *"I know where you live"* (used by Watt, Kelly and Llamas (2013)) compared to *"I'm warning you about/there's a bomb at York Station which will go off this afternoon"*. It could be argued that listeners were potentially more unwilling to dismiss a speaker producing an utterance discussing

a bomb being detonated at a certain place and time as communicating either little or no intent to harm.

6.3.2 Effects of linguistic and phonetic variables on listeners' threat and intent to harm judgements

As was the case in previous chapters, statistical analysis probing the effects of the chosen phonetic and linguistic variables on listener judgements of how threatening speakers sounded was conducted using R (R Core Team, 2015) using random-intercept linear mixed effects regression models (hereafter lmer) constructed using the *lme4* package (Bates et al., 2015). Main effect *p*-values were calculated via likelihood ratio model comparisons tests, using the *anova* function in R. This method was used to assess the influence of the tested aspects of voice on listeners' assessments of how threatening speakers sounded, and how much intent to harm they were judged to convey. In the first model, listener threat ratings formed the dependent variable, with *median F0*, *F0 range*, *speaker accent*, *utterance*, *emphasis pattern*, *perceived pitch*, *perceived speed* and *speaker sex* included as fixed effect predictor variables. Given that the experiment involved multiple speakers and multiple listeners, *listener* and *speaker* were also included as random effects. In the second model, the same fixed and random effects were used, with listener intent to harm ratings forming the dependent variable. Table 6.1 displays the output of the both lmer models.

	Threat evaluations			Intent to harm evaluations		
	χ^2	Df	p	χ^2	df	p
Perceived pitch	**22.98**	**1**	**<0.001**	**13.48**	**1**	**<0.001**
Perceived speed	2.86	1	0.09	3.21	1	0.07
Emphasis pattern	3.36	1	0.07	3.36	1	0.11
F0 range	2.23	1	0.14	2.66	1	0.10
Speaker sex	0.22	1	0.64	1.67	1	0.20
Speaker accent	3.58	2	0.17	1.59	2	0.45
Utterance	0.53	1	0.47	0.03	1	0.87
Median F0	0.40	1	0.53	0.001	1	0.99

Table 6.1 - Effects of tested variables on listener evaluations of threat and intent to harm. Significant effects are displayed in **bold**.

For listeners' perceptions of how threatening speakers sounded, the results in Table 1 highlights a significant effect of perceived pitch on listeners' perceptions of both how threatening speakers sounded and how much intent to harm was conveyed through their speech. No other variable had a significant effect on listeners' judgements of either threat or intent to harm. The relationship between listeners' judgements of perceived pitch and threat ratings is plotted in Figure 6.2. Due to the pre-existing sex differences in vocal pitch, this was done separately for male and female speakers. The plots reveal a trend for higher-pitched voices to be judged as sounding less threatening and intentful compared with lower-pitched voices. Additionally, the effect is more prominent for the male speakers than for the female speakers.

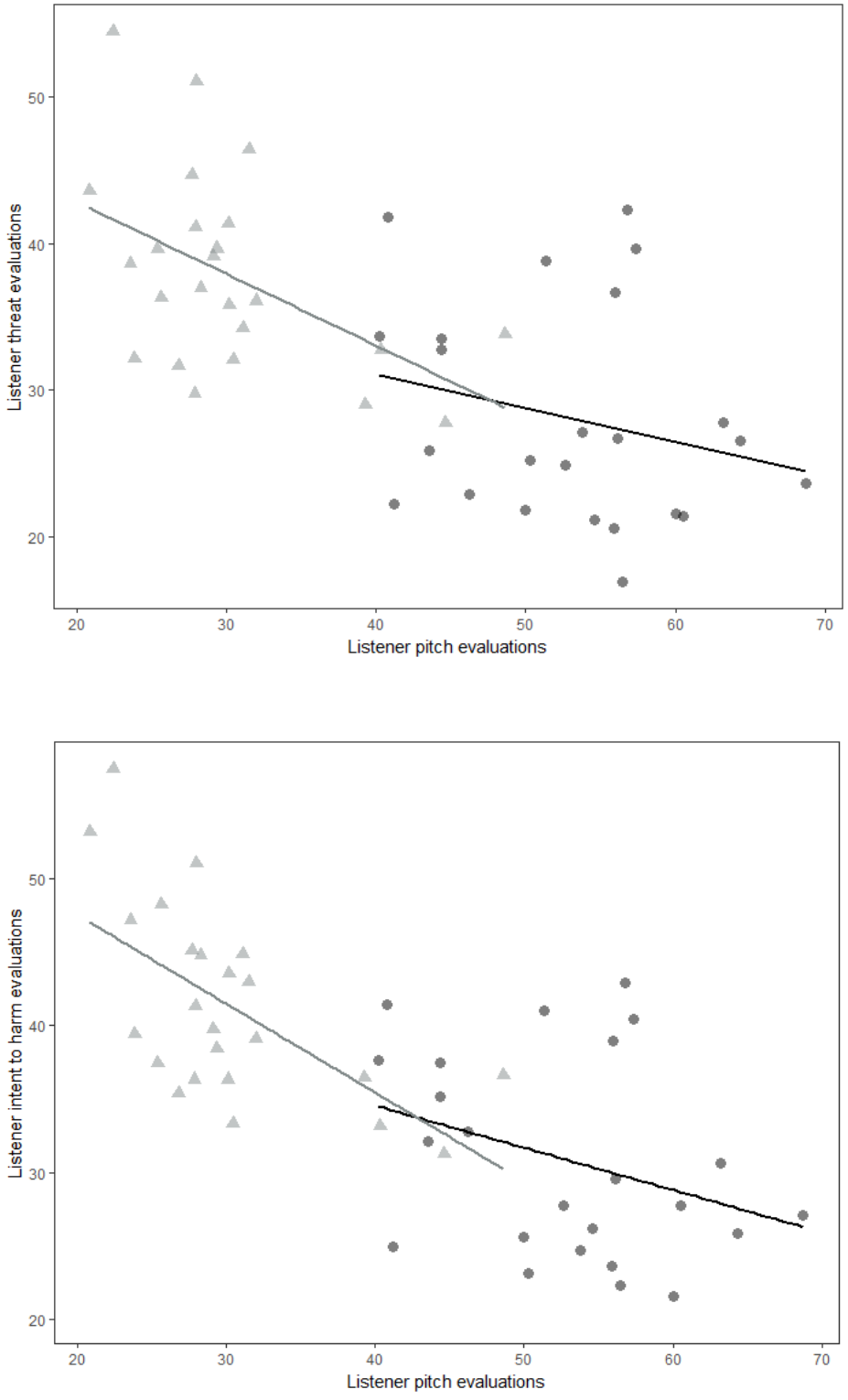


Figure 6.2 – Relationship between perceived pitch and threat evaluations (top) and between perceived pitch and intent evaluations (bottom) in Experiment 1. Points are averaged across listener for each utterance and split in accordance with speaker sex. Male speakers are represented by triangles and female speakers are represented by circles.

In addition to testing for the significance of the fixed effect predictors, the random effects of *speaker* and *listener* within the model were also analysed. This was done in order to evaluate whether the specified random effects significantly affected listeners' evaluations of both how threatening speakers sounded and how much intent to harm was conveyed through their speech. The results of this analysis are displayed in Table 6.2, which shows significant effects for both *listener* and *speaker* on listeners' evaluations of threat and intent to harm. This suggests that characteristics of both the 'threatener' and the hearer can significantly influence how utterances are perceived with respect to the level of perceived threat and intent to harm. While Tagliamonte and Baayen (2012) state that a large amount of variation being attributable to individual experimental participants is commonplace in psycholinguistic experiments, the effect of *listener* can be considered particularly noteworthy as a guard against the notion that spoken threats are likely to be interpreted in the same way by different listeners (cf. the assertion made by Gingiss (1989)).

	Threat evaluations			**Intent to harm evaluations**		
	χ^2	**df**	**P**	**χ^2**	**df**	**p**
Listener	214.82	1	<0.001	350.61	1	<0.001
Speaker	7.71	1	0.005	8.76	1	0.003

Table 6.2 - Effects of *speaker* and *listener* on listener evaluations of threat and intent to harm

In order to further evaluate the amount of variation in the data attributable to individual listeners, permutation tests were conducted on each utterance in the dataset using random samples of 12 listeners. This was done in order to analyse the amount of variation within any given subset of listeners, and random samples of 12 listeners were chosen given that this is the number of people required to sit on a jury panel in the UK. Given that juries are instructed to reach a unanimous decision in criminal cases in UK courts, it was considered interesting to see how varied listeners' threat and intent to harm evaluations would be within any random set of 12. These tests were conducted using MATLAB software, with 1000 random permutations of 12 listeners ran for each utterance. From this, the mean threat and intent to harm ratings for each group were recorded, along with the range and interquartile range for each random subset. Table 6.3 displays the average values for the 1000 subsets, for each speaker in the data. Average values were taken across the four utterances produced by each speaker.

Speaker	Mean		Interquartile range		Range	
	Threat	*Intent to harm*	*Threat*	*Intent to harm*	*Threat*	*Intent to harm*
Speaker 1	40	36	34	41	63	67
Speaker 2	24	25	28	40	60	66
Speaker 3	39	42	40	44	73	77
Speaker 4	40	47	37	35	74	76
Speaker 5	37	40	34	40	73	80
Speaker 6	23	25	31	38	69	72
Speaker 7	38	42	37	37	68	73
Speaker 8	30	31	37	43	66	73
Speaker 9	45	43	36	37	72	83
Speaker 10	24	28	31	43	65	73
Speaker 11	35	41	40	41	70	79
Speaker 12	31	34	35	40	70	72

Table 6.3 – Mean, interquartile range and range values showing averaged levels of listener agreement within subsets of 12 listeners' threat and intent to harm evaluations for each speaker. These values are produced by averaging the four subsets with the highest and lowest levels of agreement from each utterance to produce one value per speaker.

The values in Table 6.3 show a high level of variation between listeners' evaluations of threat and intent to harm. The data shows that the average interquartile range for both threat and intent to harm, for all 12 speakers in the experiment, extended beyond 30% of

the 100-point rating scale. The lowest average threat score range across the 1000 random trials of 12 listeners was 60 (Speaker 2), which equates to 60% of the total scale available to listeners, while the highest average threat score range in Table 6.3 was 74 (Speaker 4), which equates to almost three quarters of the total available scale. These values show a high overall level of disagreement between listeners within the random samples of 12 created for this analysis.

To further supplement the analysis in Table 6.3, Tables 6.4 and 6.5 display interquartile ranges for the listener subsets with the highest and lowest levels of agreement for each speaker. Threat evaluations are displayed in Table 6.4 and intent to harm evaluations are displayed in Table 6.5. The best-performing set was classed as the set with the smallest interquartile range, while the worst-performing set was classified as the set with the largest interquartile range. When two or more subsets had the same interquartile range, the range was used to differentiate and classify the subsets with the highest and lowest level of agreement. As before, averaged values were taken across the four utterances produced by each speaker.

Speaker	Interquartile range	
	Highest agreement	***Lowest agreement***
Speaker 1	16	50
Speaker 2	12	46
Speaker 3	15	60
Speaker 4	19	56
Speaker 5	16	57
Speaker 6	13	46
Speaker 7	16	50
Speaker 8	18	54
Speaker 9	14	56
Speaker 10	11	50
Speaker 11	15	57
Speaker 12	14	52

Table 6.4 – Interquartile ranges showing the averaged highest and lowest levels of agreement within subsets of 12 listeners' threat evaluations for each speaker. These values are produced by averaging the four subsets with the highest and lowest levels of agreement from each utterance to produce one value per speaker.

Speaker	Interquartile range	
	Highest agreement	***Lowest agreement***
Speaker 1	16	55
Speaker 2	13	57
Speaker 3	19	63
Speaker 4	16	51
Speaker 5	16	60
Speaker 6	17	52
Speaker 7	14	52
Speaker 8	15	58
Speaker 9	15	56
Speaker 10	14	57
Speaker 11	17	63
Speaker 12	18	58

Table 6.5 – Interquartile ranges showing the averaged highest and lowest levels of agreement within subsets of 12 listeners' intent to harm evaluations for each speaker. These values are produced by averaging the four subsets with the highest and lowest levels of agreement from each utterance to produce one value per speaker.

The values in both Table 6.4 and Table 6.5 illustrate the wide discrepancy between the subsets with the highest and lowest level of listener agreement. The worst performing subsets of 12 listeners for both threat and intent to harm evaluations show interquartile ranges spanning between 46% and 63% of the total scale available to listeners. In contrast, the best performing subsets show interquartile ranges spanning between 11% and 19% of the total scale available to participants. This analysis suggests that different levels of agreement would be achieved depending on which listeners were tasked with judging the level of threat and intent to harm within an indirect threat utterance. This

further strengthens the case for the view that not all listeners will interpret a potentially threatening utterance in a comparable or similar way.

6.3.3 Listeners' descriptions of speakers' voices

6.3.3.1 Pitch perception

To assess how accurate listeners' pitch judgements were, their perceived pitch scores were compared to the corresponding measured median F0 values for the voices of the speakers in the experiment. Figure 6.3 plots listeners' pitch judgements against the measured median F0 values, separated in accordance with speaker sex, given that F0 is a sexually dimorphic aspect of voice (see, for example, Puts et al., 2006).

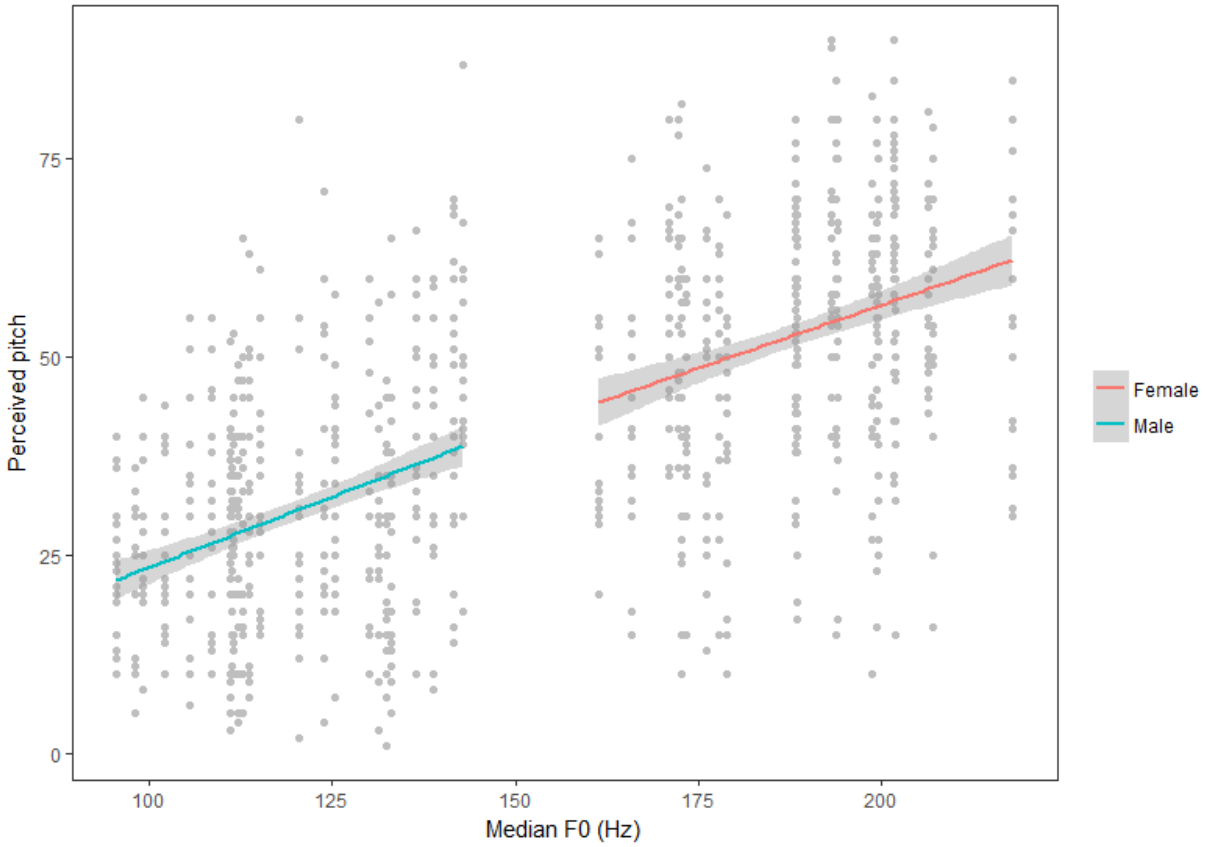


Figure 6.3 - Relationship between listeners' perceived pitch scores and median F0. The axis units are Hz (x axis) and listeners' subjective pitch ratings on a scale between 0 ('very low-pitched') and 100 ('very high-pitched') (y axis). Each dot represents a single listener judgement, and each column of dots represents an individual recording (four produced by each speaker; male and female speakers are treated separately).

Figure 6.3 shows a weak positive correlation between listeners' pitch judgements and the measured median F0 values for both male (Pearson's $r = 0.33$, df = 492, $p < 0.001$) and female (Pearson's $r = 0.28$, df = 479, $p < 0.001$) speakers. A small to medium effect sizes for the relationship between median F0 and perceived pitch for female and male speakers can be posited for the data in the present experiment. Additionally, although the relationship is reported by the models to be statistically significant for both male and female speakers, the graph in Figure 6.3 shows that a high level of variation exists between listeners' perceptual pitch scores and the corresponding measured F0 values.

The relatively high level of variation in the sample is also evident in the r^2 values for the relationship between measured mean F0 and perceived pitch. For the male speakers, 10% of the variation ($r^2 = 0.10$) in the sample was accounted for by the relationship between measured mean F0 and perceived pitch. For the female speakers, 7% ($r^2 = 0.07$) of the total variation was accounted for by this relationship. Figure 6.3 also illustrates that while male voices were, overall, perceived to be lower-pitched than the female voices, there was a relatively high degree of overlap between the perceived pitch judgements for the male and female voices in the experiment. This was, however, not mirrored in the measured median F0 values, which showed complete separation between male and female speakers.

Three potential explanations can be proposed to explain the results seen in Figure 6.3. The first is that individual listeners interpreted the perceived pitch scale differently, and that a given value on the scale did not, therefore, map onto the perceived pitch scale equivalently for each listener. Secondly, it could be the case that other aspects of speakers' voices, such as voice quality or the relative distribution of formants, could also influence pitch perception. This would mean that making direct comparisons between average F0 and perceived pitch is a rather crude one-dimensional measure of the accuracy of listeners' pitch judgements. Thirdly, it could be the case that listeners are both inconsistent and inaccurate when tasked with gauging how high-pitched the voice of a given speaker is.

In an attempt to reduce the influence of individual differences in how listeners interpreted the scale used to elicit perceived pitch judgements, standardised scores were calculated for each listener's judgements of the high-pitchedness of speakers' voices using the scale() function in R (Baayen, 2008:61). Figure 6.4 plots the standardised

perceived pitch scores against the corresponding measured median F0 values. The figure reveals that there was some reduction in variation when standardised scores were used, in comparison to the raw data displayed in Figure 6.4. Analysis of the correlation coefficients showed a slightly tighter positive correlation and increased effect size between perceived pitch and median F0 for both male (Pearson's $r = 0.40$, df = 492, $p < 0.001$) and female (Pearson's $r = 0.32$, df = 479, $p < 0.001$) speakers when standardised scores were used. However, the r^2 values for both male ($r^2 = 0.16$) and female ($r^2 = 0.10$) speakers showed that only a limited amount of variation was accounted for by the relationship between standardised perceived pitch scores and measured mean F0.

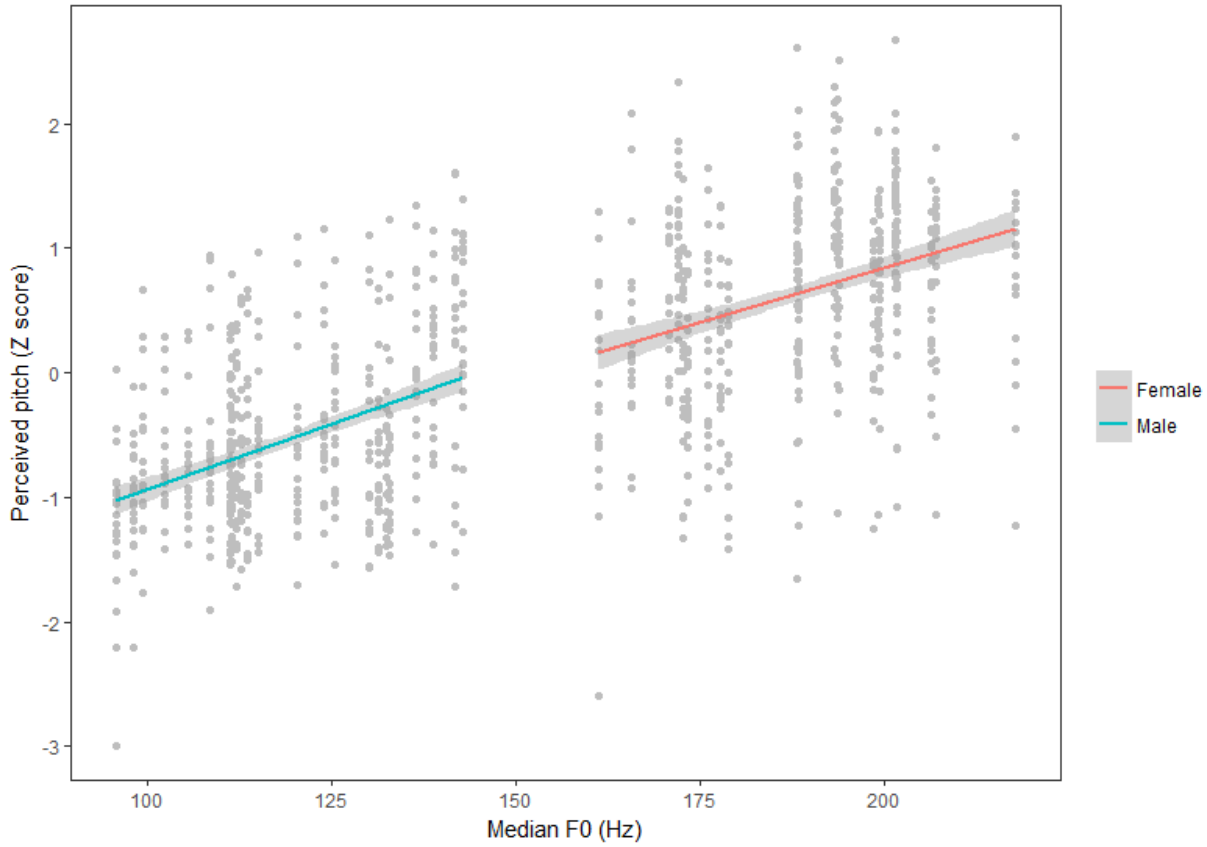


Figure 6.4 - Relationship between listeners' standardised perceived pitch scores and median F0. The axis units are Hz (x axis) and listeners' standardised subjective pitch ratings on a scale between 0 ('very low-pitched') and 100 ('very high-pitched') (y axis). Each dot represents a single listener judgement, and each column of dots represents an individual recording (four produced by each speaker; male and female speakers are treated separately).

The second reason that was proposed above for the weakness of the relationship between perceived pitch and measured mean F0 was that other variables, such as voice quality or the dispersion of formants across the frequency spectrum, could influence listener pitch perception alongside average F0. In order to assess the relationship between multiple acoustic phonetic variables and listeners' pitch judgements, multiple linear regression models were constructed using the lm() function in R. These contained listeners' perceived pitch scores as the dependent variable, and measurements of *median F0*, *F0 range*, *formant dispersion* (measured as the "average distance between adjacent

formants up to F3" (Xu, 2013)), *jitter* (an index of variability in glottal cycle duration), *shimmer* (glottal cycle amplitude variation), and *harmonic-to-noise ratio* of each speaker's voice in each utterance. These additional measurements were also extracted using the ProsodyPro Praat script (Xu, 2013). These variables were used to capture a range of information about each speaker's vocal tract resonances and phonation qualities. Separate models were constructed for male and female speakers. Analysis of the r^2 values from the models for both male ($r^2 = 0.15$) and female ($r^2 = 0.13$) speakers showed that a greater proportion of variance was accounted for when the additional acoustic measures were considered, although the respective models still only accounted for 15% and 13% of the variation in the data. The proportion of variance accounted for in the relationship between perceived pitch and acoustic aspects of voice was further enlarged by using the standardised pitch judgement scores instead of the raw judgement scores, with 20% of the variation being accounted for by the model for male speakers ($r^2 = 0.20$), and 19% by the model for female speakers ($r^2 = 0.19$). However, in order to capture this level of variation, multiple judgements made by the same listener were required in order to calculate the standardised pitch judgement scores. To some degree this could be considered unrealistic for users of documents such as the NCTSO bomb threat checklist, which is designed to obtain earwitness evaluations from a single listener about a single speaker on a single occasion.

Further questions arise from this analysis relating to the role of the individual listener in the pitch judgement task. Specifically, is it simply the case that some listeners are good at the task, and some are not? If this were indeed the case, then there might be some merit in testing the ability of an earwitness to distinguish aspects of voice, for instance pitch, before his/her earwitness evidence is further used. In order to address this question using the data from the current experiment, a subset was created containing the

responses of all those participants who provided a pitch judgement for three of the male speakers in the experiment, hereafter labelled Speaker 1, Speaker 2 and Speaker 3. These speakers were chosen because their average median F0 values (a) spanned the range found in the data for male speakers, (b) reflected the population statistics for English speakers' average F0 reported by Hudson et al. (2007), and (c) were almost equally spaced from each other along the pitch continuum (Speaker 1 - 99Hz, Speaker 2 - 120Hz and Speaker 3 - 140Hz). Given that the question randomisation process meant that not all listeners evaluated the voices of all speakers, some listeners were excluded from this analysis.

In total, 26 listeners provided at least one perceived pitch judgement for utterances produced by the three speakers described above. Table 6.6 shows the perceived pitch scores for each listener. The interest in this analysis is not in the absolute values, but rather in the relative pitch judgements provided by listeners. Given the 20Hz gaps between the three speakers' average median F0 values, it was expected that listeners would provide a lower perceived pitch score for Speaker 1 (99Hz) than for Speaker 2 (120Hz), and that the score applied to Speaker 2 would, in turn, be lower than the score for Speaker 3 (140Hz). If a listener met these criteria, they were classified as an accurate listener, shown in bold type in Table 6.6. This analysis showed there were 14 accurate listeners within the subset. This would support the view that some listeners are simply unable to judge pitch accurately according to the present criterion, while other listeners are capable of performing the task adequately or well.

Listener	Perceived pitch scores		
	Speaker 1 (99Hz)	**Speaker 2 (120Hz)**	**Speaker 3 (140Hz)**
P10	**30**	**31**	**58**
P11	19	18	41
P12	16	9	60
P13	19	51	49
P16	**20**	**30**	**70**
P17	**10**	**20**	**50**
P20	**10**	**35**	**45**
P25	**15**	**20**	**33**
P26	**21**	**28**	**40**
P28	20	20	20
P29	30	25	10
P3	**20**	**58**	**68**
P36	**10**	**23**	**30**
P40	**44**	**45**	**49**
P46	29	20	66
P47	26	18	18
P50	**11**	**26**	**49**
P52	**20**	**30**	**55**
P53	37	24	10
P6	**20**	**41**	**46**
P61	35	38	25
P63	22	12	33
P64	**29**	**35**	**51**
P69	28	13	60
P8	**25**	**34**	**55**
P87	39	37	46

Table 6.6 – Listeners' perceived pitch scores for Speaker 1, Speaker 2 and Speaker 3. Bold type denotes listeners who assigned the 'correct' ranking of the three speakers from low to high pitch, irrespective of the spacing on the perceptual scale between Speakers 1 and 2 and Speakers 2 and 3, or the placement of the scores on the 0-100 scale.

6.3.3.2 Body size perception

This experiment also elicited listeners' judgements of perceived body size, with the aim of analysing whether these would align with judgements of perceived pitch. This question was also addressed in Chapter 5, with these data showing that voices which were described as sounding 'high-pitched' by listeners were more likely to be described as being physically smaller than those speakers whose voices were described as sounding 'deep'. The analysis in this chapter builds on the work in Chapter 5, as listeners were not restricted to selecting from either 'deep' or 'high-pitched' options when judging perceived pitch, but could instead opt to use any value from the 0-100 scale when assessing how high-pitched a given speaker's voice was. Table 6.7 shows the total number of times each descriptor was used by listeners within the experiment, and displays relatively high usage of all descriptors apart from *very short* and *very tall*, which sit at the extreme ends of the available perceived height options.

Height descriptors		Build descriptors	
Descriptor	**Total no. of uses**	**Descriptor**	**Total no. of uses**
Average height	494	Average build	579
Taller than average	321	Small build	376
Shorter than average	302	Slim	348
Very short	61	Stocky	133
Very tall	41	Large build	115

Table 6.7 – Total number of times each body size descriptor was used by listeners

The relationship between body size evaluations and perceived pitch are displayed in Figures 6.5 and 6.6, and show a clear link between perceived pitch and both perceived height and perceived build. The trend in the data was for voices judged to be high-pitched to also be perceived as belonging to a physically smaller speaker, with the reverse pattern being observed for voices judged to be lower in pitch. The results highlight a trend for perceived lower-pitched voices to be more commonly associated with descriptors such as *stocky*, *large build*, *taller than average* and *very tall*, with perceived higher-pitched voices more commonly associated with descriptors such as *slim*, *small build*, *very short* and *shorter-than-average*. These results further strengthen

the view that perceived pitch is strongly linked to listeners' perceptions of body size, when judgements are made using auditory stimuli only.

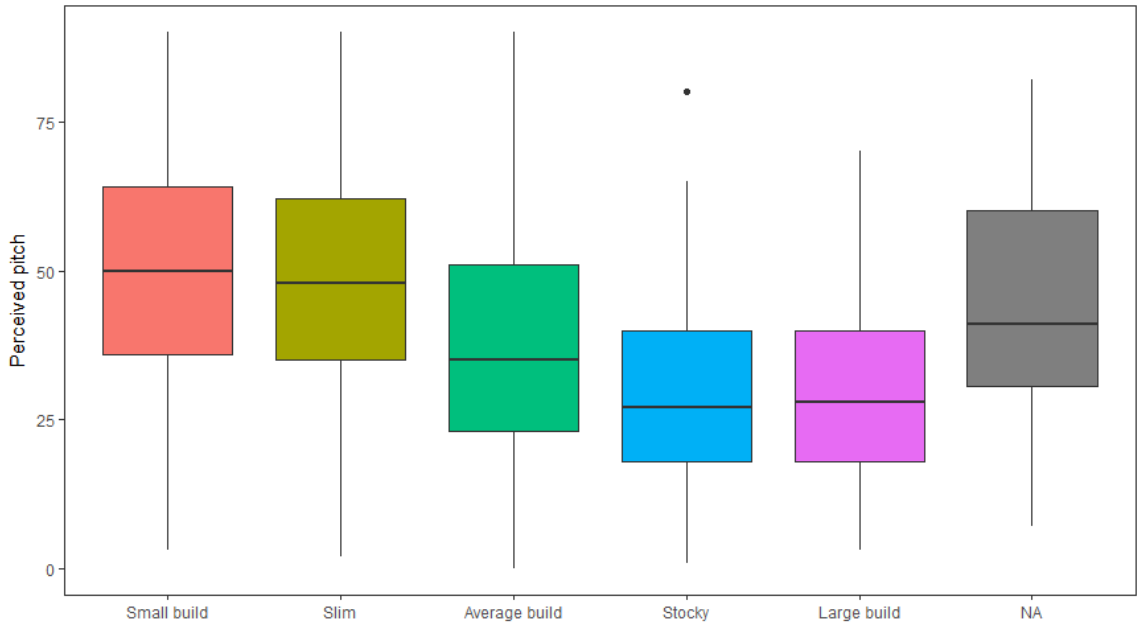


Figure 6.5 – Relationship between listeners' evaluations of perceived speaker build and listeners' evaluations of perceived pitch in speakers' voices

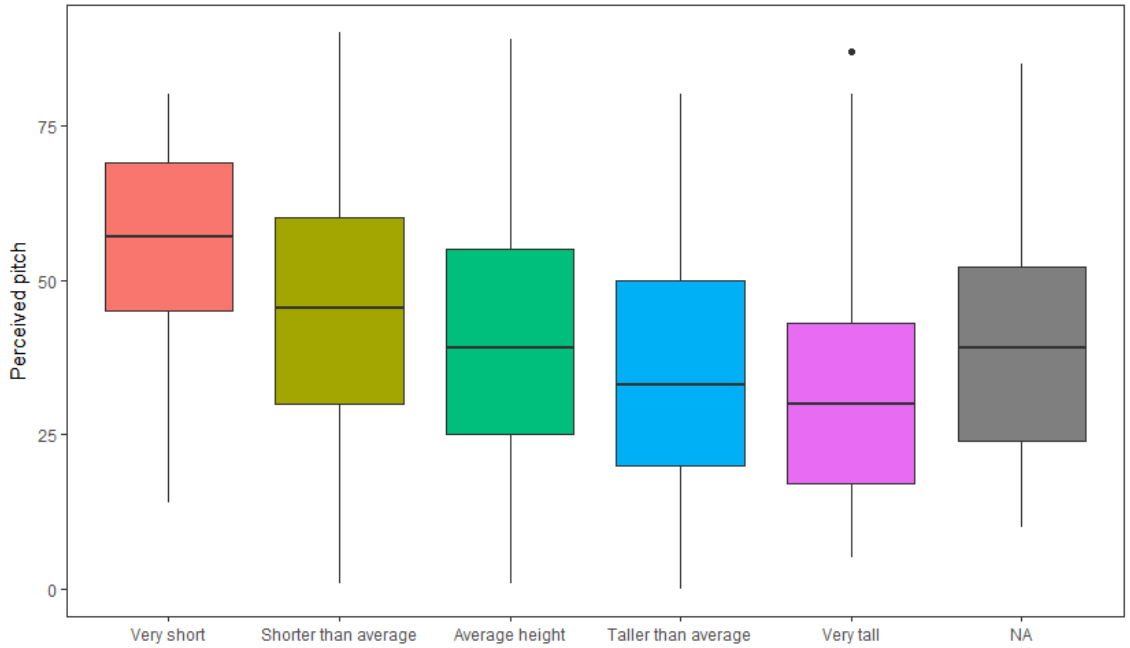


Figure 6.6 – Relationship between listeners' evaluations of perceived speaker height and listeners' evaluations of perceived pitch in speakers' voices

Given the relationship between perceived pitch and perceived body size, alongside the significant effects found for perceived pitch on listeners' evaluations of both threat and intent to harm in Section 6.3.2, this analysis also considered whether a predictable relationship would exist between body size evaluations and perceptions of threat and intent to harm. Although the context of the utterances in the experiment did not dictate that physical dominance would be a prerequisite for the ability to carry out the threatened act of planting a bomb, in the same way that it might be for a threat which pre-empted direct physical contact, it was predicted that perceived larger speakers may also be perceived to sound more threatening and to convey a greater level of intent to harm through their speech. Figures 6.7 and 6.8 show the relationship between listeners' judgements of how threatening speakers sounded and evaluations of perceived height and build.

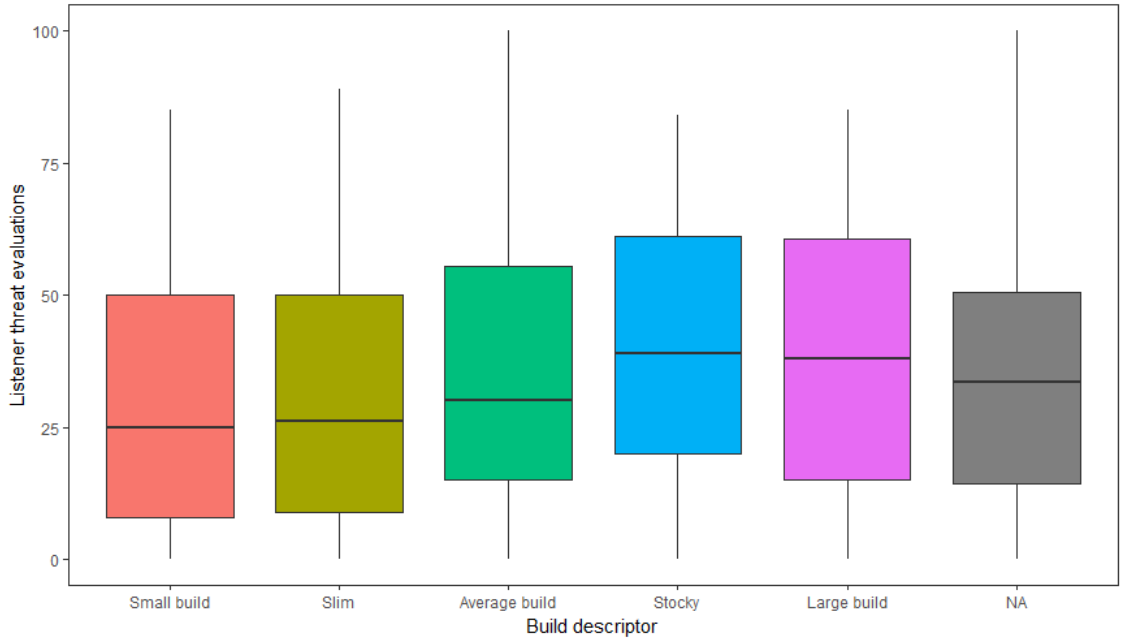


Figure 6.7 – Relationship between listeners' evaluations of perceived speaker build and listeners' evaluations of how threatening speakers sounded

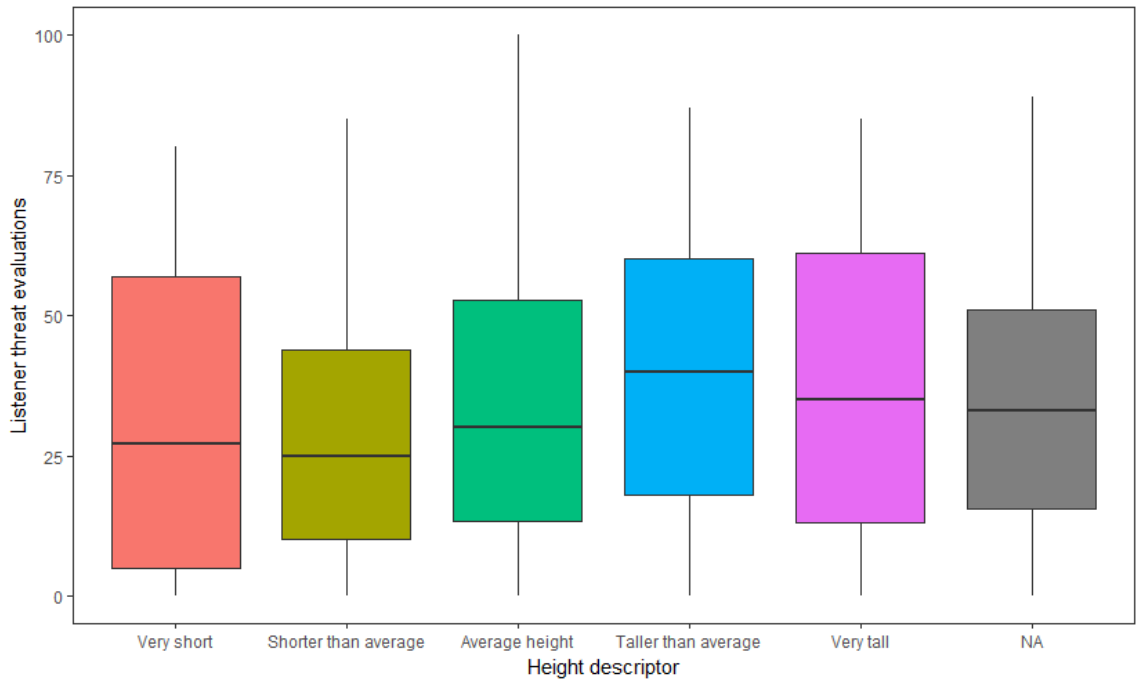


Figure 6.8 – Relationship between listeners' evaluations of perceived speaker height and listeners' evaluations of how threatening speakers sounded

The data in Figures 6.7 and 6.8 show some evidence for the predicted relationship between listeners' evaluations of how threatening speakers sounded and perceived speaker height and build. Although not as strong as the relationship between perceived pitch and perceived body size, both figures show a trend in line with the expectation that speakers who were perceived to be physically larger would also be perceived as sounding more threatening, and vice versa. The lack of uniformity with respect to the *very tall* and *very short* descriptors could be attributable to the small number of listeners who selected these descriptors in their evaluations. A similar pattern can be seen for the relationship between listeners' body size assessments and their evaluations of how much intent to harm was conveyed by speakers through their utterances. These effects are plotted in Figures 6.9 and 6.10.

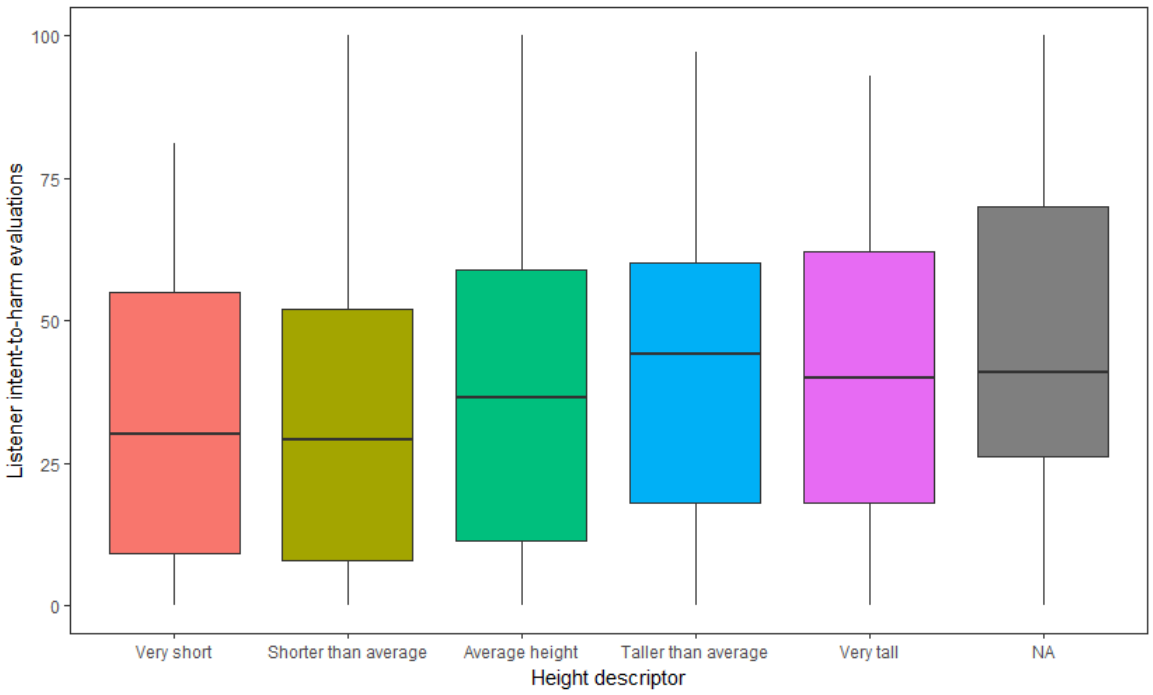


Figure 6.9 – Relationship between listeners' evaluations of perceived speaker height and listeners' evaluations of intent to harm

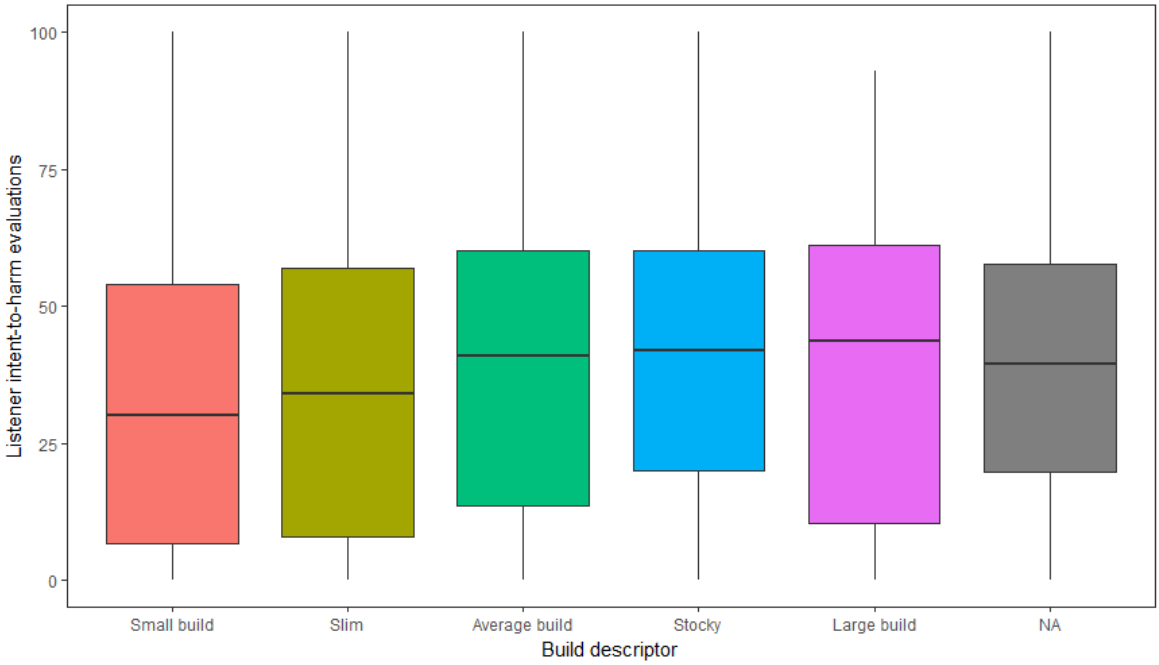


Figure 6.10 – Relationship between listeners' evaluations of perceived speaker build and listeners' evaluations of intent to harm

6.3.3.3 Accent perception

The data in this study also permitted an assessment of how accurately listeners could describe a speaker's accent via the responses to the question *"What accent do you think this speaker has? Leave the box blank if you are unsure"*. The experiment made use of three different accents: Standard Southern British English (SSBE), Northern Irish English, and 'Middle Eastern'-accented English. Listeners were also asked to state how similar they felt their own accent was to a series of other UK accents using a 0-100 scale (very different - very similar). The list of accents included *Oxford/Cambridge* and *Belfast* so as to facilitate an assessment of how closely aligned listeners thought their own accents were to the British target varieties in the experiment.

Responses to the question which obtainened listeners' assessments of how similar they thought their own accent was to the accents of both *Oxford/Cambridge* (SSBE) and *Belfast* (Northern Irish English) showed that listeners in the experiment aligned their own accents much more closely to SSBE than to Northern Irish English. The mean similarity score across the sample for SSBE was 40.4 (range 0-100), whereas the mean similarity score for Northern Irish English was 4.6 (range 0-47). Additionally, 51 listeners provided a similarity score of 0 for *Belfast*, in contrast to 17 listeners who gave a similarity score of 0 for *Oxford/Cambridge*. The full set of results of this analysis is shown in Table 6.8, below.

Accent group	Mean similarity score	Range of similarity scores
Oxford/Cambridge	40.4	0-100
Belfast	4.6	0-47
Glasgow	5.2	0-65
London Cockney	21.5	0-83
Yorkshire	38.0	0-100
Liverpool	13.9	0-85
Newcastle	14.9	0-95

Table 6.8 – Mean and ranges of similarity scores provided by listeners in response to the question asking how similar they thought their own accent was to people from the listed places.

A summary of the accent attributions for the SSBE speakers is shown in Figure 6.11. The results shown in Figure 6.11 reveal that listeners appeared to describe the accent of the SSBE speakers relatively accurately when they opted to describe it, although the most commonly chosen option was to leave the box blank to indicate uncertainty. When labelling SSBE, the most common way of answering besides selecting Unsure (Blank) was to choose one of the set of accent labels relating to SSBE or RP. These answers included "Southern", "Southern accent", "Southern England", "Southern English", "SSBE", "Standard Southern British", "RP", "Roughly RP" and "RP but grew up in London/'Estuary English' area" (Estuary English being the relatively newly-emerged 'hybrid' of RP and working-class London English; see Altendorf 2011). An association between the SSBE speakers and the prestigious university towns of Oxford and Cambridge was also found in the data, as was a link between the SSBE accent and the capital city of London. More specific places in southeast England, including Kent,

Chelsea and Surrey, were occasionally listed by listeners. More general terms such as "British" and "English" were also used, possibly owing to the generalisable nature of the accent, or to the presence of other accents in the experiment which were non-English.

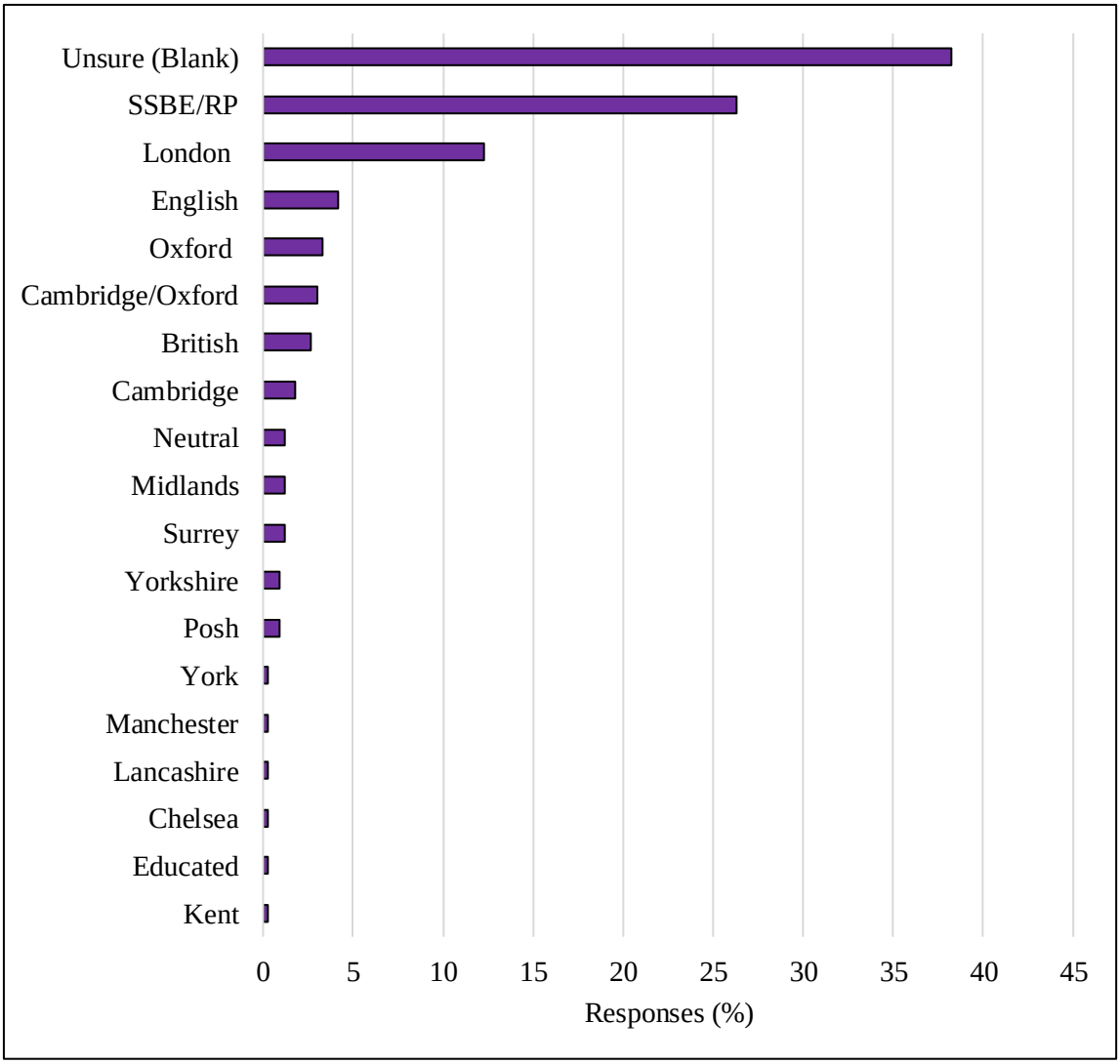


Figure 6.11 – Percentage distribution of responses to the question *"What accent do you think this speaker has? Leave the box blank if you are unsure"* for the Standard Southern British English speakers

Given the lack of a fixed geographical location for SSBE, and the position of Received Pronunciation (RP) as a social rather than a geographical accent of the UK (Hughes, Trudgill and Watt, 2012), it can be argued that an association with any location within

the south or south east of England could validly be considered an accurate description of an SSBE accent. It can also be argued that if a listener was unsure about a speaker's accent, then providing no answer rather than risking an inaccurate description was an appropriate strategy. Furthermore, it could be contended that the explicit instruction to provide no answer when the listener was unsure about a speaker's accent was a useful means of allowing listeners to express their uncertainty, rather than implicitly encouraging listeners to provide an accent label solely because the question asks for one. It is also possible that a listener's decision not to provide an answer was based on the perception that speaking with an SSBE accent means the talker has 'no' accent, a belief which is commonly held among laypeople in the UK (Mugglestone, 2003). It could also be the case that because SSBE is not confined to a specific locality in Southern England, it was not possible for listeners to define the speaker's accent to a specific town, city or region. Indeed, one participant in the experiment (Participant number 65) described her own accent as "no accent - plain southern but not posh", which further illustrates these possible explanations.

Figure 6.11 also shows that a negligible number of more inaccurate labels, including "Yorkshire", "Manchester", "York" and "Lancashire" were provided by listeners. While it is certainly true that some people from these places speak with RP/SSBE accents, or accents phonetically very close to the standard model, they are not areas where the majority of speakers would have such an accent. The descriptors "York" and "Yorkshire" could also be attributable to the fact that participants in the experiment were students at the University of York, an institution attracting significant numbers of students - many of whom have SSBE accents - from southern England and/or from affluent middle-class backgrounds. For a northern English city, York and its surrounding area is home to an unusually high proportion of university graduates and

people in professional occupations, and as such has a demographic profile that is markedly different from those of other urban areas of Yorkshire (Dorling 2010). These factors mean that students have numerous opportunities to be exposed to SSBE accents within their university city.

Figure 6.12 shows the responses provided by listeners when they were asked to describe the accent of the Northern Irish speakers in the experiment.

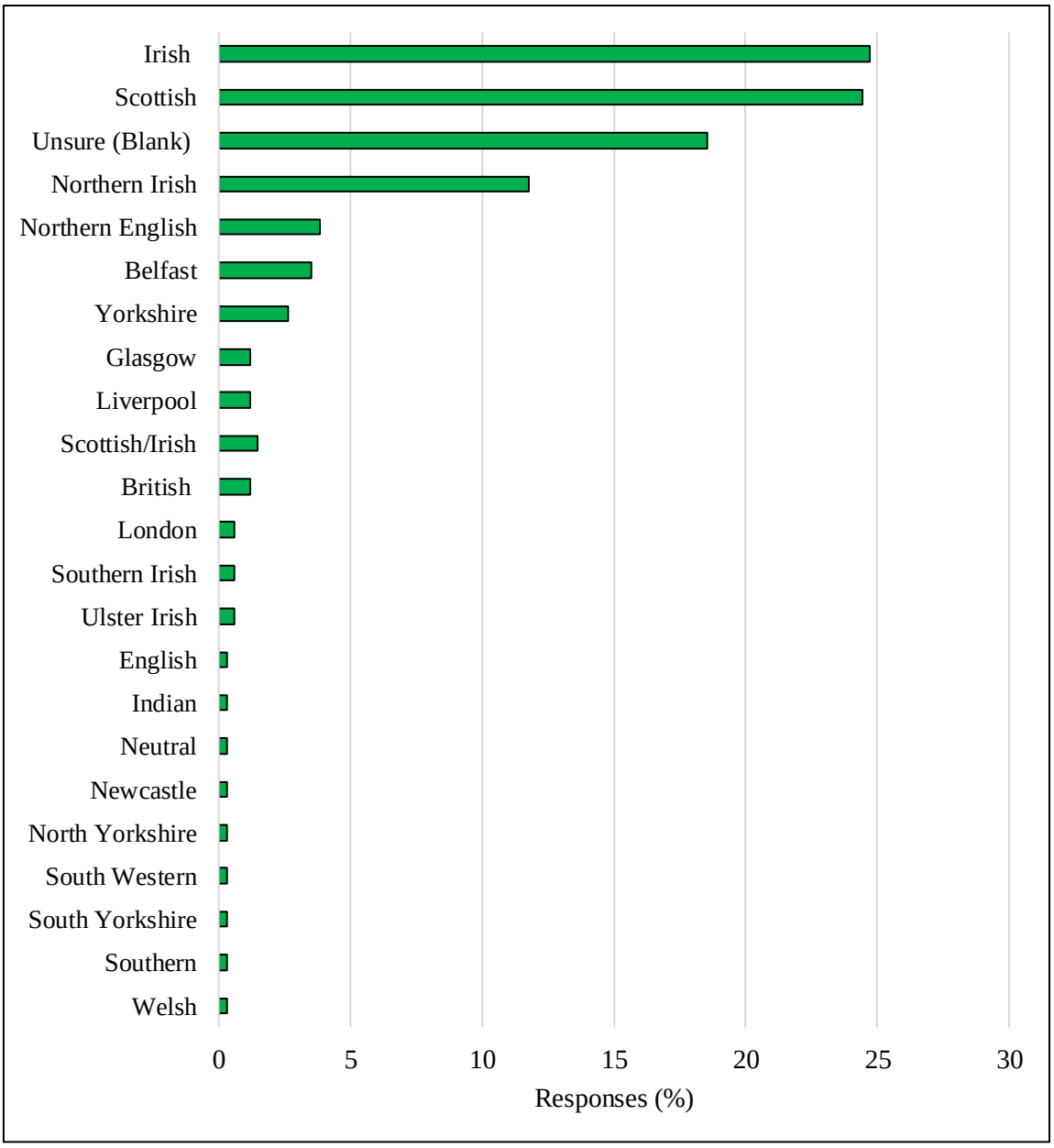


Figure 6.12 – Percentage distribution of responses to the question *"What accent do you think this speaker has? Leave the box blank if you are unsure"* for the Northern Irish speakers

In contrast to the SSBE accent description, the "Unsure (Blank)" classification was not the most popular label provided by listeners for the Northern Irish-accented speakers. Figure 6.12 shows that the accent was, instead, most commonly described as "Irish". There was also a much greater proportion of "Irish" labels compared with the number of "Northern Irish" and "Southern Irish" labels. This suggests that many speakers either

could not, or were unwilling to, determine the speaker's accent more precisely than to say he/she had an Irish accent. They may alternatively have thought the "Northern" qualifier to be superfluous, just as listeners from outside England might not think it necessary to specify whether an evidently English speaker is from the north or south of England. The results in Figure 6.12 also show that there appears to be confusion between listeners' perceptions of Northern Irish and Scottish accents. The Northern Irish speakers in the sample were frequently reported to have a Scottish accent, which was either indicated using a generic "Scottish" label or a more specific label such as "Glasgow". Given that the trend in the data was for listeners to say that their own accent was dissimilar to both Northern Irish English (mean similarity score to *Belfast* = 4.6/100) and Scottish English (mean similarity score to *Glasgow* = 5.2/100), the confusion is perhaps explainable by the relative lack of perceived similarity to and/or familiarity with, the target varieties.

Subsequent analysis was conducted to assess whether the confusion between the Northern Irish and Scottish accents was either speaker-specific, listener-specific, or both. Figure 6.13 shows the number of 'Scottish' and 'Irish' labels assigned to each of the four Northern Irish speakers in the sample. For the purposes of this analysis, labels were grouped so that the "Northern Irish", "Southern Irish", "Ulster Irish", "Irish" and "Belfast" descriptors were all grouped into the 'Irish' category, while the "Scottish" and "Glasgow" labels were grouped into the 'Scottish' category. Figure 6.13 shows that while the proportions of 'Scottish' and 'Irish' classifications were not the same for each speaker, no single speaker was consistently misidentified as sounding particularly Scottish by the listener group. The range of Scottish misidentifications was 32% to 45%. This suggests that the confusion between the two accents seen in Figure 6.12 was not the consequence of one or two speakers in the study being frequently mistaken for

Scottish speakers, but rather that the misidentification applied more or less consistently across all the speakers in the study.

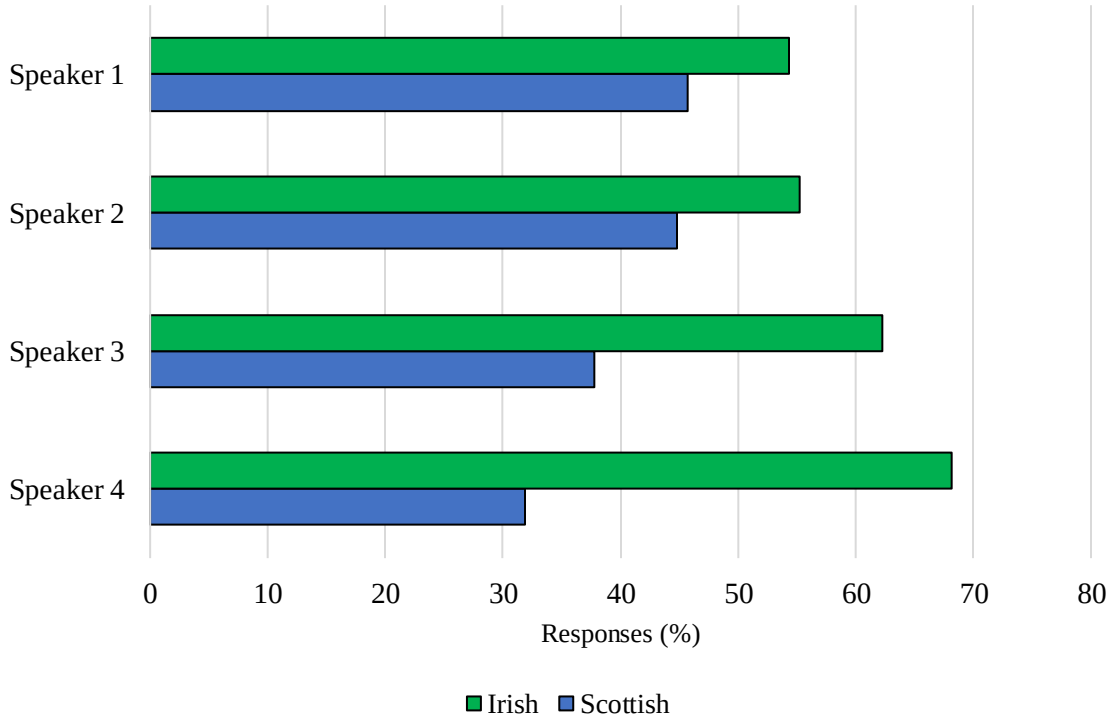


Figure 6.13 - Percentage of Scottish and Irish accent labels assigned to each of the four Northern Irish speakers in the sample

Given that the confusion of the Northern Irish and Scottish accents was not closely associated with any particular speaker, analysis was also conducted to assess how good individual listeners were at attributing the relevant accent labels correctly. The responses of each individual listener were assessed, with a count taken for the number of 'Irish' labels assigned to the Northern Irish voices. These results are shown in Table 6.9. Due to the automatic question randomisation process, results are displayed as percentages, as different listeners heard different numbers of the Northern Irish recordings (range = 1-8; mean = 4). Listeners were grouped according to the extent to

which they assigned an 'Irish' label to the voices of the Northern Irish speakers (in percent).

Percentage of 'Irish' attributions for the Northern Irish speakers	Number of listeners
0-20	33
21-40	12
41-60	9
61-80	9
81-100	20

Table 6.9 - Percentages for the number of listeners who provided 'Irish' labels for the Northern Irish speakers' accents

The data in Table 6.9 show that 20 listeners classified the Northern Irish accent using 'Irish' labels between 81 and 100% of the time. Conversely, 33 listeners classified the Northern Irish accent using 'Irish' labels between 0 and 20% of the time. This shows that the majority of listeners within the sample performed either very accurately or very inaccurately when assigning accent labels to the Northern Irish voices, and it suggests that labelling inaccuracies within the data shown in Figure 6.12 were the result of some listeners being consistently unable to provide a correct label.

The third accent included in this experiment was 'Middle Eastern'-accented English. Figure 6.14 shows the accent labels provided for the 'Middle Eastern' speaker samples. Given the large number of accent labels used to describe the 'Middle Eastern' speakers' voices, Figure 6.14 excludes labels which were used on just one occasion. These excluded labels were *African, American, British Arabic, Central European, Central*

Asian, *Korean*, *Automated*, *Greek*, *Hispanic*, *South American*, *Leeds*, *Northern British*, *Malaysian*, *Non-regional*, *Welsh*, *Swedish*, *Scandinavian*, *Scottish*, *South Africa*, *South East*, *Turkish*, and *Thai*.

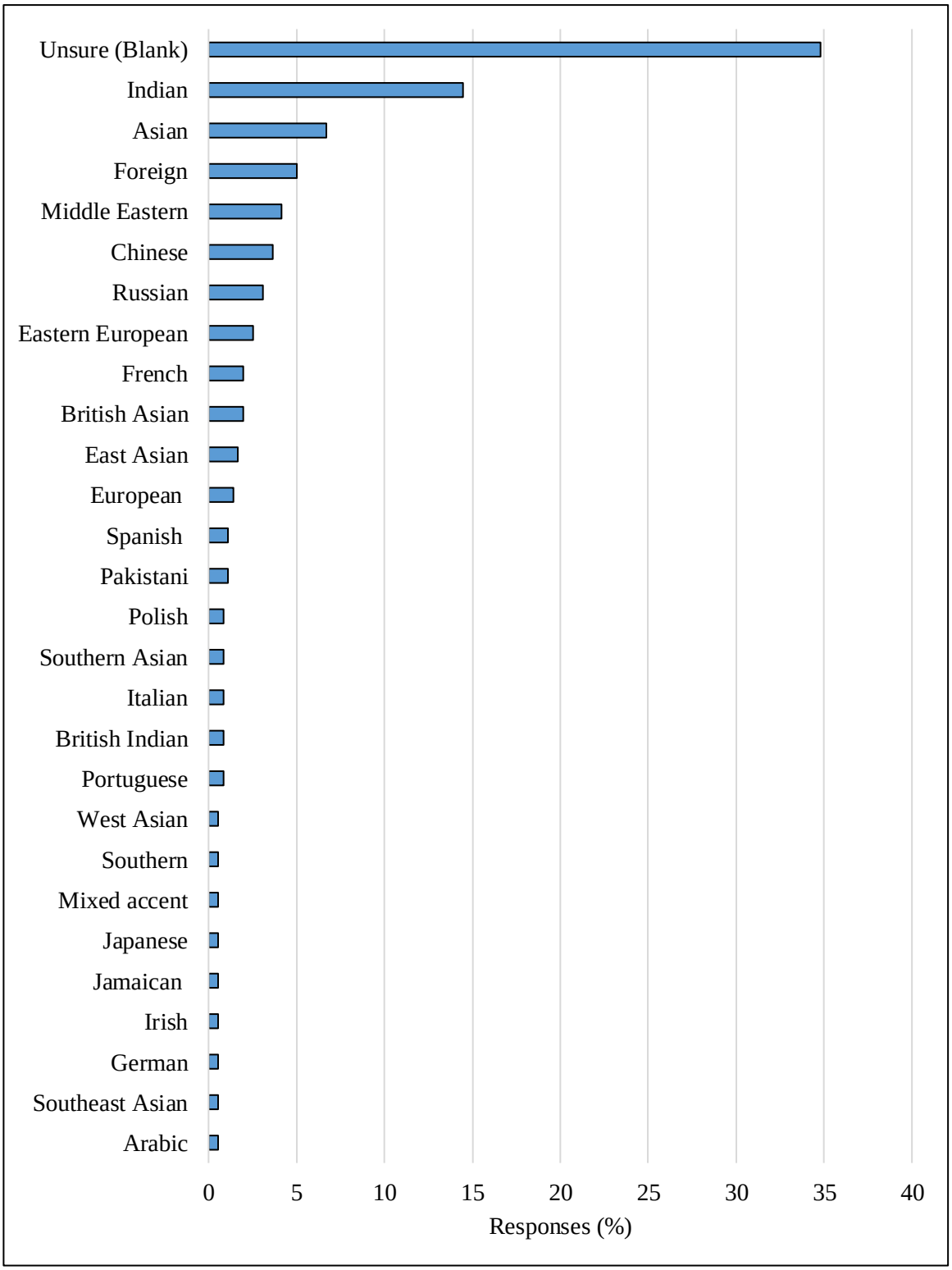


Figure 6.14 – Percentage distribution of responses to the question *"What accent do you think this speaker has? Leave the box blank if you are unsure"* for the Middle Eastern speakers

Figure 6.14 shows that, like the SSBE accent, the most popular accent label assigned to the 'Middle Eastern' speakers was *"Unsure (Blank)"*. There was also a greater number of different labels assigned to the Middle Eastern speakers (n=40) than to the SSBE speakers (n=19) or the Northern Irish speakers (n=23), suggesting a greater level of inconsistency among listeners when assigning accent labels to the foreign accent compared to the British accents in the experiment. The overwhelming majority of responses named a non-British location for the accent of the 'Middle Eastern' speakers in this experiment, but there was a high level of inconsistency within the labels assigned, which made reference to 35 different countries spanning five continents. Additionally, relatively few responses (n=17) pinpointed the 'Middle Eastern' speakers' accents as having any Arabic, Persian or Middle Eastern origin, with *Indian*, *Asian* and *foreign* the most commonly assigned labels. While the *foreign* descriptor is non-specific, it can be considered an accurate description in so far as listeners were able to say that the speakers in the 'Middle Eastern' recordings were not native British English speakers. It may also not be unreasonable to suggest that *Asian* is a relatively accurate accent label for the Middle Eastern speakers, given the proximity of parts of the 'Middle East' to the Asian subcontinent.[3] Again, however, the descriptor is relatively broad and arguably would be of rather limited use within a forensic investigation.

6.3.3.4 Unconstrained voice descriptions

In addition to providing assessments of how high-pitched speakers' voices were and an assessment of the accent of each speaker in the experiment, listeners were also able to

[3] The term *Asian* in the UK tends to be used to denote people with origins in the countries of the Asian subcontinent – chiefly India, Pakistan, and Bangladesh – rather than people of East Asian ancestry (China, Korea, Japan, Vietnam, etc.). It is also recognised that the Middle Eastern countries, including those of the Arabian Peninsula, are conventionally said to be part of the continent of Asia.

provide unconstrained descriptions of speakers' voices by responding to the question *"What other descriptors would you use to describe the speakers' voice, if any?"* In total, 584 descriptors were provided across the experiment, with multiple listeners describing multiple voices. The purpose of obtaining this data, following the work in Chapter 5, was to assess whether a relationship existed between these unconstrained descriptions and listeners evaluations of both threat and intent to harm. In order to conduct this analysis, the descriptors provided by participants in the study were categorised using the National Research Council of Canada's Emotion Lexicon database (EmoLex) (Mohammad et al., 2013; Mohammad and Turney, 2013a; 2013b). The EmoLex database is described as a "large word-emotion association lexicon" (Mohammad and Turney, 2013a:1). It contains 14,382 English words and labels whether each word is associated with the emotional states of *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise* and *trust*. The full data collection and analytical procedure for the database is outlined in Mohammad and Turney (2013b). Emotional associations are categorised in a binary way, with a given word being either associated with a certain emotion, or not associated with that emotion. Words are also classified by polarity as either *positive*, *negative*, or *neither positive or negative*. For example, *threat*, *threaten* and *threatening* are all classified as being negative words which are associated with anger and fear. Additionally, *threaten* is also associated with *anticipation*, and *threatening* is additionally associated with *disgust*.

Mohammad and Turney (2013b) highlight a range of uses for the EmoLex database, including a variety of marketing and technological applications. Perhaps more relevant to the work in this thesis is the argument that the database could be used in "detecting how people use emotion-bearing-words and metaphors to persuade and coerce others" (Mohammad and Turney, 2013b:3). For the data in this chapter, the primary use of the

EmoLex database was to assess whether listeners' unconstrained voice descriptions were biased towards emotional content related to threats or the act of being or sounding threatening. The work in this section uses the lexicon database to assess the emotional content of the unconstrained voice descriptions provided by listeners in this study. The aim of this analysis was to test how closely the descriptions provided by listeners aligned with listeners' evaluations of how threatening speakers sounded and how much intent to harm speakers conveyed through their speech. For example, it was hypothesised that there would there be a bias towards *negative* polarity in the set of voice descriptions provided in the experiment on account of threats being associated with *negative* polarity in the EmoLex database. In order to undertake this analysis, it was necessary to simplify and/or re-categorise some of the voice descriptions in order to align a voice description with a corresponding EmoLex descriptor. This often resulted in the removal of adverbs and other modifiers from the voice descriptions provided by listeners in the experiment. Additionally, there were several examples of voice descriptions which could not be aligned with an EmoLex entry. A range of examples which illustrate these issues are provided in Table 6.10.

Original voice description	EmoLex classification
A little worried	Worried
Bored	Boredom
Like she's anxious to warn the police	Anxious
Sense of urgency in voice	Urgent
Fed-up	Unhappy
Monotonous	Monotony
The speaker sounds very young	Young
Also adds a slight emphasis on the word "this"	N/A
Upper class	N/A
Straight to the point	N/A
Posh	N/A

Table 6.10 – Original voice descriptions provided by listeners and the corresponding EmoLex classification

From a total of 584 unconstrained voice descriptions, 466 were aligned with an EmoLex entry. This equated to 80% of the total amount of voice descriptions provided by listeners in the experiment. The ten most common EmoLex descriptors in the voice

description data were ***female*** *(82),* ***male*** *(62),* ***calm*** *(24),* ***monotony*** *(22),* ***worried*** *(14),* ***clearness*** *(13),* ***foreign*** *(12),* ***boredom*** *(10),* ***inform*** *(10)* and ***panic*** *(10).* The total number of times each EmoLex descriptor was used is presented in brackets next to the descriptor, meaning that, for example, *female* appeared 82 times in the data. A total of 105 different EmoLex descriptors were assigned to the voice descriptions provided by listeners, with the majority of descriptors (n=60) occurring just once.

Table 6.10 shows the number of *positive* and *negative* descriptors in the voice description data. Positive and negative classifications were mutually exclusive, with the exception of the word *intense*, which is classified as being both positive and negative in the EmoLex database. Analysis of positive and negative descriptors in the data shows that, perhaps surprisingly, more positive descriptors were used to describe the indirect threat stimuli compared with the number of negative descriptors. One possible skew in the data, however, is that *female* is classified as a positive word by EmoLex, whereas *male* is classified neutrally. These were the two most frequent descriptors used by listeners in the experiment, with 82 and 62 occurrences respectively. Table 6.11 also shows the average threat and intent to harm ratings assigned to the *positive* and *negative* descriptors. Stimuli in which the speaker's voice was described using a *positive* word were assigned a lower mean threat rating than stimuli in which the speaker's voice was described using a *negative* word. The same pattern was also seen in the average intent to harm ratings, with a lower mean threat rating assigned to voices which were described using a *positive* EmoLex descriptor.

	Number	Mean threat rating	Mean intent to harm rating
Positive descriptors	167	27	30
Negative descriptors	115	36	36

Table 6.11 - Average threat and intent to harm ratings assigned to the *positive* and *negative* EmoLex descriptors

A Mann-Whitney U test was conducted to further test whether the difference between the threat and intent to harm ratings in the *positive* and *negative* descriptor groups reached statistical significance. The output of these tests showed that the difference between threat ratings assigned to voices for which a *positive* descriptor was used, compared with those to which a *negative* descriptor was used, was statistically significant (p=0.004). However the difference between intent to harm ratings assigned to voices for which a *positive* descriptor was used, compared with those to which a *negative* descriptor was used, did not reach statistical significance (p=0.06). These results suggest a lack of separation between listeners' descriptions of speakers' voices and their evaluations of traits such as threat and intent to harm, with the effect being stronger for ratings of how threatening speakers sounded compared with how much intent to harm they conveyed through their speech.

In addition to examining the relationship between *positive* and *negative* voice descriptions, this section also considers the relationship between the emotional categories in the EmoLex database, the voice descriptions provided by listeners in the

experiment, and listeners' evaluations of perceived threat and intent to harm. Table 6.12 shows the number of voice descriptors that were associated with each of the eight emotions in the EmoLex database, alongside the average threat and intent to harm ratings assigned to descriptors classified into each of the emotional categories.

	Total number	**Mean threat rating**	**Mean intent to harm rating**
Anger	26	47	49
Disgust	14	45	50
Trust	35	33	33
Fear	61	32	34
Anticipation	31	30	25
Surprise	30	29	27
Joy	17	29	28
Sadness	45	28	28

Table 6.12 – Number of voice descriptions aligned with each emotion in the EmoLex database, alongside the mean threat and intent to harm ratings for descriptors in each category

The data in Table 6.12 illustrates that considerably higher threat and intent to harm ratings were assigned to voices that were also described using words associated with *anger* and *fear* according to the EmoLex database. Given that the EmoLex database associates *threat*, *threaten* and *threatening* with both anger and fear, the results in Table 6.12 could be seen as further evidence that listeners' descriptions of speakers' voices are not distinct from their evaluations of traits such as threat and intent to harm. The results in Table 6.12 also show that voice descriptions associated with those emotions which could be viewed as being dissimilar to anger and fear, such as joy and sadness, received comparatively lower mean threat and intent to harm ratings.

6.4 Discussion

The experiment presented in this chapter aimed to build on work presented in previous chapters by exploring how aspects of speakers' voices can influence decisions about how threatening speakers sound and how much intent to harm was conveyed in a given utterance.

The most important fixed effect found within the experiment was the significant effect of *perceived pitch* on listeners' judgements of both threat and intent to harm. The finding that voices perceived to be lower in pitch were generally evaluated as sounding more threatening than those perceived to be higher pitched would support previous research identifying a link between lowered pitch and the perception of dominance, aggression and other related traits (Ohala 1984; Tusing and Dillard, 2000; Puts et al., 2006; Puts et al., 2007; Mileva et al., 2018). However, the result also illustrates the potential importance of engaging with listeners' subjective perceptual scales when considering social evaluations based on aspects of voice, with listeners' own opinions about how high-pitched a speaker's voice was acting as a much stronger predictor than measured average F0.

The significant random effects of both *listener* and *speaker* were considered to be important findings with respect to highlighting the dangers that accompany assumptions surrounding the interpretation of potential language crimes. These results suggest that there is the potential for disagreement between listeners about the level of threat or intent to harm in a given utterance, alongside differences between speakers that were not captured by the fixed-effect variables. This result could have important implications for the evaluation of threats in courtrooms, which rely upon judges and juries to assess

the severity of potentially threatening utterances spoken by one or more speakers. In highlighting *listener* as a significant random effect, it is argued that caution should be advised around any assumption that all people will evaluate either how threatening an utterance sounds, or how much intent to harm was conveyed in a given utterance, in a comparable or similar way to one another. This finding was strengthened by the analysis which examined variation in threat and intent to harm evaluations within random subsets of 12 listeners – the number of people required to sit on a UK jury panel. Analysing the variation within these random sets of 12 listeners illustrated that a large amount of variation existed between listeners' threat and intent to harm evaluations, with average ranges sometimes spanning as much as 80% of the scale available to listeners making judgements about speakers' voices. This result further cements the importance of evidence-based, objective decision making regarding the analysis and interpretation of potential language crimes, rather than an over-reliance on the notion that language users will simply "know a threat when they hear one" (Gingiss, 1986:153). The strong effect of listener also limited the proportion of variation in the data captured by the fixed effects. This suggests that caution should be adopted in assuming the automaticity of any of the reported significant effects within the experiments.

The lack of an effect of speaker accent was, to some extent, surprising given the large quantity of previous research emphasising the importance of accent in social evaluations of speakers (Giles, 1970; Labov, 1972, Preston, 2002; Coupland and Bishop, 2007; Watson and Clark, 2015). This lack of effect may be attributable to the relative strength of the other tested effects, the choice of stimuli, or the use of a non-matched-guise design, which may have limited the perceptual strength of some of the accent features exhibited by speakers. It may also be that the group of listeners tested were not

susceptible to bias based on the hypothesised stereotypes about the accents presented in the study.

In this experiment, judgements of threat and intent to harm were strongly correlated. This supports the analogous finding in Watt, Kelly and Llamas (2013). However, in this study it was found that among the intent to harm judgements, there was a greater proportion of scores that exceeded the threat score given to the same stimulus. This was not the case in Watt, Kelly and Llamas (2013), who found the opposite pattern when obtaining listener assessments of the utterance *"I know where you live"*. It is argued that the use of a different indirect threat stimulus, which mentions a bomb, along with a time and specified place for the threatened action, could have resulted in this difference, with listeners being more unwilling to dismiss such an utterance as conveying either no or minimal intent to harm.

In addition to examining listeners' judgements of threat and intent to harm, the research in the chapter also assessed how accurately a group of listeners could perceive different aspects of a speaker's voice, with a view to evaluating the usefulness of such a practice in certain forensic contexts With respect to assessments of vocal pitch in line with measured average F0, the analysis showed small to medium-sized correlations in the data between median F0 and listeners' judgements of how high-pitched speakers' voices were. The coefficients improved when standardised pitch scores were used, and when other acoustic measurements relating to pitch and voice quality were included alongside average F0 measurements. However, the best-performing model – that for male speakers – still only accounted for 20% of the total variation. The analysis also illustrated how some listeners within the sample seemed unable to correctly appraise the relative differences between three speakers' voices with average median F0 values of

99Hz (Speaker 1), 120Hz (Speaker 2) and 140Hz (Speaker 3). Of the 26 listeners who evaluated these speakers' voices, only 14 assigned relative pitch judgements in accordance with the increase in average F0 across the three speakers' voices. This suggests that some listeners lack the ability to reliably judge how high-pitched a speaker's voice is, while some listeners are able to accurately estimate vocal pitch in line with measured acoustic correlates.

With regards to the description of accents, the analysis suggests that listeners' abilities to describe accents decreases as the degree of unfamiliarity or geographical distance increases. There were relatively few inaccurate labels used to describe the accents of the SSBE speakers, with a higher number of confusions shown when listeners were asked to describe the Northern Irish speakers' accents, and further confusion when listeners were asked to describe the accents of the 'Middle Eastern' speakers. Given the trend in the data for speakers to identify the SSBE accents as being more similar to their own in comparison to the Northern Irish English, these data would support the idea that the more geographically distant or unfamiliar an accent is, the greater the scope for confusion or otherwise inaccurate accent labelling (the L1 accents of English spoken in Australia and New Zealand are obvious likely counterexamples to this generalisation, but they do not invalidate the general 'proximity effect' patterns observed in numerous perceptual dialectology studies. See e.g. Montgomery, 2015; Shen & Watt, 2015; Preston, 2018). Had this study tested listeners from Glasgow, Edinburgh or Belfast, then the misclassification of the Northern Irish-accented speakers as Scottish speakers would not have been expected as the listeners would have presumably been more familiar with the tested varieties.

Additionally, given the wide variety of answers provided by listeners in this study when they were asked to describe the accents of the Middle Eastern speakers, the data lend support to the view that there is limited value in asking phonetically untrained non-native listeners to assess the geographical provenance/nationalities of speakers based on vocal information alone. As listeners rarely assigned a "British" label to the Middle Eastern accents, it could be argued that listeners were adequately equipped to assess whether a speaker was a native or non-native speaker of English. However, any information beyond this was unreliable. This generalisation is especially important in view of the fact that the NCTSO bomb threat checklist encourages users to give an opinion concerning a speaker's possible nationality. The results also urge caution in regional accent identification by lay-listeners owing to the poor performance of some listeners in the Northern Irish accent classification task. They also suggest that factors such as the background of the listener and their general accent classification ability should also be considered.

It can also be contended that the use of information about a speaker's accent obtained through asking non-linguist earwitnesses to describe the voice of a given speaker should also be used in conjunction with the knowledge that not all accents have a well-defined corresponding geographical location. For example, the spread of geographical locations that listeners associated with the SSBE speakers in this study spanned much of the south of England, and yet it cannot be considered 'inaccurate' to suggest that SSBE speakers could come from any of those places. It is argued, therefore, that the use of speech-based evidence in the form of phonetically untrained listeners' descriptions of voices and accents should be treated with due scepticism by default, and that such information should be used in conjunction with empirically verified data about UK and international varieties of English that have been collected by professional linguists.

One possible improvement to the practice of eliciting information from earwitnesses would be the development of a set of materials designed to test a listener's abilities to identify various aspects of speakers' voices. Given that the evidence recorded in documents such as the NCTSO bomb threat checklist would, in many cases, be based only on the perceptions of a single listener, it would potentially be useful to assess the capability of that earwitness to make reliable observations of different aspects of speaker's voices. This would allow the police and other investigative agencies to verify whether the checklist user can consistently and accurately identify different aspects of voice before any use is made – either in court or for the purposes of further investigative work – of subsequent checklist evidence he or she might produce (cf. the recommendations laid out in Nolan (2003) concerning testing of earwitness reliability using the voice parade paradigm). However, such a recommendation would require more research to be implemented in practise, specifically regarding the finer details of how such a test could be standardised and implemented by those working in the criminal justice sphere.

There is plentiful scope for expansion of the design of this study in future work, which could focus on other aspects of voice such as speech rate, variation in the F0 contour as a cue to how 'monotonous' or 'lively' a speaker's utterances are perceived to be, nasality, disfluency features (e.g. hesitations, filled pauses, etc.), and the use of paralinguistic markers such as clicks. As was referred to earlier, it is also acknowledged that the experimental conditions in this study created a more favourable earwitness environment than would be expected in certain real-world scenarios, such as the handling of bomb threats in emergency service control rooms, hospitals or schools.

Chapter 7 - Discussion

7.1. Introduction

In Chapter 1, the main aim of this research project was identified as to examine, using an experimental methodology, potential phonetic and linguistic factors which could cause listeners to perceive greater or lesser levels of threat, intent-to-harm and other traits in speakers' voices. In doing so, it was hoped that the research presented through this thesis would help to facilitate a greater level of understanding surrounding the perception of spoken threats. This was considered important in view of the assertions made by Watt, Kelly and Llamas (2013) that limited knowledge exists surrounding the extent to which phonetic parameters of speakers' voices can influence listeners' perceptions of how threatening speakers sound. Additionally, the research presented in this thesis also investigated the abilities of listeners to describe aspects of speakers' voices such as pitch, regional accent and voice quality. This aspect of the project was conducted following the assertion made by Griffiths (2012) that it is important for linguists to gain a broader and more comprehensive understanding of how people who do not have advanced-level linguistic training perceive and describe speakers' voices.

This chapter evaluates the findings presented in Chapters 3, 4, 5 and 6 with respect to providing answers to the overarching research questions set out in Chapter 1. Firstly, a summary of the key findings from the experimental work is provided, before a discussion of the implications of those findings and how successful they have been in answering the research questions set out at the start of the project. The discussion will also highlight areas for future research into both spoken threats and listener evaluations of speakers' voices.

7.2. Summary of key findings

Before turning to an overall discussion of the research findings, this section summarises the results of the experimental chapters in this thesis with respect to the overarching research questions outlined at the outset of this project.

Chapter 3 presented results from two experiments which were designed primarily as concept exploration tests for ideas which were subsequently developed in later chapters. In Experiment 1, listeners were presented with a series of utterances containing two indirect threats – *"I know where you live"* and *"I wouldn't do that if I were you"*. These target utterances were produced by a single speaker using a matched guise design, which created two contrasting speaker accent groups. A non-standard London Cockney guise was used alongside a standard Received Pronunciation accent guise. The target stimuli were also altered to create contrasting mean fundamental frequency (F0) levels. Stimuli were resynthesized to create three mean F0 levels – low (90Hz), mid (115Hz) and high (140Hz). The goal of the experiment was to assess whether F0 and speaker accent would influence listeners' judgements of how threatening the matched guise stimuli sounded. Results showed that both mean F0 and speaker accent significantly affected listeners' threat evaluations. These effects were in the expected direction, with the low F0 stimuli rated as sounding the most threatening and the high F0 stimuli rated as sounding the least threatening. Additionally, the non-standard London Cockney accent guise was rated as sounding more threatening than the standard RP guise. This was expected following the results of previous accent evaluation studies (see, for example, Dixon, Mahoney and Cocks, 2002). It was also noteworthy that the effects for mean F0 and speaker accent were stronger for the utterance category which was rated as sounding less threatening overall (*"I wouldn't do that if I were you"*). The link between F0, perceived body size and perceived threat was also examined in Experiment 1, with

results of the qualitative evaluations of speaker body size showing a trend for speakers with lower F0 to be perceived as having a larger body size in terms of both height and build. However, for evaluations of speaker build, the RP guise with low F0 was evaluated as having a physically smaller build than the London Cockney guise with the same low F0 level. This suggested a link between body size and threat evaluations, with speakers who were evaluated as being physically larger also perceived as sounding more threatening. It was also suggested that evaluative accent judgements could also play a role in listeners' body size evaluations. Finally, the experiment also showed strong positive correlations between listeners' judgements of threat and aggression, threat and menace, and threat and anger, as well as a moderate negative correlation between threat and friendliness ratings.

The design of the study in Experiment 2 adopted the approach taken by Watt, Kelly and Llamas (2013) in order to control for the interpretation of the verbal channel through the use of unfamiliar foreign language stimuli. The stimuli for this experiment were produced by both male and female speakers. German and Polish speakers were used in this experiment, with alterations performed on the stimuli to create contrasting low and high F0 levels (90/140Hz for male speakers, 170Hz/250Hz for female speakers), and slow (-20% of original speaking tempo) and fast (+20% of original speaking tempo) versions of each utterance. The effects of mean F0, speech rate and F0 range on listeners' evaluations of how threatening speakers sounded were analysed. Results showed that both mean F0 and speech rate had a significant effect on listeners' threat evaluations. There was also a significant interaction shown between F0 and speech rate. There was a trend for low mean F0 recordings to be evaluated as sounding more threatening than their high F0 counterparts, and a bigger difference in threat ratings for the slow and fast recordings between the high and low F0 categories. The experiment

also showed a difference in ratings assigned to male and female speakers, with the male speakers perceived to sound more threatening, and a difference in ratings assigned to the different languages, with the Polish utterances judged as sounding more threatening than the German utterances.

Both experiments presented in Chapter 3 were designed as initial explorations of the notion that different aspects of speakers' voices could cause listeners to infer different levels of threat in those speakers' voices. However, as initial experiments, there were some limitations that the studies presented in subsequent chapters aimed to address and expand upon. In Section 2.6, the assertion made by Milburn and Watman (1981) that situational context is a core element in the communication of threats was reiterated. However, listeners were given no sense of any context in which the judgements they provided in Experiments 1 and 2 should be made. The third experiment presented in this thesis, in Chapter 4, aimed to address this weakness by examining whether providing experimental participants with contextual information would influence their evaluations of how threatening speakers sounded. The effect of different phonation qualities on listeners' evaluative judgements of how threatening speakers sounded was also tested in Experiment 3, alongside differences in vocal pitch and speaking tempo. The experimental design followed that of Experiment 2 by using unfamiliar foreign language stimuli, although listeners were informed in the experiment that the language they would hear would be German. This information was provided so that all listeners were aware they were listening to the same language. Half of the listeners in the experiment were told that the speaker in each recording was stating that there was a bomb at a local football stadium that would be detonated unless certain demands were met, while the other half were given no contextual information other than that in some forensic cases, listeners can be asked to evaluate the voices of speakers they hear. In

addition to evaluating how threatening speakers sounded, listeners were also asked to describe each voice using an adapted version of the NCTSO bomb threat checklist, and to indicate what they thought the speaker in each recording looked like based on their voice. This was done and using a list of predetermined options for speakers' perceived height and build. Results showed significant differences between threat ratings assigned to the different phonation qualities, with the creaky and harsh voice samples perceived as sounding the most threatening. In contrast, the falsetto recordings perceived as conveying the lowest amount of threat.

Experiment 3 also showed that listeners were capable of describing some aspects of phonation quality with a reasonably high level of accuracy, such as whispery voice. However, potentially problematic descriptive terms such as *angry* and *calm* were also frequently used by listeners when they were tasked with describing the voices that they heard. When listeners' descriptive evaluations were analysed alongside their evaluations of how threatening speakers sounded, results showed that voices labelled as sounding *angry*, *harsh*, *creaky*, and *hoarse* were perceived as sounding more threatening than voices for which those descriptors were not used. In contrast, the reverse effect was seen for voices described as sounding *calm* and *high-pitched*. As was the case in Experiment 1, voices judged to be higher in pitch were more frequently associated with the perception of a smaller speaker, while the reverse effect held for voices perceived to be lower in pitch. A difference in threat ratings assigned to the same utterances in the different experimental contexts was also observed. This contextual effect was in the expected direction with higher ratings assigned by the group who were told that the recordings they evaluated were bomb threats. While this result was not surprising, it showed that providing participants with contextual information in an experimental setting could influence judgements about how threatening speakers sound. Other results

from Experiment 3 showed that the NCTSO checklist evaluations revealed a wide variety of age estimations for the non-modal voice samples, and that the falsetto recordings frequently resulted in a male speaker being incorrectly classified as a female speaker.

Experiment 4, presented in Chapter 5, was designed to build on the work presented in the previous chapters by expanding the scope of the research to include contentful utterances produced by multiple speakers. The experimental design was again based around bomb threat evaluations, following the effects for experimental context seen in Experiment 3. The analysis presented in Experiment 4 focussed on listeners' own productions of what they considered to be a 'threatening tone of voice' (Watt, Kelly and Llamas, 2013). It also focussed on the degree of phonetic emphasis placed on realisations of the word 'will', which had previously been linked to increased commitment on the part of threateners (Napier and Mardigian, 2003; Gales, 2010; Nini, 2017). The work in Chapter 5 also elicited further descriptions of speakers' voices using the NCTSO bomb threat checklist framework. It was also the first experiment in this thesis to link analysis of speakers' unconstrained productions of a 'threatening tone of voice' to listeners' perceptions of how threatening speakers sounded. The results showed that across the sample of speakers, there were significant differences in the expected direction for mean F0, duration and intensity across the produced tokens of 'will' between the threatening and neutral tone of voice productions of the target utterances. This difference was not mirrored in other tokens, and suggests that there was a trend for speakers to identify 'will' as a word which could be used to signal a threatening tone of voice. The results of the perception experiment showed that listeners rated the 'neutral tone of voice' utterances as sounding less threatening than the 'threatening tone of voice' utterances. While this result was unsurprising, it is important

in emphasising a link between speakers' productions and listeners' perceptions of what makes an utterance sound threatening. This relates to the notion that shared knowledge between speaker and hearer is a requirement for the successful communication of a given threat. Analysis of listeners' own descriptions of speakers' voices showed that, as was the case for the experiment in Chapter 4, there was a trend in the data for voices described as sounding 'angry' and 'harsh' to be rated as sounding more threatening than voices for which those descriptors were not used. The reverse pattern was seen for the 'calm' descriptor, with voices described as sounding calm assigned significantly lower threat ratings than voices for which the 'calm' descriptor was not used. Results also suggested that the term 'harsh' was more closely aligned to perceptions of anger than it was to other terms which may have been used to describe a phonetically harsh voice quality. The research presented in Chapter 5 also highlighted further inconsistencies in listeners' judgements of pitch and speech rate for certain voices, while also showing greater levels of agreement between listeners for those voices which were at the more extreme ends of the measured F0 and speech rate scales.

The final experiment in this thesis was outlined in Chapter 6, and aimed to present the most comprehensive experiment in the thesis with respect to an examination of how listeners infer traits about speakers from indirect spoken threat productions. Experiment 5 followed the design of Experiment 4 and incorporated vocal stimuli from multiple speakers producing multiple indirect threat utterances. The same utterances used in Experiment 4 – *"There's a bomb at York Station which will go off this afternoon"* and *"I'm warning you about a bomb at York Station which will go off this afternoon"* – were re-recorded by different speakers for use in this experiment. Speakers were asked to produce each utterance twice, once where they emphasised the word 'will', and once where they emphasised the word 'this'. The experiment used speakers of three different

varieties – SSBE, Northern Irish English, and foreign-accented English – in order to further test the findings from Experiment 1 that speaker accent can influence threat judgements. The effect of these variables, along with a range of other phonetic and linguistic variables on listeners' evaluations of both how threatening speakers sounded and how much intent to harm was conveyed through their speech was tested. These other variables included average F0, F0 range and the indirect threat utterance. Given the finding from Experiment 4 that listeners' judgements of pitch and speech rate did not always align with measured average F0 and measured speech rate respectively, the experiment elicited judgements of pitch and speaking tempo from listeners. The effect of these measures on listeners' evaluative judgements of threat and intent to harm were also tested. The research in Chapter 6 also assessed the accuracy of listeners' judgements of pitch and speaker accent against measured acoustic correlates of pitch and ground truth knowledge of the speakers' accents respectively.

The final aspects of the research included within this experiment were listeners' assessments of speakers' body size, and an analysis of a series of unconstrained voice descriptions provided by listeners. The results of these facets of the experiment showed a significant effect for perceived pitch on listeners' evaluations of both threat and intent to harm, with a trend in the data for lower-pitched voices to be evaluated as sounding more threatening than higher-pitched voices. The results also showed that the effect was stronger for male speakers than for female speakers. With respect to the inference of threat and intent to harm, the other key finding from this study was the significant random effects of both speaker and listener, with the effect of listener being particularly strong. Through analysing 1000 random samples of 12 listeners, the number required to sit on a jury panel in the UK, a high level of individual variation was shown between the different listeners in the experiment. This result was used as evidence that the

analysis of spoken threats by listeners does not reflect the view that individuals will always "know a threat when they hear one" (Gingiss, 1986:153). The results from the experiment in Chapter 6 showed a strong correlation between listeners' judgements of how threatening speakers sounded and how much intent to harm was conveyed through their speech. With respect to listeners' own evaluations and descriptions of speakers' voices, the analysis suggested that some listeners have the ability to assess vocal pitch in line with measured acoustic correlates, but others did not. A similar result was seen for the regional accent judgements, with those accents that were more geographically distant and/or unfamiliar being described less accurately. Furthermore, listeners assessments of speakers' body size were closely aligned to perceived pitch judgements in the expected direction, with higher pitched voices correlating with the perception of a larger speaker. Finally, the analysis of the unconstrained voice descriptions provided by listeners were also aligned to judgements of threat and intent to harm, with voices described with terms associated with anger and fear in the EmoLex database assigned higher overall threat and intent ratings by listeners.

7.3. Discussion of results

One core motivation for the work presented in this thesis was to address the issue highlighted in the UK Parliamentary Office of Science and Technology report on forensic language analysis surrounding the disjointed relationship between jurors and linguistic experts in relation to threat assessment. The report stated that "jurors expect certain procedures to be possible which experts assert are not, such as personality analysis, determining truth and falsity, and assessing threat in speech intonation (although this is a research interest)" (POST, 2015:3). It was pointed out in Chapter 1 that a core aim of work of the type presented in this thesis was to address this apparent

lay-listener belief that aspects of voice can be used to determine threat. This aim was specifically linked to the phonetic features investigated in the research chapters of this thesis, as summarised above in Section 7.2 of this chapter.

It should be stated at the outset of this discussion that research findings should **not** be interpreted as attempting to illustrate any kind of automaticity regarding links between aspects of voice and spoken threats, particularly if applied directly to a given case without the use of a linguistic expert. The working framework outlined at the end of Chapter 2 was designed to guard against over-interpretation of the findings in this thesis by delimiting the scope of the research to a narrower set of threat criteria. As previously stated, a research project of this size and scope was not able to deal with all the different permutations from the general threat typology shown in Figure 2.4. The goal of the thesis was, under the working threat framework, to evaluate the types of judgements that listeners make about indirect threats from anonymous or unfamiliar speakers. When contextual information was introduced, it was done in the form of framing the experiments within the context of potential bomb threats made to emergency service institutions. Applying the results of this thesis to the interpretation of threat judgements in, for example, a stalking case with a specific level and direction of personal relationship between a threatener and a victim, would therefore clearly not be appropriate. However, I argue here that the range of findings presented through the thesis can be useful within similar real-world contexts to which the experimental designs aimed to mirror.

Perhaps the most prominent finding in this thesis with direct potential for real-world application was the result from the experiment in Chapter 6 which highlighted very little listener agreement with respect to judgements of how threatening and intentful speakers

sounded. This was particularly noteworthy given the previously-highlighted concern surrounding the idea that listeners will "know a threat when they hear one" (Gingiss, 1986:153). By showing that a high level of disagreement existed between listeners evaluating the same voices, both with respect to the whole set of participants and to the 1000 random samples of 12 listeners tested, the experimental work in Chapter 6 should serve to promote caution in any kind of automatic assumptions surrounding the ways in which a potentially threatening utterance will be perceived by listeners. Considering this in view of the working threat framework and general threat typology outlined in Chapter 2, it can be argued that in the case of anonymous threats made to institutions where there is no established relationship between speaker and hearer, one level of information which could be used to interpret the meaning of a given utterance is stripped away or neutralised. Furthermore, in the context of emergency service control rooms, the threat could be received by any number of listeners, and the threatener would not be aware of who the recipient was aside from their role within the institution. The lack of agreement in listeners' judgements of threat and intent-to-harm shown in Chapter 6 could be important for contexts in which unfamiliar listeners are tasked with making decisions on, or judgements about, a threatener who is unknown to them. Furthermore, I argue that in settings such as an emergency service control room centre where multiple listeners could evaluate a threat without any prior warning or background information, it is important to be aware that there is the potential for disagreement between the judgements and evaluations that those listeners provide.

As the research in this thesis provides as an opposing stance to the argument that people will simply "know a threat when they hear one" (Gingiss, 1986:153), I wish to argue that it is better to begin from a position where it is assumed that listeners will disagree about how threatening or intentful speakers sound, and then see that their opinions and

judgements converge, rather than to make the assumption from the outset of any kind of investigation into crimes involving spoken threats that people will agree on the severity or the interpretation of a potentially threatening utterance. The experimental results from Chapter 6 serve as evidence to promote this view, albeit in the restricted setting outlined by the working threat framework for this thesis. The results also align with the analysis of speakers' productions of spoken threats conducted by Kelly (2014), which showed limited cross-speaker commonalities in productions of a so-called 'threatening tone of voice'.

Additionally, in Chapter 2, Section 2.2, a difference between the concepts of 'making' and 'communicating' threats was proposed, placing more focus on how listeners perceived potentially threatening utterances. Borrowing from Shuy's (1993:17) figure which highlights how ambiguity in meaning can lead to different interpretations, it can be argued that the work in this thesis has helped to show that a group of individual listeners will likely approach the task of interpreting utterances differently to one another. Individuals' prior beliefs and knowledge of the world can result in different listeners evaluating the same utterance produced by the same speaker with the same contextual information very differently. This was shown to be the case throughout the work presented in this thesis. Again, I argue against the idea that listeners will automatically interpret utterances in the same way as one another. Accepting this view would be a much better basis on which to conduct both real-world investigations and future research on spoken threats. The view that listeners will always agree on threat interpretations is also problematic for the 'reasonable person' position put forward by Danet et al. (1980), which argued that is a reasonable person would interpret an utterance as a threat, then a threat has been made. However, assuming that all listeners who took part in the experiment in Chapter 6 would qualify as being 'reasonable

people', the divergence in evaluative judgements highlights the potential problems with such an argument. The findings would instead support Gingiss' (1986) criticisms of this position, and reinforce the argument that conducting threat assessment based on a 'reasonable person' interpretation could be potentially problematic.

The results in this thesis have shown that throughout the experiments conducted, different aspects of voices could significantly influence listeners' judgements of how threatening speakers sounded. Although some of the findings were not consistent through the different experiments, such as the effect of speaker accent in Experiment 1 and Experiment 5, the influence of different aspects of voice on listeners' threat perception lends support for the view that when a threat is spoken, more than just the words can contribute to the meaning and interpretation of that threat. This would be particularly applicable to the types of indirect threats analysed in this project.

The range of judgements provided by listeners in each of the experiments presented in this thesis have spanned the entire range of the available judgements scale. While acknowledging that different listeners can interpret the same scale differently, the wide range of evaluations of the indirect threat stimuli shows that, particularly in the case of indirect threats, more than just the words spoken influence listeners' judgements. As highlighted in Chapter 2, these differences can be brought about by a wide range of factors, and the experimental work presented in this thesis has also highlighted the potential influence of certain phonetic variables on the perception of indirect spoken threats. However, while acknowledging Al-Shorafat's (1988) assertion than prosodic factors should be included in the conditions for defining spoken threats as a worthwhile aim, the work presented in this thesis has shown a lack of simplicity regarding the influence of such variables. However, the fact that variables such as vocal pitch, speaker

accent and voice quality were shown to influence listeners' judgements of how threatening speakers sounded in the experiments presented in this thesis should not be ignored. I argue that a reasonably safe assertion to make regarding this would be to highlight that certain phonetic variables have the potential to influence listeners' perceptions of how threatening speakers sound, and that they are among a series of factors, alongside the words spoken and the context, which can shape listeners' evaluations. The relative importance of these factors would likely differ on a case-by-case basis, and could be different if a different threat context outside of the working framework for this thesis was investigated.

However, while the results from this thesis have highlighted that different aspects of speakers' voices can influence listeners' perceptions of traits such as how threatening speakers sound, the results have also shown that not all cues will be the same for all listeners. I therefore wish to argue that the results in this thesis should serve to promote awareness on the part of those tasked with investigating crimes involving spoken threats that different aspects of speakers' voices, such as how high-pitched a voice was perceived to be, could affect listeners' perceptions of threat in unfamiliar speakers. It is important that such judgements, if and when they are highlighted, do not bias investigations into spoken threat crimes. After all, factors such as the pitch of a speaker's voice or a speaker's accent do not objectively reflect a speaker's level of intent to cause harm. It would also be incorrect to assume that all listeners have the same evaluative biases of speakers' voices, but it is hoped that the results of this thesis could, at the very least, help to raise some awareness of the potential dangers of allowing such judgements to prevail over more objective evidence in cases involving spoken threats.

In addition to the real-world implications discussed above, it is hoped that the research presented in this thesis will also help to advance knowledge of how aspects of voice can affect threat evaluation and person perception from a more theoretical perspective. Such outcomes, in accordance with McIntyre and Price's (2018) assertion about the value of research which lacks direct real-world impact, were also an integral aim of the research presented in this thesis.

Research throughout this thesis has shown that in any experiment where an effect was seen for either perceived pitch or measured pitch on listeners' evaluations of how threatening speakers sounded, the direction of the effect was that speakers with lower pitched voices were perceived to sound more threatening. These findings support previous literature (see, for example, Ohala, 1984; Tusing and Dillard, 2000; McAleer et al., 2014; Mileva et al., 2018), which have highlighted a relationship between lowered pitch and traits such as dominance and aggression. Although a link between increased threat perception and lowered pitch, either perceived or measured, was not shown in every experiment, the direction of the effects were consistent with the frequency code hypothesis first discussed in Ohala (1984). The work in this thesis helps advance work in this area by applying such a hypothesis to contentful utterances rather than stimuli made up of, for example, low-pass filtered speech or elongated vowel sounds. Both of these points reflect the ideas put forward by Banzinger and Scherer (2005) and Gussenhoven (2004) that the frequency code hypothesis can be socially constrained, and that different listeners may bring different biases and views to evaluative judgements of speakers' voices and the contributing linguistic features that may drive these judgements.

This view would also apply to the other variables investigated in this thesis, such as speaker accent and voice quality. With regards to speaker accent, the divergence between the results in Experiments 1 and 5 with respect to the effect of speaker accent on listeners' threat evaluations shows that while accent is a potentially influencing factor, this was not always the case. Future research in this area could work on developing the methodologies used in this thesis and applying them to both more accents and listeners from different social and demographic groups. The latter would arguably be useful for research into varieties which are linked to historically negative stereotypes. For example, in Chapter 6, threat evaluations of the Northern Irish speakers producing indirect bomb threats did not significantly differ from the other two accent groups, despite historical connections between bomb attacks and the period of The Troubles in Northern Ireland. However, the mean age of the participants who took part in Experiment 5 was 20 years old, meaning that most of the participants were not alive for the vast majority of the period of conflict in Northern Ireland. It could be argued that had a group of older listeners been tested, the results could have been different.

Additionally, had the experiment in Chapter 6 been conducted with listeners from Northern Ireland, then different results would also have been expected to the ones found in Chapter 6, following Coupland and Bishop's (2007) finding that people tend to rate accents similar to their own more favourably. Furthermore, it is also acknowledged that the social demographic of participants who were predominantly White British and attending university does not reflect the demographic makeup of the wider population. Repeating the experiments with listeners of different ages and from different social backgrounds would be a useful addition to the studies presented in this thesis, although the availability of, and access to, such participant groups remains somewhat problematic.

Methodologically, one area of research from this thesis which yielded promising results was the use of listeners' own evaluations as markers for judging the effect of various linguistic and phonetic variables on evaluations of how threatening speakers sounded. Given the findings throughout that listeners' evaluations of aspects of speakers' voices such as pitch and speaking tempo were often not in line with measured acoustic correlates, it can be argued that using listeners' own judgements of these dimensions of speakers' voices as fixed effects provided a more accurate assessment of the factors that caused them to infer greater or lesser levels of threat and intent to harm. This approach could be used in other studies examining the effect of different aspects of voice on listeners' evaluative judgements, as it is arguably less driven by the researcher and more by the views of individual listeners. As the research in this thesis has shown that it is possible for the same utterance production to be judged as sounding, for example, 'slow' by one listener and 'rapid' by another, it is argued that using listeners' own evaluative judgements as fixed effects on ratings for traits such as threat to harm is a worthwhile process for future research on voice evaluations.

In addition to examining the influence of phonetic variation on listeners' threat evaluations, the second main strand of research presented in this thesis was an investigation of how accurately listeners could describe specific aspects of speakers' voices. This analysis was conducted in relation to speakers' regional accent (Chapters 3 and 6), voice quality (Chapter 4), vocal pitch (Chapters 4, 5 and 6), speaking tempo (Chapters 4 and 5), alongside other unconstrained descriptions of speakers' voices (Chapters 5 and 6). The motivation for conducting this analysis was to critically assess the usefulness of documents which aim to elicit descriptions of speakers' voices, such as the National Counter Terrorism and Security Office (NCTSO) bomb threat checklist.

One practical suggestion for the implementation of these research findings would be the creation of a document which could provide linguistically-informed guidance to police officers and other legal practitioners tasked with using earwitness evidence elicited from documents such as the NCTSO bomb threat checklist. It can be argued that the real danger in obtaining such evidence is not the evidence itself, but the weight attached to the evidence by those investigating crimes which involve earwitness evidence. Of course, an ideal standard in this area would be the creation of a universally accepted standard earwitness elicitation document, compiled jointly by both police officers and linguists. However, given the differences in the earwitness frameworks that do exist (see, for example, Handkins and Cross, 1985; NCSTO, 2016), alongside the potential for earwitnesses to be asked direct verbal questions about the voices of perpetrators during the course of an investigation (Nolan, 2003), this seems an unlikely and unrealistic aim at this stage. The creation of guidance for investigators surrounding the interpretation of earwitness evidence could hopefully help to address the issue stated at the end of Chapter 6, that:

> *"[I]t is hoped that the availability of systematically-collected data of the sort described above* [the research findings from Chapter 6] *will serve to encourage more discriminating, better-informed evaluations of the utility of earwitnesses' voice descriptions on the part of members of the law enforcement and intelligence communities."*

In Table 7.1, below, I use the findings from the research presented in the experimental chapters of this thesis to suggest a series of guidance notes which could be provided to police officers or other relevant professionals tasked with eliciting descriptions of speakers' voices from earwitnesses. The terms that are provided on the NCTSO bomb

threat checklist under the section entitled "caller's voice" are used as a basis for this guidance, and the advisory notes are designed as an initial attempt to help those tasked with using earwitness evidence within the legal sphere. The work in Table 7.1 is not designed to be something which is finalised or concrete, and it is hoped that it could be used to promote further discussion between linguists and police officers about both the collection and implementation of earwitness evidence in criminal investigations. It should also be stated that the guidance primarily refers to unrecorded crimes involving speakers' voices and earwitness evidence. Should a recording be available for speaker profiling, then the advice and assessments of expert phoneticians should be sought, rather than the investigators solely relying on either descriptions provided by linguistically untrained earwitnesses or intuitions formed by listening to the material.

Descriptor(s)	Research findings	Chapter	Advisory notes
Deep **High-pitched**	Listeners were capable of more accurately labelling voices as 'deep' or 'high-pitched' when the voice in question had measured F0 values which were objectively high or low. Higher disagreement existed between listeners when they evaluated more 'regular' pitched voices.	Chapter 5	*If an earwitness states that a voice sounds either **deep** or **high-pitched**, be aware that research has shown that some listeners are more capable than others of accurately identifying how high-pitched a speakers' voice is in line with measured phonetic correlates. Caution is advised in assuming the automatic correctness of such judgements without further testing of the evaluative abilities of the witness. Vocal pitch can also be influenced by emotional state, such as whether the speaker was shouting, and technical*
	Limited agreement existed between listener evaluations of how high-pitched a speaker's voice was when a scale was used to elicit judgements.	Chapter 6	
	Some listeners were inherently better at evaluating pitch using a scalar judgement score than other listeners.	Chapter 6	

	Even phonetically trained experts can struggle to accurately perceive how high-pitched a speakers' voice is in line with F0 measurements, with other aspects of voice such as phonation quality also influencing judgements alongside measured F0.	Kirchhübel (2018, personal communication).	*factors such as whether the speaker was talking on the telephone.*
Slow **Rapid**	Listeners tasked with assessing speaking tempo using the NCTSO **slow** and **rapid** descriptors showed more agreement in relation to measured speech rate (syllables per second) with voices at the higher and lower end of the expected normal range (see Gold, 2014). Voices towards the middle of the normal range were often described as 'slow' by one listener and 'rapid' by another. Listeners linked read aloud speech with slower speaking tempo.	Chapter 5 Chapter 4	*Research has shown that sometimes, the same voice can be described as sounding **slow** by some listeners and **rapid** by other listeners. This is particularly true of speakers talking at what could be considered a 'regular' tempo. Slowed speech rate is one property, alongside a lack of hesitations, pauses and repetitions, of read-aloud speech. It may be useful to explore whether listeners were able to pick up on this as a cue to the speech being read from a pre-prepared text rather than being spontaneous. Playing both types of speech (spontaneous and read aloud) to an earwitness may be one way to test whether their judgements of speaking tempo are likely to be accurate.*

Calm **Angry**	The International Association of Forensic Phonetics and Acoustics (IAFPA) code of practice strongly discourages members from conducting psychological profiling or sincerity assessments. This could include inferring information about a speaker's emotional state.	IAFPA Code of Practice (IAFPA, 2004)	*Experimental research has shown that when tasked with evaluating speakers' voices using the NCTSO bomb threat checklist, listeners frequently used the **calm** and **angry** labels. However, these offer information about emotional state rather than descriptive aspects of voice. Professionals working in voice analysis are advised against inferring information about speakers' emotional states. It may, however, be useful to explore the reasons behind the use of a **calm** or **angry** descriptor with an earwitness. Such a discussion may, for example, reveal whether a speaker was shouting or not. **Calm** speech could also be linked to factors such as slowed tempo and read-aloud speech. Such information could be more linguistically-relevant than assuming that a speaker was either **calm** or **angry**.*
	Both the **calm** and **angry** labels were frequently assigned by listeners in the experiments presented in this thesis.	Chapters 4 and 5	
	Calm and **angry** were consistently linked to perceptions of lower and higher threat, respectively.	Chapters 4 and 5	
	A potential link was identified between **calm** attributions and read-aloud, non-shouted speech.	Chapter 4	

Regional accent (confined in this case to accents of the United Kingdom)	Listeners could identify London Cockney and Received Pronunciation accent guises with a relatively high degree of accuracy.	Chapter 3	*Research has shown that it is not always easy for a listener to accurately describe someone's regional accent. The background of the listener is also likely to affect their ability to describe an accent. For example, Northern Irish and Scottish accents are more likely to be confused by listeners from England compared to listeners from Northern Ireland or Scotland. Caution is advised in any over-reliance on earwitness accent descriptions. Non-specific accent descriptions such as "not from around here" or "Southern" could also be as useful in investigations as those which provide a specific geographical location.*
	Confusion was shown between Scottish and Northern Irish accents by listeners from England.	Chapter 6	
	Listeners were comparatively more unsure when describing Standard Southern British English compared to other accents such as Northern Irish English.	Chapter 6	
Nationality	Listeners provided a wide range of descriptions for foreign-accented speakers who had English as their second language.	Chapter 6	*Extreme caution should be advised when using earwitness assessments of a speaker's **nationality** if that speaker has communicated in English only. Research has shown that listeners have very limited success when assessing the nationality of a foreign speaker who speaks in English. Assessing a speaker's **nationality** may be more useful if the earwitness*

			recognizes that a particular foreign language has been spoken by a perpetrator. In such cases, it would be advised to check that the listener has sufficient knowledge to recognize the language being spoken.
Male/Female	When speaking in what can be considered a regular voice, speaker sex attributions were relatively accurate. However, falsetto and whispered speech caused listeners to misidentify the sex of the speaker on occasions.	Chapter 4	*Sometimes, male speakers can attempt to sound like a female speaker by using falsetto voice as a form of disguise. Research has also shown that it can also be difficult for listeners to accurately determine the sex of a speaker if that speaker is whispering. Consider these factors when assessing earwitness evaluations of speaker sex.*
Crying **Clearing throat** **Slurred** **Stutter** **Lisp** **Laughter**	These descriptors were used extremely infrequently by listeners when they were describing speakers' voices in the experiments presented in this thesis. None of the features were present in the samples, and it can be argued that listeners were providing accurate descriptions by not using them.	Chapters 4 and 5	*Research has shown that listeners are unlikely to use these descriptors when the features are not present in the speech samples they were evaluating.*

Unconstrained voice descriptions	When provided, these were highly variable between listeners. Unconstrained descriptors of speakers producing bomb threat utterances often made use of emotional descriptors which were linked to the perceived level of threat in the utterance.	Chapters 5 and 6 Chapter 6	*Research has shown that these kinds of descriptors can be highly variable between listeners, sometimes providing useful information and sometimes not providing useful information. When used to assess threateners' voices, these descriptors have been shown to link to perceived emotional states and threat levels rather than specific aspects of speakers' voices. The usefulness of these descriptors should be considered on a case-by-case basis.*

Table 7.1 – A series of potential guidance notes for law enforcement professionals tasked with evaluating earwitness evidence based on the findings from this thesis

As previously stated, the work detailed in Table 7.1 is designed to provide a series of advisory notes for those tasked with implementing lay-listener voice evaluations from the NCTSO bomb threat checklist document into practice during investigative work. One hope for such work is that it provides a base document which could be amended, edited and improved through further research. Furthermore, it is also hoped that such a document could act as a catalyst for facilitating further discussions between linguistic researchers, police officers and policy makers regarding the use of earwitness evidence, particularly related to spoken threats. This is similar to the work proposed by Smith et

al. (2018), and it is hoped that projects such as these could further advance the usability of earwitness evidence by law enforcement agencies.

One further possibility to help improve the usability of earwitness evidence highlighted both in Chapter 6 and in the advisory notes in Table 7.1 was the creation of material designed to test the abilities of a given earwitness to determine certain aspects of speakers' voices. The development of such material would require more research and development, but I argue that given the emerging availability of voice databases designed with forensic relevance in mind (see, for example, the Dynamic Variability in Speech (DyViS) project (Nolan et al., 2006) and the West Yorkshire Regional English Database (WYRED) (Gold et al., 2018) projects), the aim of developing test materials designed to ascertain the voice description abilities of an earwitness is not altogether unrealistic. For example, a research team at Nottingham Trent University, UK, are currently using the DyViS database for the purpose of earwitness voice description research (see Smith et al., 2018). This shows that precedent already exists for the use of such material in voice description tasks. Furthermore, one of the principle aims of the WYRED project is identified as being "to build a database of British English speech which will be made publically available for wider use by researchers and any other interested parties" (WYRED, 2018). Here, I argue that these databases of speakers' voices could be made available for the testing of earwitness voice description accuracy, especially given that ground truth information about the phonetic composition of the voices would be known in advance to the testers. I also argue that it should be incumbent on linguists to design and create such a test, and make it both accessible and available to investigators tasked with using earwitness voice description evidence in their work.

7.4. Scope for future work

Given the inherently developmental nature of the work presented in this thesis, there is plentiful scope for future work which could build on this research. The advantages and limitations of conducting experimental perceptual research was discussed in Chapter 2, and future work could look to expand the designs used in this thesis. It is somewhat difficult to create controlled experimental stimuli and conduct ethical voice evaluation experiments in a way that more accurately replicates real-world practice. However, future work on listeners' evaluations of spoken threats could take the findings from the experiments in this thesis and test them in different scenarios. One future possibility for bridging the gap between experimental and real-world conditions would be to use a controlled, experimental setting to test people who are used to dealing with real-world speaker and voice evaluation tasks in their professional lives. One such group of people could be emergency call handlers, whose profession involves dealing with calls which are wide-ranging, unpredictable and often highly sensitive. While such a participant group would clearly be aware that they were taking part in an experiment, it would be expected that their prior background, professional experience and knowledge of the world could influence their evaluative judgements of speakers' voices. It is hoped that the initial findings from this thesis relating to threat evalautions and earwitness voice descriptions could be used to spark meaningful dialogue between police officers and emergency call handlers about how to further advance the findings and make them more applicable to everyday real-world law enforcement tasks. By engaging in such a dialogue, it is also hoped that the knowledge of those people who work in law enforcement and investigation could help to further advance the kinds of findings presented throughout this thesis. This would be one way of using the research presented in this thesis to create relevant and meaningful impact.

There are two other factors which could also be addressed through further research aiming to improve on the kinds of experimental designs shown in this thesis. The first issue would be the difference between the way in which groups and individuals make decisions on how threatening and intentful speakers sound. This issue was raised in Chapter 6 when the analysis of variation in the threat and intent to harm scores of groups of 12 listeners was conducted, and showed considerable variation both within and between the random subsets of 12 listeners with respect to threat and intent to harm judgements. However, research has shown (see, for example, Myers and Lamm (1976)) that the dynamics of group deliberation means that judgements made by groups are not the same as those made by individuals. Given the lack of agreement shown between listeners' evaluations of threat and intent to harm in Chapter 6, it would be worthwhile to consider a similar task but obtain evaluative decisions made by a group as opposed to individual listeners. This would facilitate an assessment of whether or not the dynamics of a group, akin to a jury panel, would change the degree of listener agreement over how threatening and intentful speakers sounded compared to the individuals' results presented in Chapter 6 of this thesis.

The second area which could improve understanding of listeners' threat assessments through experimental research of the kind presented in this thesis would be the variation of evidential information provided to participants making evaluative judgements. In a follow-up to Dixon, Mahoney and Cocks' (2002) work investigating the influence of speaker accent on attributions of guilt, Dixon and Mahoney (2004) conducted a similar experiment but also presented listeners with strong and weak evidential information against the suspect. Encouragingly, Dixon and Mahoney (2004) found that evidence strength was a significant predictor of guilt attributions, with the effect of speaker accent being non-significant. A similar follow-up study would make a useful addition to

the research presented in this thesis. Varying the level of evidence against a 'threat perpetrator' in an evaluative judgement task would facilitate an assessment of how important aspects of voice are in shaping listeners' threat and intent to harm evaluations alongside other types of evidence. Such 'evidence' could also be based on a real-world case, such as the Middlesbrough Crown Court trial outlined by Watt, Kelly and Llamas (2013), and discussed in Chapter 1 of this thesis. Given that prescribed contextual information influenced listeners' judgements in Chapter 4, there would be methodological scope for such a design in future research on spoken threats.

In the later experimental chapters presented in this thesis, research was conducted which focussed on establishing links between phonetic aspects of speech and previous work on linguistic features of spoken threats. One such example from the research in Chapter 5 was establishing whether placing emphasis through raised F0, raised intensity and increased duration on the word 'will' would influence both productions and perceptions of a 'threatening tone of voice'. This analysis was conducted owing to previous work (Gales, 2010; Nini, 2017) having linked the use of modal 'will' to increased commitment on the part of threateners. It is hoped that future linguistic analysis of spoken threats could take this approach and apply an integrated linguistic-phonetic approach to further develop knowledge surrounding both the production and perception of spoken threat utterances.

With respect to the voice description strand of the project, the previously-highlighted need for more work to investigate how earwitnesses describe and evaluate the voices of perpetrators is reiterated here. There is plentiful scope for more research in this area, looking at both the perception of specific features such as pitch, speech rate, intonation, voice quality and speaker accent, alongside more general research looking at the effects

of aspects such as providing different instructions and memory recall on the quality of listeners' descriptions. Working alongside law enforcement agencies should also be a priority for linguists interested in voice description research to ensure that the results from research projects can have meaningful practical applications.

Chapter 8 - Conclusion

8.1. Conclusion

The body of work presented in this thesis has taken an experimental approach to examining listeners' perceptions of how threatening speakers sound when producing a range of indirect threat utterances. This was conducted with particular focus on how phonetic aspects of voice could influence perceptual judgements of traits such as threat and intent to harm. It also focussed on earwitness voice descriptions and the abilities of linguistically-untrained listeners to accurately describe speakers' voices.

In summary, I see the findings of this thesis as an extension to initial work exploring the influence of aspects of speakers' voices on listeners' inference of threat and intent to harm (see, for example, Watt, Kelly and Llamas (2013)). Given the multitude of environments and contexts in which threats can be made, and the unknown and probably very large number of variables which could influence how they are perceived, it would clearly be unwise to over-generalise the findings of this thesis to any genuine situation involving spoken threats.

However, I align with the assertion made by Watt, Kelly and Llamas (2013) that although the subjectivity surrounding the inference of phenomena such as threat and intent to harm makes drawing generalised conclusions difficult, research of the kind presented in this thesis can help provide both an empirical grounding and a measure of objectivity to a situation where both of these have, so far, been lacking.

Another aim of the work in this thesis was to generate empirical data as a basis upon which to make recommendations about how earwitness evidence can be better collected and later deployed by those tasked with gathering such information. It is hoped that this approach could be helpful in guarding against the use of erroneous, redundant, vague or otherwise low-value earwitness testimony in the sphere of criminal investigation. At the very least, it is hoped that the availability of systematically-collected data of the sort described above will serve to encourage more discriminating, better-informed evaluations of the utility of earwitnesses' voice descriptions and evaluations on the part of members of the law enforcement and intelligence communities.

References

Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, 15(1), 38-59.

Agha, A. (2007). *Language and social relations*. Cambridge: Cambridge University Press.

Allport, G. W., & Cantril, H. (1934). Judging personality from voice. *The Journal of Social Psychology*, *5*(1), 37-55.

Al-Shorafat, M. O. (1988). Indirect Threats. *Word*, 39(3), 225-227.

Antoniou, M. (2010). *Scale intensity (energy) with output*. Northwestern University. Script for Praat.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37(5), 715-727.

Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bach, K., & Harnish, R. (1979). *Linguistic communication and speech acts*. MA: MIT Press.

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current directions in psychological science*, 8(2), 53-57.

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech communication*, *46*(3-4), 252-267.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

BBC News. (2016). *EgyptAir: Man held after using fake suicide belt*. BBC News website. Online resource: https://www.bbc.co.uk/news/world-middle-east-35921579. Accessed 13th September 2018.

BBC News. (2018). *'Ex-soldier' raiding Home Counties houses at gunpoint.* BBC News website. http://www.bbc.co.uk/news/uk-england-42863601. Accessed 28 February 2018.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). MIT Press.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.22.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82, 55-59.

Breitenstein, C., Lancker, D. V., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition & Emotion*, *15*(1), 57-79.

Broeders, A., & van Amelsvoort, A. (2001). A practical approach to forensic earwitness identification: constructing a voice line-up. *Problems of Forensic Sciences, 47,* 237-245.

Carter, N. R. (2010). We Shall Be Watching You, You're Going to Die, and Other Threats: A Corpus-Based Speech Act Approach. *UTA Working Papers in Linguistics*, (3), 48-61.

Catford, J. C. (1988). *A practical introduction to phonetics*. Oxford: Clarendon Press.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2009). Encoding emotions in speech with the size code. *Phonetica*, 65(4), 210-230.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Coupland, N. & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics*, 11(1), 74-93.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Culpeper, J., Iganski, P., & Sweiry, A. (2017). Linguistic impoliteness and religiously aggravated hate crime in England and Wales. *Journal of Language Aggression and Conflict*, *5*(1), 1-29.

Danet, B., Hoffman, K. B., & Kermish, N. C. (1980). Threats to the life of the president: An analysis of linguistic issues. *Journal of Media Law and Practice*, 1(2), 180-190.

Dimos, K., Dick, L., & Dellwo, V. (2015). Perception of levels of emotion in speech prosody. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow.

Dixon, G., Foulkes, P. & LaShell, P. (2013). Lay listener judgments of pitch. Paper presented at the *International Association of Forensic Phonetics and Acoustics (IAFPA) conference*, Tampa, Florida, USA.

Dixon, J. A. & Mahoney, B. (2004). The effect of accent evaluation and evidence on a suspect's perceived guilt and criminality. *The Journal of Social Psychology*, 144(1), 63-73.

Dixon, J. A., Mahoney, B. & Cocks, R. (2002). Accents of guilt? Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology*, 21(2), 162-168.

Dorling, D. (2010). Persistent North-South divides. In N. M. Coe and A. Jones (Eds.) *The Economic Geography of the UK*. London: Sage, pp. 12-28.

Douglas, J., Burgess, A. W., Burgess, A. G., & Ressler, R. K. (2013). *Crime classification manual: A standard system for investigating and classifying violent crime*. John Wiley & Sons.

Eckert, H. & Laver, J. (1994). *Menschen und ihre Stimmen: Aspekte der Vokalen Kommunikation*. Weinheim: Beltz Psychologie Verlags Union.

Edwards, J. R. (1982). Language attitudes and their implications among English speakers. In E. B. Ryan & H. Giles (eds.), *Attitudes towards language variation*. London: Arnold. 20–33.

Eriksson, A. (2005). Tutorial on forensic speech science. In *Interspeech*, Lisbon, Portugal.

Fecher, N. (2015). *Praat pitch alteration script*. Department of Language and Linguistics, University of York. Script for Praat.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal behaviour*, 69(3), 561-568.

Fisher, B. (2018). *Fundamental frequency judgement over studio and telephone recordings, and across laryngeal heights: a perceptual study*. Unpublished Master's dissertation. University of York.

Fraser, B. (1975). Warning and threatening. *Centrum*, 3(2), 169-180.

Fraser, B. (1998). Threatening revisited. *International Journal of Speech, Language and the Law*, 5(2), 159-173.

Freese, J., & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society*, *27*(2), 195-219.

French, P., Harrison, P., & Windsor Lewis, J. (2007). R v John Samuel humble: The Yorkshire ripper hoaxer trial. *International Journal of Speech Language and the Law*, *13*(2), 255-273.

French, P., & Watt, D. (2018). Assessing research impact in forensic speech science casework. In McIntyre, D., & Price, H. (Eds.). *Applying Linguistics: Language and the Impact Agenda*. Oxford: Routledge.

Gales, T. (2010). *Ideologies of Violence: A Corpus and Discourse Analytic Approach to Stance in Threatening Communications*. PhD Thesis. University Of California, Davis.

Gales, T. (2011). Identifying Interpersonal Stance in Threatening Discourse: An Appraisal Analysis. *Discourse Studies*, 13(1), 27-46.

Gales, T. (2012). Linguistic analysis of disputed meanings: Threats. In C.A. Chapelle (Ed), *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell.

Gales, T. (2015). The stance of stalking: a corpus-based analysis of grammatical markers of stance in threatening communications. *Corpora,* 10(2), 171-200.

Gales, T. (2016). Threatening Stances: A corpus analysis of realized vs. non-realized threats. *Language and Law*, 2(2), 1-25.

Giles, H. 1970. Evaluative reactions to accents. *Educational Review*, (22), 211–227.

Giles, H., & Billings, A. C. (2004). Assessing language attitudes: Speaker evaluation studies. *The handbook of applied linguistics*, 187-209.

Gingiss, P. (1986). Indirect threats. *Word*, 37(3), 153-158.

Gobl, C., & Chasaide, A. N. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech communication*, *40*(1-2), 189-212.

Gonzalez, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of phonetics*, 32(2), 277-287.

Gold, E. A. (2014). *Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters.* PhD thesis, University of York.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London/New York: Academic Press.

Grabe, E., Post, B. & Nolan, F. (2001). *The IViE Corpus*. Department of Linguistics, University of Cambridge. http://www.phon.ox.ac.uk/IViE.

Greenawalt, K. (1989). *Speech, crime, and the uses of language*. Oxford: Oxford University Press.

Griffiths, M. (2012). 'Did he have an accent?' Forensic speaker descriptions of unknown voices. In *Proceedings of The International Association of Forensic Linguists' Tenth Biennial Conference*. Birmingham, UK: Aston University.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Speech Prosody 2002*, International Conference.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Handkins, R. E., & Cross, J. F. (1985). Can a voice lineup be too fair?. In *annual meeting of the Midwestern Psychology Association, Chicago, IL*.

Hollien, H., & Michel, J. F. (1968). Vocal fry as a phonational register. *Journal of Speech, Language, and Hearing Research*, *11*(3), 600-604.

Hothorn, T., Bretz, F., Westfall, P., & Heiberger, R.M. (2008). *Multcomp: Simultaneous Inference in General Parametric Models*. http://CRAN. R-project.org. R package version 1.0-0.

Hudson, T., De Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. In *Proceedings of the 16th International Congress of Phonetic Science*, Saarbrücken: Germany, pp.1809-1812.

Hughes, A., Trudgill, P., & Watt, D. (2013). *English accents and dialects: An introduction to social and regional varieties of English in the British Isles*. Routledge.

International Association of Forensic Phonetics and Acoustics. (2004). *Code of Practice*. Online resource: https://www.iafpa.net/the-association/code-of-practice/. [Accessed 5th March 2018].

Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education*, 38(12), 1217-1218.

Kalin, R. (1982). The social significance of speech in medical, legal and occupational settings. In E.B. Ryan and H. Giles (Eds.), *Attitudes towards language variation: Social and applied contexts*. London: Arnold, pp. 148-163.

Kaplan, J. P. (2016). Case report: Elonis v. United States. *International Journal of Speech, Language & the Law*, 23(2), pp. 275-292.

Kelly, S. (2014). *An Analysis of the Prosodic Properties of Neutrally-Worded Threat Productions*. Unpublished Master's dissertation, University of York.

Kirchhübel, C. (2013). *The acoustic and temporal characteristics of deceptive speech. PhD thesis*, University of York.

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, 46(1-3), 117-125.

Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *International Journal of Speech language and the Law*, 4, 48–83.

Künzel, H. J. (2001) 'Beware of the "telephone effect": the influence of telephone transmission on the measurement of formant frequencies', *International Journal of Speech, Language and the Law*, 8(1): 80–99.

Langdridge, D., & Hagger-Johnson, G. (2009). *Introduction to research methods and data analysis in psychology*. London: Pearson Education.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. Academic Press.

Larner, S. (2015). From intellectual challenges to established corpus techniques: introduction to the special issue on forensic linguistics. *Corpora*, *10*(2), 131-143.

Laver, J. (1994). *Principles of phonetics*. New York, NY: Cambridge University Press.

Laver, J. (1980). The Phonetic Description of Voice Quality. Cambridge: Cambridge University Press.

Law, J., & Martin, E. A. (2009). *Oxford dictionary of law* [Seventh Edition]. Oxford: Oxford University Press.

Lindh, J. (2006). Preliminary F0 statistics and forensic phonetics. *Proceedings of the 15th annual International Association of Forensic Phonetics and Acoustics conference*, Department of Linguistics, Göteborg University: Sweden.

Mackenzie Beck, J. (2007). *Vocal Profile Analysis Scheme: A User's Manual*. Edinburgh: Queen Margaret University College-QMUC, Speech Science Research Centre.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS One*, 9, e90779.

McIntyre, D. & Price, H. (2018). Linguistics, language and the impact agenda. In McIntyre, D., & Price, H. (Eds.). *Applying Linguistics: Language and the Impact Agenda*. Oxford: Routledge.

Meloy, J.R., Hart, S.D., & Hoffman, J. (2013). Threat assessment and threat management. In J. R Meloy & J. Hoffmann, (Eds). International handbook of threat assessment. Oxford: Oxford University Press, pp. 3-17.

Milburn, T. W., & Watman, K. H. (1981). *On the nature of threat: A social psychological analysis*. New York: Praeger.

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 128-138.

Mohammad, S. M., & Turney, P. D. (2013a). NRC emotion lexicon. *NRC Technical Report*. National Research Council, Canada.

Mohammad, S. M., & Turney, P. D. (2013b). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, *29*(3), 436-465.

Montgomery, C. (2007). *Northern English Dialects: A perceptual approach*. PhD thesis, University of Sheffield.

Montgomery, C. (2015). Borders and boundaries in the north of England. In R. Hickey (ed.) *Researching Northern English*. Amsterdam: Benjamins, pp. 345-368.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 855-869.

Myers, D. & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83(4), 602-627.

Napier, M., & Mardigian, S. (2003). Threatening messages: The essence of analyzing communicated threats. *Public Venue Security*, 16-19.

National Counter Terrorism Security Office. (2016). Bomb threat checklist. Online resource: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/552301/Bomb_Threats_Form_5474.pdf. [Accessed 15 December 2017.]

Nini, A. (2017). Register variation in malicious forensic texts. *International Journal of Speech, Language & the Law, 24*(1).

Nolan, F. (2003). A recent voice parade. *International Journal of Speech Language and the Law, 10*(2), 277-291.

Nolan, F., & Grabe, E. (2013). Preparing a voice lineup. *International Journal of Speech Language and the Law, 3*(1), 74-94.

Nolan, F., McDougall, K., & Hudson, T. (2013). Effects of the telephone on perceived voice similarity: implications for voice line-ups. *International Journal of Speech, Language & the Law, 20*(2), 229-246.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education,* 15(5), 625-632.

Office for National Statistics. (2018). *Crime Survey for England and Wales, 2018*. Online resource: http://www.crimesurvey.co.uk/en/SurveyResults.html. [Accessed 14th September 2018.]

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, 41(1), 1-16.

Ogden, R. (2009). *Introduction to English Phonetics*. Edinburgh: Edinburgh University Press.

Parliamentary Office of Science and Technology. (2015). *Forensic Language Analysis*. POSTnote number 509. Online resource: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0509 [Accessed 13th September 2018].

Pear, T. H. (1931). *Voice and personality, as applied to radio broadcasting*. Oxford, England: Wiley.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of sociolinguistics*, *11*(4), 478-504.

Preston, D. C. (2018). Perceptual dialectology. In C. Boberg, J. Nerbonne and D. Watt (eds.) *The Handbook of Dialectology*. Oxford: Wiley-Blackwell, pp. 173-207.

Preston, D. R. (2002). Language with an Attitude. In J. K. Chambers, P. Trudgill and N. Schilling-Estes, (Eds). *The Handbook of Language Variation and Change*. Oxford: Blackwell, pp. 40-66.

Public Order Act. (1986). London: Her Majesty's Stationery Office. Online resource: http://www.legislation.gov.uk/ukpga/1986/64/pdfs/ukpga_19860064_en.pdf

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*, 283-296.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. Evolution and Human Behavior, 28, 340-344.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech communication*, *40*(1), 227-256.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Shen, C., & Watt, D. (2015). Accent categorisation by lay listeners: Which type of "native ear" works better? *York Papers in Linguistics Series 2, 14*, 106-131.

Sherrin, C. (2015). Earwitness evidence: the reliability of voice identifications. *Osgoode Legal Studies Research Paper Series. Paper 101*. Online resource: http://digitalcommons.osgoode.yorku.ca/olsrps/101. [Accessed 28 February 2018.]

Shuy, R. W. (1993). *Language crimes: The use and abuse of language evidence in the courtroom*. Oxford: Blackwell.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621-640.

Smith, H., Braber, N., Robson, J., Wright, D., & Kelly, S. (2018). *Developing a procedure for eliciting accurate detailed and consistent voice descriptions from lay witnesses.* Paper presented at the International Association of Forensic Phonetics and Acoustics (IAFPA) conference, Huddersfield, UK.

Solan, L. M., and Tiersma, P. M. (2015). Threats. *Speaking of Language and Law: Conversations on the Work of Peter Tiersma*, 223-229.

Storey, K. (1995). The language of threats. *International Journal of Speech, Language and the Law.* 2(1), 74 – 80.

Sykes, S., and Perring, R. (2016). *Panic at 25 schools across UK as sick caller warns 'bomb will take children's heads off'*. Express.com. Online resource: https://www.express.co.uk/news/uk/672926/Canterbury-school-bomb-threat-scare-terrorism-UK. Accessed 14th September 2018.

Tompkinson, J. (2016). *Accent evaluation and the perception of spoken threats*. Unpublished MSc dissertation. University of York.

Traunmüller, H., and Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. *Unpublished manuscript. [Available at http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf]*.

Trudgill, P. (1990). *The Dialects of England*. Oxford: Blackwell.

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. Perception, 45, 946–463.

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance. Human Communication Research, 26, 148-171.

Vaissière, J. (2005). Perception of intonation. In Pisoni & Remez (2005) *The handbook of speech perception*, 236-263.

Wales, K. (2000). North and south: An English linguistic divide? *English Today*, 16(01), 4-15.

Watson, K., & Clark, L. (2015). Exploring listeners' real-time reactions to regional accents. *Language Awareness*, 24(1), 38-59.

Watt, D. (2010). The identification of the individual through speech. In C. Llamas and D. Watt (eds.) *Language and Identities*. Edinburgh: Edinburgh University Press, pp.76-85.

Watt, D., & Burns, J. (2012). Verbal descriptions of voice quality differences among untrained listeners. *York Papers in Linguistics Series*, *2*, 1-28.

Watt, D., Kelly, S. & Llamas, C. (2013). Inference of threat from neutrally-worded utterances in familiar and unfamiliar languages. *York Papers in Linguistics*, (13), 99-120.

Watt, D., Kelly, S., Tompkinson, J., & Weinberg, K. (2016, Febuary) Anyone for menace? *Babel magazine, volume 14*, 18-23.

Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France. 7-10.

Xu, Y., & Kelly, A. (2010). Perception of anger and happiness from resynthesized speech with size-related manipulations. In *Proceedings of the 5th International Conference on Speech Prosody (SP2010)*. Chicago, IL.

Xu, Y., Lee, A., Wu, W. L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signalled by body size projection. *PLoS one*, 8(4), e62397.

Yamanaka , N. (1995). On indirect threats. *International Journal for the Semiotics of Law*, 8(2), 37-52.

Yarmey, A. D. (2001). Earwitness descriptions and speaker identification. *International Journal of Speech Language and the Law*, *8*(1), 113-122.

Yuasa, I. P. (2010). Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women?. *American Speech*, *85*(3), 315-337.