Evolutionary and Cognitive Approaches to Voice Perception in Humans: Acoustic Properties, Personality and Aesthetics.

Kristen K. Knowles

A thesis submitted for the degree of

Doctor of Philosophy

School of Natural Sciences, Psychology

University of Stirling

2014

DECLARATION

I, Kristen Kelley Knowles, declare that the thesis herein has been written by me and that it is the record of work carried out by myself. No part of this thesis has been previously submitted for another degree at the University of Stirling or other institutions.

Kristen Kelley Knowles

ACKNOWLEDGEMENTS

I would like to thank my parents, Douglas and Patricia Knowles, for their encouragement and unblinking optimism, and for firstly believing that I could rise to the challenge of doing a PhD. You raised a curious child (in more than one sense), a task which is unlikely to have always been pleasant.

Special thanks also go to my partner Neil for his love, unfailing reassurance and positivity. Your patience with me and my thesis likely exceeds my own. My parents' curious girl is now your curious other half. See above re: tasks which are not easy.

Heartfelt thanks also go to my supervisor, Dr. Tony Little, for his advice and guidance, and for being my cheerleader when I needed it. I had so much fun chasing interesting ideas with you.

Thanks to the University of Stirling, the Experimental Psychology Society (EPS), and the European Human Behaviour and Evolution Association (EHBEA) for providing me with resources to travel for seminars, courses and academic conferences. The skills and relationships I have developed as a result of this will be of immense value throughout my career.

Finally, my most profound thanks go to all of my study participants, who have been subjected to a battery of voice recordings over the past four years. They have my apologies as well as my gratitude.

ABSTRACT

Voices are used as a vehicle for language, and variation in the acoustic properties of voices also contains information about the speaker. Listeners use measurable qualities, such as pitch and formant traits, as cues to a speaker's physical stature and attractiveness. Emotional states and personality characteristics are also judged from vocal stimuli. The research contained in this thesis examines vocal masculinity, aesthetics and personality, with an emphasis on the perception of prosocial traits including trustworthiness and cooperativeness. I will also explore themes which are more cognitive in nature, testing aspects of vocal stimuli which may affect trait attribution, memory and the ascription of identity.

Chapters 2 and 3 explore systematic differences across vocal utterances, both in types of utterance using different classes of stimuli and across the time course of perception of the auditory signal. These chapters examine variation in acoustic measurements in addition to variation in listener attributions of commonly-judged speaker traits. The most important result from this work was that evaluations of attractiveness made using spontaneous speech correlated with those made using scripted speech recordings, but did not correlate with those made of the same persons using vowel stimuli. This calls into question the use of sustained vowel sounds for the attainment of ratings of subjective characteristics. Vowel and single-word stimuli are also quite short – while I found that attributions of masculinity were reliable at very short exposure times, more subjective traits like attractiveness and trustworthiness require a longer exposure time to elicit reliable attributions. I conclude with recommending an exposure time of at least 5 seconds in duration for such traits to be reliably assessed.

Chapter 4 examines what vocal traits affect perceptions of pro-social qualities using both natural and manipulated variation in voices. While feminine pitch traits (F_0 and F_0-*SD*)

were linked to cooperativeness ratings, masculine formant traits (D_f and P_f) were also associated with cooperativeness. The relative importance of these traits as social signals is discussed.

Chapter 5 questions what makes a voice memorable, and helps to differentiate between memory for individual voice identities and for the content which was spoken by administering recognition tests both within and across sensory modalities. While the data suggest that experimental manipulation of voice pitch did not influence memory for vocalised stimuli, attractive male voices were better remembered than unattractive voices, independent of pitch manipulation. Memory for cross-modal (textual) content was enhanced by raising the voice pitch of both male and female speakers. I link this pattern of results to the perceived dominance of voices which have been raised and lowered in pitch, and how this might impact how memories are formed and retained.

Chapter 6 examines masculinity across visual and auditory sensory modalities using a cross-modal matching task. While participants were able to match voices to muted videos of both male and female speakers at rates above chance, and to static face images of men (but not women), differences in masculinity did not influence observers in their judgements, and voice and face masculinity were not correlated. These results are discussed in terms of the generally-accepted theory that masculinity and femininity in faces and voices communicate the same underlying genetic quality. The biological mechanisms by which vocal and facial masculinity could develop independently are speculated.

CONTENTS

Chapter 1:
Introduction & Aims ... 1-18
1.1 The Mechanics of Speech Production ... 2
1.2 Links Between Voice and Face Perception ... 6
1.3 Physical, Aesthetic and Social Perception Based on Voices ... 7
1.3.1 Masculinity, Dominance and Antisocial Behaviour ... 7
1.3.2 Attractiveness ... 8
1.3.3 Personality Traits and Prosociality ... 11
1.4 Identity ... 13
1.5 Overview of Experimental Chapters ... 15

Chapter 2:
A Within-Subjects Investigation of Voice Perception Methodologies: Perceptions of Voice Attractiveness Using Vowels, Scripted and Spontaneous Speech ... 19-44
2.1 Introduction ... 19
2.2 Methods ... 25
2.2.1 Stimuli ... 25
2.2.2 Recording & Stimuli Manufacture ... 25
2.2.3 Measurements ... 26
2.2.4 Participants ... 27
2.2.5 Procedure ... 27
2.3 Results ... 28
2.3.1 Measurements Across Speech Types ... 28
2.3.1.1 Pitch and Pitch Variation ... 29
2.3.1.2 Formants ... 30
2.3.2 Ratings Across Speech Types ... 32
2.3.3 Relationships Between Vocal Measurements and Attractiveness Ratings ... 33
2.4 Discussion ... 34
2.4.1 Measurements Across Speech Types ... 35
2.4.1.1 Pitch and Pitch Variation ... 35
2.4.1.2 Formants and Harmonics ... 37
2.4.2 Ratings Across Speech Types ... 38
2.4.3 Measurements and Attractiveness ... 39
2.4.4 Future Directions ... 42
2.5 Conclusions ... 43

Chapter 3:
The Time Course of Vocal Attribution: Masculinity, Attractiveness, and Trustworthiness Perceptions at Varied Exposure Times ... 45-73
3.1 Introduction ... 45
3.2 Study 1 ... 49

3.2.1 Methods 49
3.2.1.1 Stimuli 49
3.2.1.2 Subjects 50
3.2.1.3 Procedure 51
3.2.2 Results 51
3.2.2.1 F_0 over time 52
3.2.2.2 Perceived Masculinity 54
3.2.2.3 Perceived Attractiveness 55
3.2.2.4 Perceived Trustworthiness 55
3.2.2.5 Variance Explained 56
3.2.2.6 Relationship Between F_0 and Perceived Traits 57
3.3 Study 2 58
3.3.1 Methods 59
3.3.1.1 Stimuli 59
3.3.1.2 Subjects 59
3.3.1.3 Procedure 59
3.3.2 Results 60
3.3.2.1 F_0 over time 61
3.3.2.2 Perceived Masculinity 62
3.3.2.3 Perceived Attractiveness 64
3.3.2.4 Perceived Trustworthiness 64
3.3.2.5 Variance Explained 65
3.3.2.6 Relationship Between F_0 and Perceived Traits 66
3.4 General Discussion 67
3.4.1 Masculinity 68
3.4.2 Attractiveness 70
3.4.3 Trustworthiness 71
3.5 Conclusions 72

Chapter 4:
Vocal Fundamental and Formant Frequencies Affect
Perceptions of Speaker Cooperativeness 74-102
4.1 Introduction 75
4.1.1 Trustworthiness 76
4.1.2 Dominance 77
4.1.3 The Present Research 79
4.2 Study 1 80
4.2.1 Methods 80
4.2.1.1 Stimuli 80
4.2.1.2 Subjects 81
4.2.1.3 Procedure 82
4.2.1.4 Data Treatment 82
4.2.2 Results 83
4.2.2.1 Measurements and Cooperativeness Ratings 84
4.2.2.2 Mixed Effects Models 88
4.2.2.3 Individual Differences 90
4.3 Study 2 91
4.3.1 Methods 91

4.3.1.1 Stimuli ... 91
4.3.1.2 Subjects ... 92
4.3.1.3 Procedure ... 92
4.3.2 Results ... 93
4.4 General Discussion ... 95
4.4.1 What males a voice sound cooperative? ... 95
4.4.1.1 Pitch and Pitch Variation ... 96
4.4.1.2 Formants ... 97
4.4.2 Individual Differences ... 99
4.5 Conclusions ... 101

Chapter 5:
Effects of Vocal Characteristics on Memory for Voice Identities and Crossmodal Content Recognition ... 103-134
5.1 Introduction ... 103
5.2 Methods ... 108
5.2.1 Stimuli ... 108
5.2.2 Participants ... 110
5.2.3 Procedure ... 110
5.2.4 Data Treatment ... 111
5.2.5 Preference Data ... 111
5.3 Results ... 112
5.3.1 Male Spoken Word Recognition Tests – Experiment 1 ... 113
5.3.2 Female Spoken Word Recognition Tests – Experiment 2 ... 115
5.3.3 Male Textual Word Recognition Tests – Experiment 3 ... 116
5.3.4 Female Textual Word Recognition Tests – Experiment 4 ... 118
5.3.5 Attractiveness and Memory ... 119
5.3.6 Effects of Voice Attractiveness Across Modalities ... 121
5.4 Discussion ... 123
5.4.1 Summary of Results ... 123
5.4.2 No Effect of Manipulated F_0 in Spoken Word Trials ... 124
5.4.3 Negative Effect of Masculinised F_0 in Textual Word Trials Only ... 125
5.4.4 Reduced Accuracy with Crossmodal Content Recognition ... 127
5.4.5 Do attractive voices help us remember? ... 127
5.4.5.1 No Effect of Masculinity/Femininity Preference ... 127
5.4.5.2 Remembering Attractive Voices ... 128
5.4.5.3 Remembering Content ... 130
5.4.6 Effect of Manipulated F_0 on False-Positives for Distractor Stimuli ... 132
5.5 Conclusions ... 133

Chapter 6:
Masculinity of Voices, Faces and Videos: A Cross-Modal Matching Task ... 135-159
6.1 Introduction ... 135
6.2 Methods ... 140
6.2.1 Stimuli ... 140
6.2.2 Participants ... 141

6.2.3 Procedure ... 141
6.2.4 Data Treatment and Accuracy Variables ... 142
6.3 Results ... 144
6.3.1 Study 1: Voices and Static Face Images ... 144
6.3.2 Study 2: Voices and Congruent Videos ... 145
6.3.3 Study 3: Voices and Incongruent Videos ... 146
6.3.4 Masculinity Differences ... 146
6.3.4.1 Independent and Partial Correlations in Study 1 (Male Faces) ... 147
6.3.4.2 Independent and Partial Correlations in Study 1 (Female Faces) ... 148
6.3.4.3 Mixed Effects Model for Study 1 ... 149
6.3.4.4 Independent and Partial Correlations in Study 2 (Male Congruent Videos) ... 149
6.3.4.5 Independent and Partial Correlations in Study 2 (Female Congruent Videos) ... 150
6.3.4.6 Mixed Effects Model for Study 2 ... 151
6.3.4.7 Independent and Partial Correlations in Study 3 (Male Incongruent Videos) ... 151
6.3.4.8 Independent and Partial Correlations in Study 3 (Female Incongruent Videos) 152
6.3.4.9 Mixed Effects Model for Study 3 ... 153
6.4 General Discussion ... 153
6.4.1 Summary of Results ... 153
6.4.2 Matching Voices and Static Faces ... 154
6.4.2 Matching Voices to Videos ... 155
6.4.4 Masculinity Did Not Aid Matching ... 156
6.5 Conclusions ... 159

Chapter7:
General Discussion ... 160-172
7.1 Chapter Summary and Future Directions ... 160
7.2 Overarching Themes ... 167
7.2.1 Prosociality ... 167
7.2.2 Masculinity and Sex Differences ... 168
7.2 General Conclusions ... 171

References ... 173-191

LIST OF FIGURES

Figure 1.1 Overview of the voice production system and its major determinants. ....p. 3
Figure 1.2 The anatomical positioning of the vocal folds within the vocal tract. ....p. 4
Figure 1.3 Sexual dimorphism of the vocal tract. ....p. 9
Figure 2.1 Vocal measurements across speech types. ....p. 31
Figure 2.2 Attractiveness across speech types. ....p. 33
Figure 3.1 Mean F_0 and ratings of masculinity, attractiveness and trustworthiness plotted in all time steps (Study 1). ....p. 53
Figure 3.2 Variance explained (Study 1). ....p. 57
Figure 3.3 Mean F_0 and ratings of masculinity, attractiveness and trustworthiness plotted in all time steps (Study 2). ....p. 63
Figure 3.4 Variance explained (Study 2). ....p. 66
Figure 4.1 Cooperativeness ratings (Study 1) by high/low F_0, F_0-*SD*, D_f and P_f. ....p. 86
Figure 4.2 3D plot of cooperativeness ratings as a function of the interaction between F_0 and the dominance scores of male listeners. ....p. 89
Figure 4.3 Correlations between mean cooperativeness ratings given and listeners' dominance scores. ....p. 90
Figure 4.4 Cooperativeness ratings by pitch condition (Study 2). ....p. 93
Figure 5.1 Listener memory by pitch manipulation in spoken word and visual word trials. ....p. 114
Figure 5.2 Percent correct responses for manipulated-F_0 distractor stimuli in the test phase (spoken word trials only). ....p. 115
Figure 5.3 Correlations between mean opposite-sex rated voice attractiveness and mean opposite-sex target accuracy for male and female voices. ....p. 121
Figure 5.4 Correlations between rated voice attractiveness and differential accuracy across congruent spoken word trials and crossmodal textual word recognition trials. ....p. 122
Figure 6.1 Diagram of stimuli and masculinity difference variables. ....p. 144
Figure 6.2 Listener performance for Studies 1, 2 and 3. ....p. 145

CHAPTER 1: INTRODUCTION

"…language is not, as we are led to suppose by the dictionary, the invention of academicians or philologists. Rather, it has been evolved through time... by peasants, by fishermen, by hunters, by riders."

— Jorge Luis Borges, *This Craft of Verse*

Language and the evolution of language have been a topic of scientific interest for some time. Human language is crucial not only for our intelligence but also for our culture. This is evidenced by domain-specific modules for speech and language processing (Chomsky, 1980; Fodor, 1983), which can lead an infant child to adapt to whatever linguistic culture they are born into, and some research suggests that they are sensitive to this even before birth (DeCasper & Spence, 1986; Mampe, Friederici, Christophe and Wermke, 2009). Given that the voice is of such great import to language, it is not surprising, then, that the voice is used not only to identify familiar individuals, but also to assess characteristics of unfamiliar persons. It is often true that people can be heard before they are seen, and research has shown that we are able to make judgements about a person's physical stature, their masculinity or femininity, their attractiveness and even their personality traits using their voices alone (Apple, Streeter and Krauss, 1979; Avery & Liss, 1996; Collins, 2000; Hodges-Simeon, Gaulin and Puts, 2010; Jones, Feinberg, DeBruine, Little and Vukovic, 2010; Scherer, 1978; Tigue, Borak, O'Connor, Schandl and Feinberg, 2012). More recently, researchers have shown interest in the measured properties of voices, and the way that these relate to the formation of these kinds of trait attributions. Here, I will summarise this research, including a brief overview of how speech is produced physiologically. This will be followed by a summary of key findings which link vocal characteristics to perceptions of speaker traits, and how we assign identity based on voices. I will finish with an overview of the experimental chapters contained in this thesis, which attempt to shed light on some of the lesser-studied aspects of voice perception.

1.1 The Mechanics of Speech Production

Speaking is a complex physical action which requires the synchrony of a host of actions, including respiration, vocal tract movements, and facial movements. Voice pitch (fundamental frequency, F_0) is created by vibrations of the vocal folds (also known as the vocal chords). The thickness of the vocal folds determines the range of pitch which can be uttered by a speaker. Testosterone positively influences the development of somatic tissues in the vocal folds in humans and other animals, making voice pitch a sexually dimorphic trait (Akcam *et al.*, 2004; Beckford, Rood, Schaid and Schanbacher, 1985; Jenkins, 1980; King, Ashby and Nelson, 2001; Sassoon, Segil and Kelley, 1986). Indeed, male children at puberty experience a marked drop in voice pitch that females do not experience (Harries, Hawkins, Hacking and Hughes, 1998; Harries, Walker, Williams, Hawkins and Hughes, 1997). Furthermore, within adult males, circulating testosterone is also correlated with low voice pitch (Dabbs & Mallinger, 1999; Evans, Neave, Wakelin and Hamilton, 2008).

Within each individual, F_0 within this range is determined by many factors, including respiration. Air pressure in the lungs is at its highest at the beginning of an utterance, after the speaker has inspired and the lungs are full of air. See Figure 1.1 for a diagram of the vocal apparatus. F_0 is affected by this subglottal air pressure, with a higher F_0 corresponding to high air pressure, just as a balloon that is full of air sounds higher-pitched upon release than a balloon which is nearly empty. Thus, there is a natural pattern of intonation in speech which allows us to identify sentences and phrasing even in languages we are unfamiliar with. The first words of a sentence have a higher F_0 than the rest of the utterance, and as the utterance progresses, subglottal air pressure drops gradually. Sentences or phrases are finished with an abrupt fall in F_0 as the speaker expels remaining air in the lungs in preparation for another breath (Fant, 1960; Lieberman, 1967; Vaissière, 2005).

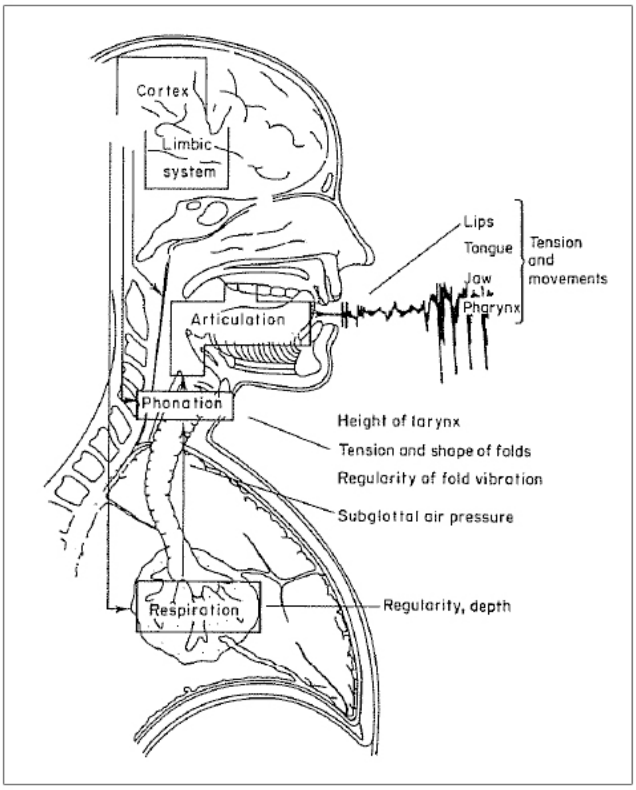

Figure 1.1
Overview of the voice production system and its major determinants. Reprinted from Scherer (1989).

Autonomic arousal also affects speaking F_0. Panic, fear, surprise and “hot” anger have been shown to activate tension in the vocal folds via stimulations of the vagus nerve, which is related to the “fight or flight” response to physical threat (Charous, Kempster, Manders and Ristanovic, 2001). Thus, the measurement of F_0 can tell us about the masculinity or femininity of a speaker, and, to a certain degree, their emotional state. Additionally, within a certain range, F_0 can be controlled deliberately by the speaker.

As explained above, F_0 varies within an utterance for a number of reasons, and this variance in pitch is an important measure itself. This is captured by measuring the within-utterance standard deviation in F_0, denoted as F_0-*SD*. Voices range from monotone (little variation in pitch) to very dynamic (high variation in pitch). Like F_0, F_0-*SD* is sexually dimorphic, with women speaking with a greater within-utterance variation in pitch than men. Pitch variation is also affected by emotional state, with high-activation emotions (anger, fear, panic) resulting in greater pitch variation than normal, and emotions like sadness and boredom resulting in monotony (Banse & Scherer, 1996; Scherer, Banse and Wallbott, 2001). F_0-*SD* is also affected by the type of utterance being measured. Researchers have shown that readings of scripted texts are more dynamic (i.e. with a higher degree of pitch variation) than natural, spontaneous utterances (Abu-Al-makarem & Petrosino, 2007; Batliner, Kompe, Kießling and Niemann, 1995; Laan, 1997; Remez, Bressel, Rubin and Ren, 1987).

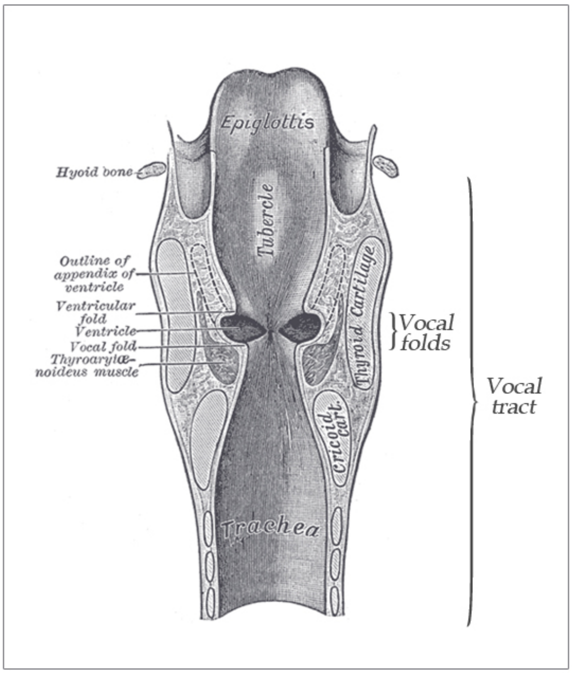

Figure 1.2
The anatomical positioning of the vocal folds within the vocal tract. Adapted from Gray's Anatomy (Gray, 1918): "Coronal section of the larynx and upper part of the trachea."

Formant measures also play an important role in perception. Formants are the harmonic overtones of F_0 as they resonate in the vocal tract (Titze, 1994). As such, formants are related to the structure of the vocal tract, which is outlined in Figure 1.2. Formant dispersion (D_f) is the distance between the formants in Hz, and the length of the vocal tract is inversely related to D_f in primates (Fitch & Hauser, 1995; Fitch, 1997). The length of the vocal tract maps on to overall body size due to an allometric relationship between the two – *e.g.* taller persons tend to have longer vocal tracts, because their bodies are larger. This relationship was true for the rhesus macaques (*Macaca mulatta*) studied by Fitch (1997), however the vocal tract in humans is not as restricted to skeletal constraints, probably due to the descended larynx, detached hyoid bone, and range of tongue and lip movements required for speech. Only weak evidence exists that there is a direct relationship between D_f and body size in humans (González, 2004), nevertheless perceptions based on these traits persist. One study has revealed a negative relationship between testosterone and formant frequencies (Bruckert, Liénard, Lacroix, Kreutzer and Leboucher, 2006). These contending results have led researchers to develop another formant measure which may be more reflective of body size than D_f. Recently, Puts and colleagues (2012) published a paper with a vocal measure which they claim is more strongly related to sexual dimorphism and height, *formant position*, denoted as P_f (Puts, Apicella and Cárdenas, 2012).

Rather than averaging the distance between the formants, P_f averages the relative frequency of each formant within a population. That is, a standardized score for each formant is assigned based on a population (or stimulus set), and these scores are averaged for each voice creating a score for the average "position" of an individual's formant frequencies relative to the population (Puts, Apicella, *et al.*, 2012). Although Puts and colleagues claim that the measurement is more sexually dimorphic than D_f, the measurement is still very new to the field of voice research, and has yet to be tested by a wide range of researchers. Despite this lack of rigorous testing, P_f has been linked to height, arm strength and upper body strength in men (Hodges-Simeon, Gurven, Puts and Gaulin, 2014; Puts, Apicella, *et al.*, 2012).

1.2 Links Between Voice and Face Perception

Some research has referred to the voice as an "auditory face," because voices offer information to listeners which aids in the ascription of interpersonal characteristics and identity of the speaker in similar fashion to faces (Belin, Bestelmeyer, Latinus and Watson, 2011; Belin, Fecteau and Bédard, 2004). As with faces, listeners can recognise persons from their voices, even after long periods of time (Papcun, Kreiman and Davis, 1989; Schweinberger, Herholz and Sommer, 1997), and listeners integrate voice and face identities to create multimodal representations of person identity, on both a conceptual and a cortical level (Campanella & Belin, 2007; Schweinberger, Robertson and Kaufmann, 2007). This same type of integration also applies to the evaluation of emotional states (Joassin, Maurage, Bruyer, Crommelinck and Campanella, 2004; Johnstone, van Reekum, Oakes and Davidson, 2006; Kreifelts, Ethofer, Grodd, Erb and Wildgruber, 2007; Pourtois, de Gelder, Vroomen, Rossion and Crommelinck, 2000). Moreover, variation in both facial and vocal structure can generate variation in perceptions of identity and affect (Belin et al., 2004; Ellis, 1989), as both voices and faces both communicate emotion based on spontaneous physiological articulations arising from autonomic responses to the environment (Banse & Scherer, 1996; Charous et al., 2001; Johnson, Emde, Scherer and Klinnert, 1986; Scherer, 1995, 2000a).

For these reasons, it may be possible to draw from the field of face research (both cognitive and evolutionary in scope) and apply these findings and methods to voice research. Indeed, many researchers have found similar perceptual results when examining facial and vocal stimuli. The testosterone-dependence of both of these traits indicates that they may both be reliable indicators of masculinity/femininity; that they are also both secondary sexual characteristics has led some researchers to believe that they operate as dual cues to mate quality (via their association with attractiveness) (Feinberg, 2008; Feinberg et al., 2005). It may be further proposed that viewers/listeners use voices in similar ways to faces in the ascription of social traits like dominance and trustworthiness. A growing body of work suggests that this is indeed the case (e.g.Puts, Jones and DeBruine, 2012; Tigue, Borak, O'Connor, Schandl and Feinberg, 2012; Todorov, Baron and Oosterhof, 2008; Todorov, Said, Engell and Oosterhof, 2008; Wolff & Puts, 2010)

1.3 Physical, Aesthetic and Social Perception Based on Voices

1.3.1 Masculinity, Dominance and Antisocial Behaviour

Because voice pitch is sexually dimorphic, a low F_0 reliably predicts ratings of men's masculinity (Collins, 2000). F_0 in men has also been related to ratings of other aspects of a masculine physical appearance such as height and hirsuteness (Collins, 2000), and a low voice pitch is also perceived as physically dominant (Puts, Hodges-Simeon, Cárdenas and Gaulin, 2007; Wolff & Puts, 2010). Furthermore, Puts, Gaulin, and Verdolini (2006) demonstrated that men lower their voice pitch when speaking to a competitor, but only if they considered themselves more dominant than their competitor. If the speaker perceived their rival as more physically dominant than themselves, the men raised their voice pitch, possibly as a sign of submissiveness in order to avoid physical altercation. These results, paired with the link between F_0 and testosterone in men (and testosterone's relationship with aggression and antisocial behaviour), suggest that voice F_0 is an honest signal of dominance and threat potential (Booth & Osgood, 1993; Dabbs & Morris, 1990; Mazur & Booth, 1998; Puts, Apicella, *et al.*, 2012).

How much a speaker varies their pitch while speaking could also be a cue to the speaker's threat potential. In a study by Hodges-Simeon et al (2010), low F_0 variation was associated with higher ratings of physical dominance. A low F_0-*SD* is also related to self-measures of physical aggressiveness (Puts, Apicella, *et al.*, 2012). Because a high F_0 variation is associated with heightened autonomic arousal and anxiety (Scherer, 1989;

Williams & Stevens, 1972), a low F_0 variation could indicate confidence in the face of competition, while a high variation may reflect nervousness and preparedness for conflict (i.e. fight or flight). This association is evident based on listener stereotypes of psychological stress, which include heightened F_0 and F_0 variation (Streeter, MacDonald, Apple, M and Galotti, 1983).

Formant traits are also associated with masculinity and dominance. Low formant dispersion (D_f) and low formant position (P_f) are related to body size, and it has been suggested that these traits serve as cues to threat potential (Feinberg, Jones, Little, Burt and Perrett, 2005; Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2007). A low formant dispersion means that the formants are spaced closely together in terms of frequency (Hz). Because D_f is sexually dimorphic, inversely related to testosterone, and a low D_f is related to a longer vocal tract, and thus, larger body size, this trait can be viewed as another cue to men's physical dominance (Bruckert *et al.*, 2006; Fitch & Hauser, 1995; Wolff & Puts, 2010). Indeed, men's voices that have been manipulated to have a lower D_f are rated as more dominant than their higher-D_f counterparts, and this trait had a greater influence on dominance ratings than F_0 (Puts *et al.*, 2007).

Taken as a whole, voice pitch, pitch variation, and formant measures all reflect sexually dimorphic traits, and low values of these traits are rightly perceived as masculine by listeners. Considering the negative relationship between testosterone and both voice pitch and formant measures, and the robust links between testosterone and aggressive behaviour, these perceptions may be highly salient in terms of intrasexual competition among men.

1.3.2 Attractiveness

Sexual dimorphism, in addition to being indicative of masculinity or femininity, affects perceptions of attractiveness, and sexually dimorphic voice traits are generally considered attractive in both male and female speakers (Feinberg, 2008; Feinberg *et al.*, 2011a). Duly, low voice pitch is considered attractive in male voices, and a high voice pitch is attractive in female voices (Collins, 2000; Feinberg, Jones, Little, *et al.*, 2005; Feinberg, Jones, DeBruine, *et al.*, 2005; Puts, Barndt, Welling, Dawood and Burriss, 2011). Sexual dimorphism of formant measures are also preferred in male and female voices (Feinberg, Jones, Little, *et al.*, 2005; Feinberg *et al.*, 2011a; Puts *et al.*, 2011). Recent research has also shown that a high pitch variation is attractive in female voices, which is congruent with the sexual dimorphism of the trait (Henton, 1995; Leongómez *et al.*, *in press*), while men with a low pitch variation report a greater number of past-year sexual partners (Hodges-Simeon, Gaulin and Puts, 2011). Furthermore, men and women may exaggerate these sexually dimorphic traits when speaking to attractive members of the opposite sex (Fraccaro *et al.*, 2011; Hughes, Farley and Rhodes, 2010; Leongómez *et al.*, *in press*; Puts *et al.*, 2006).

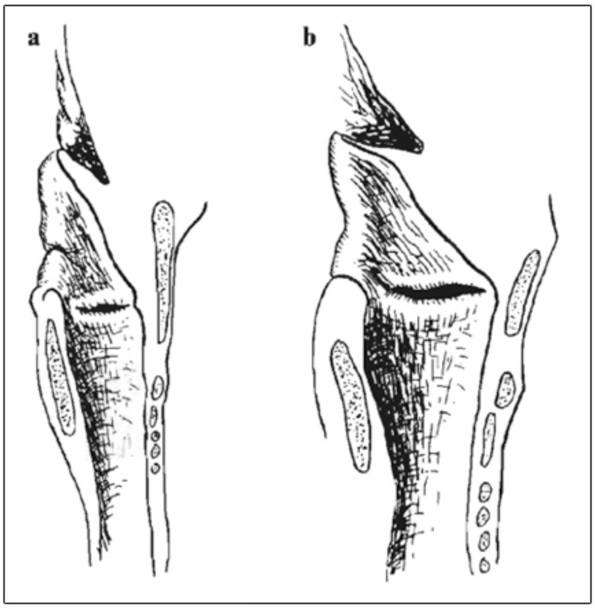

Figure 1.3
Sexual dimorphism of the vocal tract. Sagittal sections of the larynges of a 15-year-old female (a) and a 19-year-old male (b), illustrating the sexual dimorphism of the vocal folds. The vocal cords of a 19-year-old female (were such available for more precise comparison) would only be approximately 15% longer than those shown here. Reprinted from Negus (1949).

These preferences for sexual dimorphism in voices are closely tied to preferences for sexually dimorphic faces. Correlated preferences for masculine faces and voices in male and female speakers speak to this tie, and the two are considered dual cues to mate value (Feinberg, DeBruine, Jones and Little, 2008; Feinberg, 2008; Feinberg, Jones, DeBruine, *et al.*, 2005; Fraccaro *et al.*, 2010). This is because masculinity in male voices and faces are linked to testosterone and health, and femininity in female voices and faces is associated with oestrogen and fertility (Abitbol, Abitbol and Abitbol, 1999; Dabbs & Mallinger, 1999; Hill & Hurtado, 1996; Law Smith *et al.*, 2006; Penton-Voak & Chen, 2004; Rhodes, Chan, Zebrowitz and Simmons, 2003; Thornhill & Grammer, 1999). Because testosterone has immunosuppressive effects (Chen & Parker, 2004), those individuals carrying markers of high testosterone are thought to have increased resistance to infectious diseases and higher mate quality (Folstad & Karter, 2009). While most studies have found generalised preferences for masculinity in male features, some studies have shown that testosterone is correlated with perceptions of health and dominance, but not attractiveness (Rhodes *et al.*, 2003; Swaddle & Reierson, 2002). The relationship between testosterone and aggressive and antisocial behaviour may be to blame for this mismatch, and it has been found that women prefer more feminine male faces as long-term partners and when they are less fertile, but more masculine male faces as short-term partners and when they are most fertile (Jones, Little, *et al.*, 2005; Little, Jones and DeBruine, 2008; Penton-Voak & Perrett, 2000). This pattern of preferences in male faces is also mirrored in women’s preferences for male voices (Feinberg *et al.*, 2006; Puts, 2005), lending further support to the theory that voices are processed in similar ways to faces.

Another sexually dimorphic vocal trait which has been associated with attractiveness is harmonics-to-noise ratio, *HNR*. HNR captures the amount of noise in the auditory signal, and is used as a measure of vocal quality, allowing researchers to capture the degree of vocal “smoothness” and, conversely, “roughness.” A smooth vocal quality is consistent with health and youthfulness, and is positively related to listeners’ ratings of vocal attractiveness

(Bruckert et al., 2010; Eskenazi, Childers and Hicks, 1990; Ferrand, 2002). Recent work has shown this to be inconclusive, however, with lower HNR (*i.e.* rougher vocal quality) associated with increased attractiveness, possibly due to age-related associations with likeability and trustworthiness (McAleer, Todorov and Belin, 2014).

1.3.3 Personality Traits and Prosociality

Voice pitch, pitch range and speech rate change based on the emotional state of the speaker (Murray & Arnott, 1993; Scherer, 2000b; Williams & Stevens, 1972), and these emotions are accurately identified by listeners (de Gelder & Vroomen, 2000; Johnson et al., 1986; Scherer et al., 2001; Uldall, 1960). Beyond purely emotional states, listeners also judge personality traits from voices, including extraversion, truthfulness, empathy and nervousness (Apple *et al.*, 1979; Scherer, 1978). Recently, researchers have become interested in attributions of trustworthiness made using vocal stimuli.

Voices with low F_0 are considered more truthful and trustworthy than voices with high F_0, in both male and female voices (Apple *et al.*, 1979; Klofstad, Anderson and Peters, 2012; Tigue *et al.*, 2012). This may be associated with voice attractiveness via a “halo” effect, causing voices which sound attractive to be ascribed other positive traits (Eagly, Ashmore, Makhijani and Longo, 1991; Zuckerman, Miyake and Elkin, 1995). Contrary to this argument, Tigue *et al.* (2012) found that lower-pitched voices were considered trustworthy in both men and women. This suggests that trustworthiness may not be strictly related to vocal attractiveness, but rather a generalized effect of vocal masculinity being considered trustworthy in voices of both men and women. This may be due to perceived age, as the voice pitch of both men and women decreases over the life span (Linville & Fisher, 1985), and lower-pitched voices are perceived as older than high-pitched voices (Collins, 2000).

This relationship between low voice pitch and attributions of trustworthiness is somewhat unexpected in light of the links between voice pitch, testosterone and antisocial behaviour (Booth & Osgood, 1993; Dabbs & Morris, 1990) and the aforementioned relationship between low voice pitch and physical dominance (Puts *et al.*, 2007; Wolff & Puts, 2010). Researchers have not yet examined perceptions of other pro-social traits, for example cooperativeness. It may be that cooperativeness perceptions map on to trustworthiness perceptions, however, it is reasonable to suspect that higher-pitched, non-dominant sounding voices may be considered more prosocial.

In addition to pitch, pitch variation (F_0-*SD*) also influences personality perceptions, and is considered a pleasant vocal attribute (Apple *et al.*, 1979; Scherer, 1974). A high pitch variation is found in play behaviour in non-human primates, and also in human child-directed speech, which suggests that it may be used as a signal of affiliation (Goedeking, 1988; Trainor, Austin and Desjardins, 2000). An increased variation in F_0 may then also be related to perceptions of prosociality, evidenced by research which has found that a low pitch variation is associated with attributions of dominance (Hodges-Simeon *et al.*, 2010) and a high pitch variation is related to happiness (Vaissière, 2005). Thus, a high F_0-*SD* may also be considered pro-social in addition to affiliative and non-dominant.

HNR, like voice pitch, decreases with age. This results in older voices sounding more rough than younger voices, which sound smoother and healthier (Bruckert et al., 2010; Ferrand, 2002). McAleer and colleagues recently found that HNR is negatively associated with perceptions of likeability and trustworthiness; they propose that this may be driven by age-stereotypical perceptions of prosociality, such that older individuals are generally considered more trustworthy than younger individuals (McAleer et al., 2014). It should be noted, however, that this finding conflicts with research which suggests that younger voices sound warmer, more truthful, more kind, and less dominant (Berry, 1990; Montepare & Zebrowitz-McArthur, 1987; Zebrowitz-McArthur & Montepare, 1989).

1.4 Identity

Faces and voices belong to individuals, and these aspects lend themselves to person perception and help us to assign identity. This is possible because faces and voices are *memorable* and *variable* between individuals. The voice is often referred to as an "auditory face" because variation in voices aids in the ascription of identity, whether based on morphological aspects of the vocal apparatus (*e.g.* high or low pitch) or idiosyncratic aspects of intonation and delivery (*e.g.* monotony) (Belin et al., 2011, 2004). A number of studies have additionally shown that faces and voices are integrated in cognition by the listener to create multimodal representations of identity (Campanella & Belin, 2007; Schweinberger et al., 2007).

As expounded in the above sections, this variation in facial and vocal traits is largely driven by sexual dimorphism. While researchers generally believe that faces and voices communicate the same cues to mate quality (Feinberg, 2008; Feinberg, Jones, DeBruine, *et al.*, 2005), whether attractiveness in faces and voices are concomitant is somewhat unclear. While women who have attractive faces also have attractive voices (Collins & Missing, 2003), there is mixed evidence to support this relationship in men. Collins (2000) found that masculine voice traits did not correlate with masculine body characteristics, despite women's expectations to the contrary. Furthermore, Puts, Gaulin, Sporter, and McBurney (2004) found that low voice pitch was not associated with other sexually dimorphic traits in men, including measures of physical dominance, mating success, or 2D:4D, all of which are independently associated with masculinity (Apicella, Feinberg and Marlowe, 2007; Mazur & Booth, 1998; Mikach & Bailey, 1999; Neave, Laing, Fink and Manning, 2003; Puts *et al.*, 2004). At present, only one study has found that facial and vocal attractiveness are related in men (Saxton, Caryl and Roberts, 2006), but this has not been replicated (Lander, 2008). So, while faces and voices are processed similarly as cues to identity, and masculine facial and vocal traits are *expected* to co-occur by observers, whether they actually do co-occur in men is still relatively unexplored.

In addition to helping us identify and remember individuals, this variation in facial and vocal traits also seems to help us remember information not related to identity. Allan, Jones, DeBruine and Smith (2012) and Smith, Jones, Feinberg and Allan (2012) found that listeners exhibited enhanced recognition for visual objects which were associated with attractive face and voice stimuli, suggesting that attractive voices, like attractive faces, may attract more attention on part of the observer (Maner *et al.*, 2003). However, whether there is an effect of voice attractiveness on content recall in the absence of visual stimuli has received relatively little attention.

Without focusing on attractiveness, a number of studies have examined memory for spoken content; however, it may be necessary to unpick effects of memory for voice identity and memory for spoken content. Listeners have memory for specific voices and for discrete word-voice pairs (Craik & Kirsner, 1974; Goldinger, 1996; Papcun et al., 1989; Schacter & Church, 1992), which suggests that memory for voices may be tied to the specific verbal content of the stimulus, and a shift in sensory modality from exposure to recognition trials can reduce or eliminate priming effects on recall and recognition (Ellis, 1982; Gibson & Bahrey, 2005; Jackson & Morton, 1984). Thus, recall of discrete voice-word pairs may be considered a part of episodic memory (Goldinger, 1996). No study has yet examined whether voice attractiveness may enhance or diminish this type of episodic memory, although Helfrich and Weidenbecher (2011) measured listener performance on content recognition tests after having participants listen to lectures which had been manipulated in pitch. While they found that attractive male (i.e. lowered pitch) lecturing voices enhanced performance on these tests, this may be linked to procedural or academic learning rather than episodic memory *per se*, and whether these findings hold true for voice-word pairs has not yet been studied.

1.5 Overview of Experimental Chapters

This introduction has briefly reviewed the current literature on the perception of physical and social traits based on vocal attributes, the similarities between vocal and facial perception, and their functionality as dual cues to identity and to genetic quality. In the chapters to follow, I will explore these themes more thoroughly, with the aim to fill gaps in our understanding of how vocal characteristics interact with our perception of the speaker in terms of traits including attractiveness, masculinity, trustworthiness and cooperativeness, in addition to how we allocate memory and assign identity. A breakdown of the scope and aims of these chapters follows.

Chapters 2 and 3 explore systematic differences across vocal utterances. This was done in order to determine what makes a good voice stimulus. Researchers have used a number of methods in order to measure trait perception based on voices, but the stimuli they have utilized reflect a range of lengths and content types. Most researchers have utilised sustained vowels as stimuli, while others have used singular spoken words, memorised sentences, and longer scripts which are read from a page. These reflect varying degrees of ecological validity and exposure times, and the use of sustained vowel stimuli has been specifically questioned (Bruckert *et al.*, 2010; Maryn, Corthals, Van Cauwenberge, Roy and De Bodt, 2010) because isolated vowels are not normally encountered in everyday naturalistic interactions. Additionally, vowels lack dynamic speech patterns, which may contain socially important information such as speech rate and pitch variation. In Chapter 2, I explore whether there exist systematic differences in acoustic measurements across three speech types (vowels, scripted text, and natural “spontaneous” speech). I also examine whether attractiveness measurements based on vowel stimuli are related to more naturalistic speech types with the aim to settle on a method of stimulus manufacture which best reflects naturalistic encounters while maintaining a necessary degree of laboratory control.

In Chapter 3, I examine what length of exposure is required in order to obtain consistent/reliable measures of masculinity, attractiveness and trustworthiness in a series of two studies. Studies which have previously measured perceptions of these traits have used stimuli of different lengths, however studies of face perception show that reliable judgements of attractiveness and trustworthiness can be obtained after ≤ 100ms (Todorov, Pakrashi and Oosterhof, 2009; Willis & Todorov, 2006). It is reasonable to expect that attribution of traits based on vocal stimuli also develop over a time course, but no research yet exists on the topic. In Study 1, I examine a time course of attribution using a between-subjects design, with different listeners after 1-second, 5-second and 22+second exposure intervals of the same voice stimuli. Each listener rated voices on all three traits. In Study 2, I examine the same traits using a within-subjects design, where the same group of listeners rated voices in each time interval. Each listener rated voices for only one of the traits, and the time course was shifted to exposure intervals of 1, 2, 5 and 10 seconds. The results from both of these studies reveal a time course of attribution for the three traits assessed, which are discussed in terms of evolutionary psychology and human behavioural ecology. I also examine how measured voice pitch relates to the perception of the three rated traits.

Chapter 4 examines the acoustic properties of voices which sound pro-social. While previous studies have examined voices on the dimension of trustworthiness (including Chapter 3 above), no one has yet studied perceptions of voice cooperativeness. This was done in a series of two studies. In Study 1, I examined how vocal measurements in a naturalistic (i.e. non-manipulated) stimulus set are related to listener attributions of speaker cooperativeness. In Study 2, I manipulated voices in pitch and compared cooperativeness ratings between raised and lowered groups, supporting the naturalistic evidence found in Study 1. The results are discussed in terms of how these vocal traits are related to measures of physical and social dominance.

Chapter 5 questions what makes a voice memorable, and helps to differentiate between memory for individual voice identities and for the content which was spoken. In order to parse this dichotomy, I administered four experiments. In Study 1, I manipulated the pitch of voices speaking a list of singular words. Listeners rated these voices for attractiveness, and after a 3-minute distractor task, they were given an auditory recognition test. Study 2 replicated Study 1 using female voice stimuli. In Studies 3 and 4, I repeated the previous experiments using visual stimuli (textual words) rather than auditory stimuli in the test phase. The results are discussed in terms of adaptive memory and preferential attention as well as the cognitive processes which underlie these systems.

In Chapter 6, I examine vocal and facial masculinity, and whether the lack of relationships found between vocal masculinity and other masculine characteristics found by previous researchers can be extended to include facial masculinity. While facial and vocal masculinity are thought to be dual cues to mate value, whether they are concomitant has not been directly assessed, however there is mixed evidence that facial and vocal attractiveness are related. I test whether vocal and facial masculinity co-occur, and moreover whether observers *expect* them to co-occur. Using a crossmodal matching design, I test whether observers can accurately match faces to voices of real men and women. Following the matching task, the participants also evaluated the voices and faces on the dimension of masculinity. This was done in three studies. Study 1 tested matching ability for voices and static face images. I also repeated the experiment using muted videos of the individuals speaking the same content as in the voice stimulus (Study 2) and different content to the voice stimulus (Study 3). The results are linked to the generally-accepted theory that masculinity and femininity in faces and voices communicate the same underlying genetic quality, and to the hormonal changes experienced throughout puberty which influence the development of these two secondary sexual characteristics. Chapter 7 will discuss the results of these experiments in a broader scope, draw connecting themes across chapters, and reflect upon how future researchers might apply these findings to the field of voice perception, which continues to emerge as an ever-richer field of study.

Throughout all experimental chapters, I will examine both male and female voices, and where appropriate, male and female listeners. The majority of research to date has focused on female perceptions of male voices. While interest is increasing in the study of female voices, very few published experiments examine male and female voices together, and as such, researchers have missed opportunities to search out sex differences in their effects. It is reasonable to presume sex differences may occur, chiefly due to the well-documented theory that women are “choosier” than men (Trivers, 1972), and as such, variation in masculinity and attractiveness is more important for men than for women (Trivers & Willard, 1973). Thus, women may be more attentive to variation in sexually-dimorphic vocal traits when evaluating opposite-sex voices for physical, aesthetic and personality qualities. Fine-tuning of extant theories of voice perception requires examinations of sex differences in effects, and I will address these differences throughout the thesis.

CHAPTER 2: An Investigation of Voice Perception Methodologies: Acoustic Measurements and Perceptions of Voice Attractiveness Using Vowels, Scripted and Spontaneous Speech.

As mentioned in Chapter 1, stimuli of various kinds are used to elicit listener attributions of speaker traits. While sustained vowels are widely used, the ecological validity of this method has been questioned. This research tests whether measurements of vocal characteristics and perceptions of attractiveness made using sustained vowels are comparable to those made using scripted speech and spontaneous speech recordings from the same individuals. Men's voice pitch (F_0) was significantly lower in spontaneous speech while harmonics-to-noise ratio (HNR) was significantly higher in vowels. Men's formant dispersion (D_f) and formant position (P_f) were lower in spontaneous speech relative to the other types of recording. Women's F_0 was significantly lower in vowel sounds, while P_f was lower in spontaneous speech relative to the other types of recording. Pitch variation (F_0-*SD*) was more monotone in spontaneous speech than scripted speech for speakers of both genders. Spontaneous speech elicited the highest attractiveness ratings for voices of both sexes, but there was no relationship between men's vowel attractiveness and ratings made using other speech types. Women's attractiveness correlated strongly and significantly across all speech types. Our results suggest that controlled recordings do not always accurately capture a speaker's vocal traits. The reading of scripted texts may be a valid compromise between laboratory controls and ecological validity, allowing for control over lexical content while capturing a more realistic type of stimulus to present to listeners.

2.1 Introduction

Studies in the field of voice perception research use a number of methods to manufacture stimuli. These primarily consist of recording and playing back spoken vowel sounds, which are used to assess a variety of trait attributions, including attractiveness, dominance, masculinity and pleasantness (*e.g.* Borkowska & Pawlowski, 2011; Bruckert, Liénard, Lacroix, Kreutzer and Leboucher, 2006; Collins, 2000; Feinberg, Jones, Little, Burt and Perrett, 2005; Little, Connely, Feinberg, Jones and Roberts, 2011; Re, O'Connor, Bennett and Feinberg, 2012), although other types of stimuli have recently come to favour. These include scripted text passages that are read aloud (*e.g.* Puts, Gaulin and Verdolini, 2006; Tigue, Borak, O'Connor, Schandl and Feinberg, 2012; Wheatley *et al.*, 2014). Furthermore, other experiments use stimuli of one memorised sentence (*e.g.* Jones, Feinberg, Debruine, Little and Vukovic, 2008; Klofstad, Anderson and Peters, 2012; Lander, 2008) or, in some cases, a singular spoken word is used (*e.g.* Apicella, Feinberg and Marlowe, 2007; Apicella & Feinberg, 2009; O'Connor *et al.*, 2012). Whilst all of these methods are useful for standardising the content and measurement accuracy of voice stimuli, the ecological validity of such stimuli has been questioned (Bruckert *et al.*, 2010; Maryn *et al.*, 2010), as generally these types of speech are not encountered in everyday life.

Indeed, there do appear to be noticeable differences between scripted and spontaneous speech. Researchers have shown that listeners are able to classify speech as either scripted or spontaneous with a high degree of accuracy, even when the content is identical (Levin, Schaffer and Snow, 1982; Remez *et al.*, 1987). Key to this classification is within-utterance pitch variation (Laan, 1997; Remez *et al.*, 1987; Remez, Rubin and Nygaard, 1986), a measure which is characteristically not present in sustained vowel sounds, which do not contain dynamic speech patterns.

Vowel stimuli are of particular value in clinical settings (i.e. for diagnosing laryngeal disorders and dysphonia) as they are context- and content-neutral (Zraick, Wendel and Smith-Olinde, 2005), phonically stable, and precise measurements are easily extractable (Maryn *et al.*, 2010; Parsa & Jamieson, 2001). Additionally, the utterance of sustained vowels controls for regional accents, which can inform listeners of qualities such as group affiliation and social class. However, vowel sounds do not represent commonly used and encountered speech patterns (Askenfelt & Hammarberg, 1986; de Krom, 1994; Maryn *et al.*, 2010; Parsa & Jamieson, 2001), and it has been suggested that results using vowel stimuli may not be easily generalisable to naturalistic encounters (Bruckert *et al.*, 2010). Scripted speech may thus present a justifiable balance between the laboratory control of sustained vowels and the ecological validity of spontaneous speech. The reading of scripted text allows experimenters to control for spoken content which may confound results, particularly in attribution studies, by giving the listener inadvertent information regarding the speaker's personality, education, or social class. However, scripted speech may present its own confounds, including unnatural sentence intonation and pitch variation. Additionally, scripted speech does not control for the non-semantic attributes of pronunciation, articulation and fluency. These vocal qualities are linked to a variety of traits, including age, fertility status, sexual orientation, education, masculinity/femininity, and social class (Foulkes, Docherty and Watt, 2005; Kempe, Puts and Cárdenas, 2013; Kwon, 2010; Milroy, Milroy, Hartley and Walshaw, 1994; Munson, 2007; Pierrehumbert, Bent, Munson, Bradlow and Bailey, 2004; Simpson, 2009; Wadnerkar, Cowell and Whiteside, 2006; Whiteside, Henry and Dobbin, 2004; Whiteside & Marshall, 2001).

Ecological validity may also be important in terms of how listeners interpret voices, especially when used by experimenters to elicit subjective attributions such as attractiveness and dominance. Relatively few published experiments to date have used spontaneous men's speech as stimuli (*e.g.* Fischer *et al.*, 2011; Hodges-Simeon, Gaulin and Puts, 2010; Puts, Hodges-Simeon, Cárdenas and Gaulin, 2007; Puts, 2005), and to our knowledge, no study has examined attributions of women's voices using spontaneous speech.

There is little evidence that vowels or scripted speech produce similar acoustic profiles to spontaneous speech, or whether they elicit comparable assessments for traits such as attractiveness. In a series of studies by Puts and colleagues (Puts *et al.*, 2006; Puts, 2005), three recordings were obtained from male participants: one scripted passage and two spontaneous speech utterances, which were elicited as a "courtship" recording (used for opposite-sex ratings of attractiveness) and a "competitive" recording (used for same-sex ratings of dominance). While the authors found correlations across speech types for measures of voice pitch (fundamental frequency, F_0) and within-utterance pitch variation (measured as the standard deviation of F_0, F_0-*SD*), they found that speakers systematically varied both of these measures from scripted to spontaneous recordings; F_0 in spontaneous speech was either higher or lower than F_0 in scripted speech, relative to the speakers' own self-perceived dominance, and F_0-*SD* was significantly more monotone in spontaneous speech (Hodges-Simeon *et al.*, 2010; Puts *et al.*, 2006). The authors reported no significant differences in formant measures (formant dispersion, *Df*). It is unclear, however, whether the changes observed between scripted recordings to courtship and competitive recordings found in these studies can be entirely attributed to the change in context from neutral to courtship or competition, or to the change from scripted to unscripted speech. Indeed, numerous experiments have found that spontaneous speech is more monotone (*i.e.* has a lower F_0-*SD*) than scripted speech (Abu-Al-makarem & Petrosino, 2007; Batliner *et al.*, 1995; Laan, 1997; Remez *et al.*, 1987).

While these studies have illuminated important differences between scripted speech and spontaneous speech, no study to date has systematically examined whether the vocal traits of vowel utterances and scripted speech (or ratings made using these types of stimuli) are generalisable to the type of vocalisations regularly encountered outside of the psychology laboratory, or if they are more reflective of the laboratory controls used in these experiments. It is also unknown whether subjective attributions made using controlled stimuli are similar to those made using more naturalistic vocalisations. However, spontaneous, unscripted speech introduces a large degree of lexical variation between speakers, and this variation may influence perceptions of attractiveness more than acoustic measurements. Consequently, it is understandable why researchers might choose to standardise stimulus content.

Considering the inherent difference in ecological validity between vowel sounds, scripted speech, and spontaneous speech, and the acoustic differences in these three speech types , it is notable that across stimulus types attractiveness assessments follow the same directionality – low pitched voices in men are generally regarded as attractive by female listeners when presented as vowels (Collins, 2000; Feinberg, Jones, Little, *et al.*, 2005), scripted speech (Vukovic *et al.*, 2010), and in spontaneous speech (Puts, 2005). Higher-pitched female voices are likewise considered attractive presented as vowels (Feinberg, DeBruine, Jones and Perrett, 2008) and scripted speech (Puts *et al.*, 2011). These links between pitch and attractiveness are linked to the sexually dimorphic nature of the trait. F_0 is linked to testosterone in men (Dabbs & Mallinger, 1999) and is considered a secondary sexual characteristic; the average female voice is approximately 5*SD*s above the average male voice (Baken, 1987; Puts, Apicella, *et al.*, 2012). Standard deviation of voice pitch, a measure not available for vowel-sound stimuli, is also sexually dimorphic, with men speaking with less within-utterance variance in pitch (*i.e.* more monotone) than women when reading scripted passages (Henton, 1995; Puts, Apicella, *et al.*, 2012). A low F_0-*SD* has been related to measures of aggressiveness and mating success as measured by number of past-year sexual partners (Hodges-Simeon *et al.*, 2011; Puts, Apicella, *et al.*, 2012). Further to this, a high F_0-*SD* is associated with social affiliation and deference (Hodges-Simeon *et al.*, 2010; Puts, Apicella, *et al.*, 2012), and as such, this measure may play an important role in the perception of traits like dominance and attractiveness.

It is also unclear whether formant frequencies (D_f and formant position, P_f), which affect perceptions of body size and physical dominance in spontaneous speech recordings (Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2007; Wolff & Puts, 2010), and attractiveness in vowel and scripted stimuli (Feinberg, Jones, Little, *et al.*, 2005; Feinberg *et al.*, 2011b; Hodges-Simeon *et al.*, 2010), are variable across sustained vowels, scripted speech, and spontaneous speech when spoken by the same individual. The same is true of harmonics-to-noise ratio, *HNR*. HNR captures the amount of noise in the voice signal, and is commonly used as a measure of voice quality. While HNR and attractiveness ratings are positively related in vowel stimuli, perhaps because smoother voices sound healthier and younger than rough voices (Bruckert et al., 2010; Ferrand, 2002), it would be useful to examine whether these relationships remain in other stimulus types.

The present study investigated measured acoustic parameters of male and female voices across three kinds of vocal recordings: sustained vowels, scripted speech, and spontaneous speech. Attractiveness ratings of these recordings were also obtained to examine whether voice attractiveness, and the links between measurements and attractiveness, are consistent across utterance types. Measures of F_0, F_0- *SD*, D_f, P_f and HNR are included in these analyses because of their prevalence in voice perception research and their links to vocal attractiveness attributions (*e.g.* Bruckert et al., 2010; Feinberg, Jones, Little, et al., 2005; Leongómez et al., 2014); both F_0 and F_0-*SD* have been previously shown to vary between types of speech, but without implementing context-neutral recordings (Hodges-Simeon et al., 2010). While one previous study has found no differences in D_f across scripted and spontaneous utterances (Hodges-Simeon et al., 2010), the present research will also include vowel utterances in the analysis while additionally maintaining neutrality of context. P_f is also included as it is a relatively recent measure of formants which captures the mean standardised formant measure, rather than the average distance between the formants. Its increasing inclusion in current work on voice perception warrants a further critical appraisal of how this measure might (or might not) change based on the type of speech which is uttered. This study’s aim was to determine how vowel sounds and scripted speech are related to naturalistic voice measurements, and whether they elicit comparable attractiveness attributions.

2.2 Methods

2.2.1 Stimuli

Thirty-two participants (16 male, mean age = 20.3 yrs, *SD* = 2.8, range = 18-30; 16 female, mean age = 19.5 yrs, *SD* = 1.5, range = 18-23), all native English speakers, were recruited as stimulus donors. The participants, who were undergraduate students participating to fulfil a course requirement, were recorded speaking three types of vocal utterances. All phases of this experiment were approved by the University of Stirling Psychology Ethics Committee.

2.2.2 Recording and Stimuli Manufacture

Stimuli were recorded using an Audio-Technica AT-4041microphone in a cardioid pickup pattern, at a distance of approximately 65cm. Sound was recorded directly to hard disk using Windows Movie Maker v.2.1.4027.0, with a 48kHz sampling rate and 16-bit quantisation. The room was quiet and partially soundproofed with 1.5-inch thick sound-dampening foam.

First recorded were the vowel stimuli, consisting of the sustained vowels /a/, /e/, /i/, /o/, and /u/. The time space between the vowel utterances was standardised after recording such that each stimulus (including all five vowels) was 5 seconds in length (±1ms). Vocalisations were unaltered by this process; only the duration of silence between the vowels was changed, and did not result in any noticeable difference in the recording beyond altering stimulus length. Secondly, the participants read a scripted standardised text which described the weather in the UK. This script was chosen for its contextual neutrality. Lastly, participants were asked to give a response in their own words to a standardised question posed by the researchers: “If you were to be stranded on a desert island but could take with you only two of the following three items, which would you choose – a can of hairspray, a bar of chocolate, or a plastic bag?"; it was suggested that they answer in a casual way, as though they were having a conversation with their friends (see Cowan & Little, 2013).

Both scripted and uninhibited speech recordings were trimmed to 5 seconds of exposure time, beginning with the 2nd sentence uttered. This was done in order to capture honest representations of the speakers' vocal traits, and to minimise disfluency, as we considered that the speakers may have been nervous to speak into the microphone, a physiological state which has an autonomic impact on the vocal folds and raises voice F_0 (Scherer, 1989). For the scripted utterance, this resulted in a passage of roughly one sentence; "October frequently brings the first frost of the season over the greater part of the UK." As with the vowel utterance, these were limited to 5 seconds of exposure time, ±1ms.

2.2.3 Measurements

All voice measurements were obtained using Praat v. 5.3.03. F_0 measurements were taken using Praat's autocorrelation algorithm. Female fundamental frequencies were searched for between 100-600Hz, and male fundamental frequencies were searched for between 65-300Hz, per the manufacturers' recommendations (Boersma & Weenink, 2012). Formants were measured using Praat's Burg algorithm (non-LPC). For female voices, a ceiling of 5500Hz was used, and the male voice formant ceiling was 5000Hz, again per the manufacturers' recommendations. Formant dispersion (D_f) and formant position (P_f) were calculated using these mean formant measures. For formulas used to calculate D_f and P_f, see Fitch (1997) and Puts *et al.* (2012), respectively. Standard deviation of F_0 was also measured in scripted and spontaneous speech. All measurements were performed using voiced parts of the utterance only.

2.2.4 Participants

There were 214 participants (110 male, mean age = 22.5 yrs, range = 18-64, *SD* = 6.6 yrs; 104 female, mean age = 20.5 yrs, range = 18-46, *SD* = 4.3 yrs) recruited to rate the voices for attractiveness. Participants were students and members of the public who were compensated by being entered into a prize draw for £25 cash, and undergraduate psychology students at the University of Stirling, who participated to fulfil a course requirement. Participants for the prize draw were recruited by advertising at the University campus and using social contacts.

2.2.5 Procedure

Participants were given a web link to the experiment, and completed the study in a quiet room, using either computer speakers or wearing headphones. After completing basic demographic questions, participants were exposed to the experimental trials in the following blocked order: male vowel sounds, male scripted speech, male spontaneous speech, female vowel sounds, female scripted speech, and female spontaneous speech. The 16 voices within each block were fully randomised. Participants could replay the audio recordings if they wished, and participants moved through the study at their own pace, with no time pressure for ratings. Ratings were made using a 7-point Likert scale, with “1” reflecting low attractiveness, and “7” corresponding to high attractiveness.

2.3 Results

Ratings within each stimulus condition were averaged across participants, such that the responses of participants yielded six data points, which were the average attractiveness ratings for each stimulus set (three stimulus types, voiced by two sexes). The resultant data were used in our analyses. All vocal measurements were found to be normally distributed using the Kolmogorov-Smirnov test for normality, validating use of parametric statistical tests. ANOVAs which violated sphericity assumptions were corrected using Greenhouse-Geisser estimates for degrees of freedom. All measurements (with the exception of HNR in isolated vowels) showed a high degree of sexual dimorphism across all speech types. See Table 2.1.

Table 2.1 *Descriptive measurements (mean ± SD) and sexual dimorphism of acoustic parameters (independent-samples t-tests).*

	Vowel			Scripted			Spontaneous		
Measure	Male	Female	*t*	Male	Female	*t*	Male	Female	*t*
F_0 (Hz)	110.1 (±16.1)	150.0 (±34.3)	4.21***	111.2 (±19.6)	190.8 (±42.1)	6.85***	104.7 (±17.2)	177.5 (±39.6)	6.75***
F_0-*SD* (Hz)	-	-	-	14.7 (±7.1)	48.4 (±17.9)	7.00***	13.0 (±7.7)	43.6 (±19.4)	5.86***
D_f (Hz)	768 (±27)	845 (±45)	5.93***	767 (±27)	832 (±39)	5.47***	729 (±93)	847 (±40)	4.69***
P_f	-0.661 (±0.63)	0.661 (±0.53)	6.43***	-0.675 (±0.50)	0.675 (±0.58)	7.06***	-0.526 (±0.57)	0.526 (±0.54)	5.34***
HNR	8.88 (±1.76)	8.52 (±2.03)	0.54	7.15 (±1.22)	8.30 (±1.07)	2.83**	7.19 (±1.73)	8.62 (±1.93)	2.20*

* p < .05; ** p < .01; *** p < .001

2.3.1 Measurements Across Speech Types

Due to the sexual dimorphism of all traits under examination, measurements of acoustic traits were subjected to repeated-measures analyses of variance (ANOVAs) separately by speaker sex. Correlations in acoustic measurements across speech types are also included in order to capture similarity, and to examine whether any differences found by ANOVAs are due to systematic variation.

2.3.1.1 Pitch (F_0) and Pitch Variation (F_0-*SD*)

Significant differences in F_0 were observed across the three speech types in male voices, $F(2,30) = 7.67$, $p = .002$, $\eta_p^2 = .338$. See Figure 2.1. Planned contrasts revealed that men spoke in a lower average F_0 in spontaneous speech compared to vowels, $F(1,15) = 7.62$, $p = .01$, $\eta_p^2 = .337$, and to scripted speech, $F(1,15) = 13.56$, $p = .002$, $\eta_p^2 = .475$. F_0 in vowels and scripted speech did not differ from one another $F(1,15) = 0.50$, $p = .49$, $\eta_p^2 = .032$. Despite these differences, F_0 correlated highly across all three measurements (all $r > .89$, all $p < .001$, all 95% CI between [.70, .99]).

F_0 also differed across the three speech types for female voices, $F(2,30) = 13.38$, $p < .001$, $\eta_p^2 = .472$. Scripted speech F_0 was significantly higher than vowel F_0, $F(1,15) = 25.36$, $p < .001$, $\eta_p^2 = .628$, and spontaneous speech F_0 was also higher than vowel F_0, $F(1,15) = 21.68$, $p < .001$, $\eta_p^2 = .591$. Scripted F_0 did not differ significantly from spontaneous speech F_0, $F(1,15) = 1.88$, $p = .19$, $\eta_p^2 = .111$. F_0 measures were significantly correlated across the three types of stimuli; vowel-script: $r(16) = .66$, $p = .006$, 95% CI [.28, .92]; vowel-spontaneous: $r(16) = .80$, $p < .001$, 95% CI [.59, .93]; script-spontaneous: $r(16) = .55$, $p = .03$, 95% CI [.01, .92].

F_0-*SD* measures were, by nature, only available for scripted and spontaneous speech. F_0 variation was more monotone in spontaneous speech compared to scripted speech for both male and female voices, but these differences were not statistically significant, male $t(15) = 0.90$, $p = .38$; female $t(15) = 1.11$, $p = .28$. F_0-*SD* measures were significantly positively correlated between the two speech types in female voices, $r(16) = .57$, $p = .02$, 95% CI [.31, .78], and close to significant in male voices, $r(16) = .48$, $p = .06$, 95% CI [.22, .81].

2.3.1.2 Formants (D_f and P_f)

D_f was not significantly different across speech types for male voices, $F(1,15.75) = 2.32$, $p = .12$, $\eta_p^2 = .134$. Planned contrasts revealed no significant differences between individual speech types, all $F(1,15) < 2.37$, all $p > .14$, all $\eta_p^2 < .136$. D_f measurements were significantly correlated between vowel and scripted speech, $r(16) = .77$, $p < .001$, 95% CI [.48, .93], but not between vowel and spontaneous speech, $r(16) = -.17$, $p = .53$, 95% CI [-.54, .64], or between scripted and spontaneous speech, $r(16) = -.11$, $p = .69$, 95% CI [-.44, .75].

Female voice D_f did not differ across speech types, $F(2,30) = 1.14$, $p = .33$, $\eta_p^2 = .071$. Planned contrasts revealed no further significant differences between individual types of utterance, all $F(1,15) < 1.77$, all $p > .30$, all $\eta_p^2 < .106$. D_f correlated in female voices between vowels and spontaneous speech only, $r(16) = .64$, $p = .007$, 95% CI [.18, .88]. Vowel D_f and scripted D_f were not significantly correlated, $r(16) = .29$, $p = .27$, 95% CI [-.32, .69], nor were scripted and spontaneous D_f, $r(16) = .30$, $p = .26$, 95% CI [-.15, .68].

Measures of P_f did not differ significantly across speech types in male voices, $F(2,30) = 0.31$, $p = ..74$, $\eta_p^2 = .020$. Vowel P_f was not different from scripted P_f, $F(1,15) = 0.00$, $p = .95$, $\eta_p^2 = .000$, or spontaneous P_f, $F(1,15) = 0.37$, $p = .55$, $\eta_p^2 = .024$. There were no differences in P_f measures between scripted and spontaneous speech, $F(1,15) = 0.82$, $p = .38$, $\eta_p^2 = .052$. However, there also were no correlations between P_f measures of across any of the speech types (vowel-script $r(16) = -.37$, $p = .16$, 95% CI [-.76, .38]; vowel-spontaneous $r(16) = -.09$, $p = .75$, 95% CI [-.49, .42]; script-spontaneous $r(16) = .25$, $p = .35$, 95% CI [-.37, .80]). P_f also did not differ across the three speech types for female voices, $F(1.437, 21.56) = 0.97$, $p = .39$, $\eta_p^2 = .060$. Vowel P_f did not differ individually from script P_f, $F(1,15) = 0.01$, $p = .93$, $\eta_p^2 = .001$, or from spontaneous $P_{f,}$ $F(1,15) = 2.19$, $p = .16$, $\eta_p^2 = .127$. Script P_f also did not differ from spontaneous P_f, $F(1,15) = 2.00$, $p = .18$, $\eta_p^2 = .118$. P_f was correlated between vowels and spontaneous speech, $r(16) = .77$, $p = .001$, 95%

CI [.48, .92], and between scripted and spontaneous speech, $r(16) = .72$, $p = .002$, 95% CI [.32, .91], but not between vowels and scripted speech, $r(16) = .40$, $p = .12$, 95% CI [-.21, .79], although a positive directionality was observed.

HNR was significantly different across speech types for male voices, $F(2,30) = 9.75$, $p < .001$, $\eta_p^2 = .394$. HNR was higher (*i.e.* vocal quality was clearer, less noisy) in vowels compared to scripted speech, $F(1,15) = 12.71$, $p = .003$, $\eta_p^2 = .459$, and spontaneous speech, $F(1,15) = 11.89$, $p = .004$, $\eta_p^2 = .442$. There were no significant differences in HNR between scripted and spontaneous speech, $F(1,15) = 0.01$, $p = .91$, $\eta_p^2 = .001$. HNR was only correlated significantly between scripted and spontaneous speech, $r(16) = .59$, $p = .016$, 95% CI [.32, .79], all other $r < .36$, all other $p > .17$, 95% CI between [-.40, .73].

Female voice HNR was not different across speech types, $F(2,30) = 0.28$, $p = .75$, $\eta_p^2 = .019$. There were also further differences between individual types, all $F < 0.76$, all $p > .39$, all $\eta_p^2 < .048$. HNR was significantly correlated between all three speech types: vowel-script $r(16) = .52$, $p = .04$, 95% CI [.15, .80]; vowel-spontaneous $r(16) = .53$, $p = .03$, 95% CI [.11, .88]; script-spontaneous $r(16) = .68$, $p = .004$, 95% CI [.36, .90].

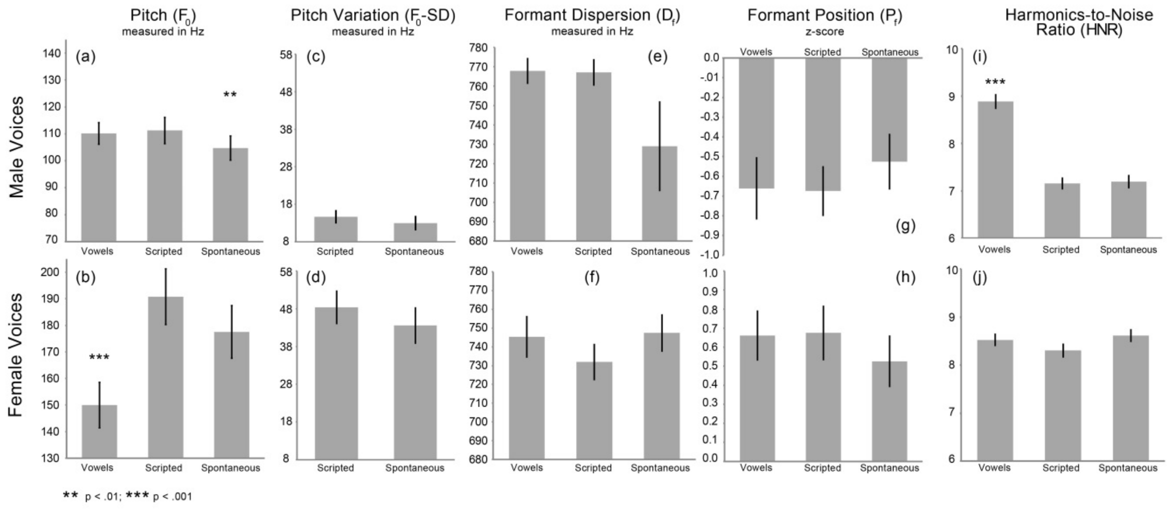

Figure 2.1
*Vocal measurements across speech types. Pitch (F_0) measurements (a and b), pitch variation (F_0-SD) measurements (c and d), formant dispersion (D_f) measurements (e and f), formant position (P_f) measurements (g and h), and harmonics-to-noise ratio (HNR) measurements (i and j) for male and female voice recordings. Error bars represent standard error of the mean; ** p < .01, *** p < .001.*

2.3.2 Ratings Across Speech Types

We examined opposite-sex ratings of male and female voices using two repeated-measures ANOVAs. Attractiveness ratings differed across stimulus types both in male voices, $F(2,206) = 4.24$, $p = .02$, $\eta_p^2 = .039$, and in female voices, $F(1.77,192.74) = 18.84$, $p < .001$, $\eta_p^2 = .147$. See Figure 2.2. Planned contrasts revealed a similar pattern of results in both sexes. Spontaneous speech was rated as significantly more attractive than both vowel sounds, male voices: $F(1,103) = 6.47$, $p = .01$, $\eta_p^2 = .059$; female voices: $F(1,109) = 21.21$, $p < .001$, $\eta_p^2 = .163$, and scripted speech, male voices: $F(1,103) = 4.59$, p = .03, $\eta_p^2 = .043$; female voices: $F(1,109) = 28.94$, $p < .001$, $\eta_p^2 = .210$. Vowel sound ratings and scripted speech ratings did not differ from one another, male voices: $F(1,103) = 0.50$, $p = .48$, $\eta_p^2 = .005$; female voices: $F(1,109) = 0.003$, $p = .96$, $\eta_p^2 = .000$.

Despite these differences in attractiveness ratings between spontaneous speech and other speech types, similarities were also observed. Comparing within stimuli, male voice attractiveness was correlated between scripted and spontaneous speech, $r(16) = .62$, $p = .01$, 95% CI [.28, .85]. However, the rated attractiveness of vowel sounds was unrelated to both scripted and spontaneous speech attractiveness ratings, $r(16) = .16$, $p = .56$, 95% CI [-.47, .60] and $r(16) = .03$, $p = .90$, 95% CI [-.47, .52], respectively. Rated attractiveness of male vowel stimuli accounted for just 0.12% of the variance (r^2) in spontaneous speech attractiveness ratings; rated attractiveness of scripted stimuli accounted for 2.46% of the variance (r^2) in spontaneous speech attractiveness ratings. Attractiveness ratings of female voices were significantly correlated across all stimulus types, all $r(16) > .67$, all $p < .005$, all 95% CI between [.36, .96]. Attractiveness of female vowel stimuli accounted for 51.6% of the variance in spontaneous speech attractiveness ratings, and scripted speech attractiveness explained 61.2% of variance.

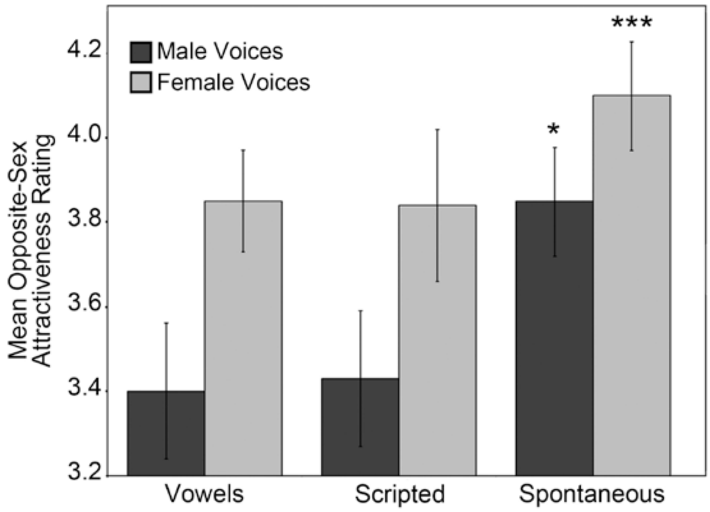

Figure 2.2
*Attractiveness ratings across speech types. Vowel sounds and scripted speech were rated similarly. Spontaneous speech was rated significantly more attractive in voices of both sexes. Bars represent mean opposite-sex attractiveness ratings. Error bars represent standard error of the mean; * p < .05, *** p < .001.*

2.3.3 Relationships Between Vocal Measurements and Attractiveness Ratings

Table 2.2 *Correlations between vocal measurements and attractiveness ratings (Pearson r). 95% confidence intervals are reported using bootstrap measures (based on 1000 samples).*

	F_0		F_0-*SD*		D_f		P_f		HNR	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Vowel	-.44	-.06	-	-	.54*	-.003	.21	-.09	-.40	.12
Lower CI	-.74	-.73	-	-	.13	-.44	-.28	-.75	-.80	-.60
Upper CI	.22	.66	-	-	.84	.50	.64	.56	.16	.76
Scripted	-.14	.37	.01	.45	.05	.04	-.003	-.17	-.15	.11
Lower CI	-.49	-.29	-.37	.05	-.55	-.64	-.58	-.65	-.58	-.54
Upper CI	.34	.77	.47	.87	.52	.57	.61	.28	.23	.87
Spontaneous	-.19	-.09	.34	.18	.04	.30	-.04	-.17	-.28	-.02
Lower CI	-.57	-.56	-.02	-.39	-.40	-.18	-.48	-.64	-.69	-.41
Upper CI	.36	.48	.67	.65	.52	.68	.45	.35	.28	.48

* p < .05

Male voice pitch was negatively but non-significantly correlated with attractiveness measurements in vowel stimuli, but not in scripted or spontaneous speech. Female voice pitch did not correlate with attractiveness ratings in any of the speech types, although there was a positive directionality observed for scripted speech. See Table 2.2.

F_0-*SD* correlated positively but non-significantly with attractiveness ratings of scripted speech in female voices. There was also a positive but non-significant directionality observed in male spontaneous speech.

D_f was significantly positively related to attractiveness ratings of male vowel sounds, but no other speech types revealed significant relationships. P_f and HNR did not significantly correlate with attractiveness ratings in any speech type, for either male or female voices, although for D_f and P_f and HNR, some non-significant relationships were observed. See Table 2.2 for complete results.

2.4 Discussion

In examining acoustic measures across three different types of stimuli, we found that the acoustic properties of vowel-sound and scripted speech recordings differ in some ways from those seen in spontaneous speech, and may not accurately reflect an individual's vocal parameters in more naturalistic encounters. While some measures were strongly correlated across types of speech, we found significant differences in some measurements across the three recordings. Furthermore, the results also suggest that vowel stimuli may not elicit comparable attractiveness ratings compared to spontaneous speech, at least in men. Of the three recordings, spontaneous speech was rated significantly more attractive than vowel sounds and scripted speech in both male and female voices. Further to this, male vowel attractiveness did not correlate with attractiveness ratings of scripted speech or spontaneous speech, although female vowel attractiveness did correlate between all three stimulus types.

Some support linking voice pitch, pitch variation and formant dispersion to ratings of attractiveness was found, but these relationships seem to be dependent on the type of stimulus used.

2.4.1 Measurements Across Speech Types

2.4.1.1 Pitch and Pitch Variation

While F_0 was strongly correlated across recordings for male and female speakers, men exhibited a significantly lower F_0 in spontaneous speech recordings, which suggests that men speak in a systematically lower voice pitch when they speak in their own words rather than in controlled utterances (including vowels). This result is somewhat supported by Puts *et al.* (2006), who found that in a competitive context, men spoke in a lower F_0 in spontaneous speech relative to (neutral) scripted speech, though only if the speakers considered themselves more dominant than their competitors, and the speakers raised their F_0 if they considered themselves less dominant than their competitors. These results show that men speak in a *systematically* lower F_0 in spontaneous speech relative to scripted speech and sustained vowels in entirely context-neutral recordings. This may be applied to the findings of Puts *et al.* (2006) to suggest that dominant men may not lower their F_0 in response to competition (as this may be their normal state when transitioning to spontaneous speech), but perhaps submissive men raise their F_0 as a sign of deference when interacting with a threatening competitor. Women's F_0 also differed across utterance type, however women spoke in a lower pitch in vowel recordings relative to the other two speech types. Perhaps scripted and spontaneous speech elicited a higher pitch in female voices because women speak with less monotony in connected speech relative to men – the variation in pitch could be contributing to the higher mean F_0 observed here. These results also show that both men and women spoke with less within-utterance pitch variation (F_0-*SD*) in spontaneous speech compared to scripted speech, indicating that their voices were more monotone when speaking naturally. While this finding was short of statistical significance, a number of previous experiments have found that scripted speech has a higher F_0 variation than spontaneous speech (Laan, 1997; Remez *et al.*, 1987, 1986). Similar results were reported by Hodges-Simeon *et al.* (2010), however, their results report that F_0-*SD* was different between scripted and non-neutral "courtship" recordings, but was the same for "competitive" recordings. Further, F_0-*SD* was shown to vary for competitive recordings relative to the speakers' self-perceived physical dominance. While both of these spontaneous speech utterances were contextualised, our spontaneous speech recordings were neutral in context, indicating that natural speech patterns are more monotone, and a higher F_0-*SD* in courtship recordings could be contrived by the speaker to maintain the interest of the listener. A high F_0-*SD* amongst speakers of both sexes communicates emotional expressiveness (both positive and negative high-activation emotions including happiness, anxiety and anger), and is perceived as flirtatious, pleasant and interested (Banse & Scherer, 1996; Jurafsky, Ranganath and Mcfarland, 2007; Scherer & Oshinsky, 1977; Scherer, 1974, 1989).

The general drop in F_0-*SD* may also be attributed to the different tasks of reading a scripted text and producing spontaneous speech. Scripted speech produces consistent patterns of intonation on the level of the sentence, with a more consistent rise and fall in F_0 across the sentence relative to spontaneous speech (Lieberman, Katz, Jongman, Zimmerman and Miller, 1985). This difference is also perceived by listeners, who are able to classify speech as either scripted or spontaneous with a high degree of accuracy, even when the content is identical (Levin *et al.*, 1982; Remez *et al.*, 1987). Perhaps the intonation patterns that listeners use as cues are contrived by the speakers due to a desire to give the appearance of spontaneous speech. However, our results, paired with previous findings (*e.g.* Abu-Al-makarem & Petrosino, 2007; Batliner *et al.*, 1995; Hodges-Simeon *et al.*, 2010; Laan, 1997;

Puts *et al.*, 2006; Remez *et al.*, 1987) show that naturalistic patterns of intonation are more monotone and lower in pitch than those produced when speakers read from a script.

2.4.1.2 Formants and Harmonics

These results show no differences in D_f across speech type for either male or female voices; D_f was correlated between vowels and scripted speech in men, and between vowels and and spontaneous speech in women. While there are no significant differences found, the lack of correlations for this measurement is somewhat surprising. If D_f is an indicator of vocal tract length (and thus, body size), a general agreement in this measure across utterances for the same individual should be expected. Indeed, one previous study has found agreement in this measure across two types of utterance (Hodges-Simeon et al., 2010), however their sample included nearly seven times as many voices relative to the present research.

We found very little correlation in P_f across speech types for male voices, however these measures were generally correlated for female voices. As with D_f, P_f was not different across different types of speech. This could be due to the methods used to calculate these measurements (see Fitch, 1997; Puts *et al.*, 2012). While D_f is the averaged distance between formant frequencies, P_f is calculated by obtaining a standardised score for each formant and averaging them, and does not reflect the distance between formants. This measure is relatively new to voice perception research - Puts *et al.* (2012) report that P_f was more sexually dimorphic than D_f in scripted speech recordings, and our data show some support for this conclusion, with P_f contributing larger effect sizes than D_f in all three types of recording, however these differences are not as substantial as those reported by Puts and colleagues. Furthermore, because P_f is a within-sample measure, reflecting the “position” of each individual formant relative to the other stimuli measured, P_f measures from one stimulus set may not be generalisable on a population level. While Puts *et al.* (2012) measured 444 individuals, our study used a sample of 36 voice donors.

HNR was shown to differ across speech types for male voices only, with vowels having a higher vocal quality than other types of speech. This may be explained by the nature of isolated vowel phonations. These are precisely controlled speech patterns with very little movement required of the vocal folds and vocal tract. Types of dynamic speech, such as scripted and spontaneous speech, require much more movement – more is required of the vocal folds, and voices may be less clear (*i.e.* with more noise in the auditory signal) as a result. Female voices, however, did not exhibit the same pattern, perhaps because female voices are generally clearer and smoother. Indeed, while in vowel sounds there was no significant degree of sexual dimorphism in HNR, these differences are significant in both scripted and spontaneous speech. This measure also generally correlated across types of speech for male and female speakers. Because HNR is reflective of vocal quality (as opposed to any kind of affect), it is not surprising that these measures are related when captured from the same individual on successive recordings.

2.4.2 Ratings Across Speech Types

Women's ratings of male vocal attractiveness made using vowel-sound stimuli did not correlate with ratings made using more naturalistic stimuli, and vowel attractiveness predicted extremely low levels of variance in attractiveness ratings of scripted and spontaneous speech. These differences in attractiveness ratings across stimulus types suggest that ratings of male voice attractiveness made using vowel sounds may not be generalisable to those made during more naturalistic encounters. Considering that the attractiveness of male vowel sounds is not a good predictor of the attractiveness of the same voices using more naturalistic speech types, researchers should perhaps reflect on the implications of using vowel sounds as stimuli. Notably, men's ratings of female voice attractiveness were correlated across all stimulus sets. This suggests that female vocal attractiveness may be accurately captured using vowel stimuli. This sex difference may be explained by the relationship between F_0 and attractiveness for male vowel stimuli, which was not present in other stimulus types, or in any stimulus type for female voices.

Spontaneous speech received higher attractiveness ratings than both vowels and scripted speech for voices of both men and women. To a certain degree, these results make intuitive sense, as vowel sounds are not regularly encountered speech patterns. The degree to which spontaneous speech stimuli are more attractive may thus be due to its nature as a common, familiar type of speech. Therefore, it may be that spontaneous speech sounds inherently more "realistic" with natural intonation patterns and the absence of restrictions on lexical content. There is also inherent variability in content for spontaneous speech, whereas in vowel sounds and scripted speech, the content is identical across speakers. This may be potentially less interesting on the part of the listeners, and in turn this could lead to lower attractiveness ratings. Lastly, the spontaneous speech recordings which we obtained were elicited by asking the speakers to respond in a casual way, as though they were speaking to friends. This method may have enhanced the perceived friendliness of the speakers, and may have thus positively influenced attractiveness ratings over the other two types of speech.

2.4.3 Measurements and Attractiveness

While spontaneous speech was rated as significantly more attractive than other speech types (for both male and female speakers), and F_0 was lower in men when speaking their own words, F_0 did not correlate with attractiveness ratings in scripted or spontaneous speech. Therefore, we cannot assume that the difference in F_0 was a driving factor for higher attractiveness ratings in spontaneous speech. Although not significant (likely due to a small sample size), the expected negative correlation between F_0 and attractiveness in male vowel sound stimuli was found, which corroborates the results of numerous studies showing that masculine F_0 is attractive using vowel stimuli (Collins, 2000; Feinberg, Jones, Little, et al., 2005; Feinberg et al., 2006, 2011a; Saxton et al., 2006). However, there were no other significant correlations between F_0 and attractiveness in voices of either sex, in any speech type. This lack of substantial relationship other than in vowel stimuli is contrary to a large body of work which has found robust links between F_0 and attractiveness in both male and female voices using scripted and spontaneous speech (e.g Hodges-Simeon *et al.*, 2010; Puts *et al.*, 2011; Puts, 2005; Vukovic *et al.*, 2010), although (Smith *et al.*, 2012) report no preference for raised-F_0 female voices using singular spoken words. It is important to note that many of these studies utilise manipulated F_0 to find generalised preferences, while our study finds no direct relationship between F_0 and attractiveness using unmanipulated, naturalistic variation in F_0 other than in male vowel stimuli. Furthermore, previous studies tend to use the listener as the unit of analysis, leading to greater statistical power for the detection of smaller effect sizes (*e.g.* Feinberg *et al.*, 2011a, 2005; Puts *et al.*, 2006).

The only other vocal measurement which we found to correlate substantively with attractiveness ratings was F_0-*SD* in female scripted speech, although again this was not significant. This positive correlation indicates that a higher variation in voice pitch (i.e. a more dynamic, less monotone voice) was found to be more attractive, which is in line with preference for feminine traits in female voices, as a higher female F_0-*SD* is congruent with the sexual dimorphism of the trait (Henton, 1995). This supports the results of Leongómez *et al.* (in press), who found that speakers affect a higher F_0-*SD* in response to attractive individuals, and these responses are considered attractive to naïve listeners. Furthermore, Fischer *et al.* (2011) found that women exhibited greater pitch variation in spontaneous speech in days prior to ovulation relative to the rest of the menstrual cycle, when conception risk at its highest, and men rated women’s voices as more attractive during the fertile phase. However, in the current study, this correlation was not present in spontaneous speech recordings. Due to the controlled nature of the scripted recordings, pitch variation may be one of few ways for listeners to comparatively judge attractiveness between speakers. Introducing individual variation in spoken content may divert attention away from acoustic differences and toward the lexical content of the recordings, which may have more strongly influenced attractiveness ratings.

D_f was positively and significantly related to attractiveness ratings in male vowel sounds although we did see a positive, nonsignificant in female scripted speech. If D_f is indicative of apparent vocal tract length, we would expect to see a *negative* relationship between D_f and attractiveness (as lower D_f measures correspond to longer vocal tracts, which are associated with a larger body size) which was found by Feinberg *et al.* (2011b) using vowel stimuli, and by Hodges-Simeon *et al.* (2010) using spontaneous speech. Our data are unable to support these findings. Additionally, although Puts *et al.* (2012) suggests that P_f is more sexually dimorphic than D_f, and formant frequencies are proposed to be indicators of physical dominance (Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2007; Wolff & Puts, 2010), we did not find P_f to be independently associated with opposite-sex attractiveness in any condition for either sex.

HNR was only associated with attractiveness in male vowel sounds, where a negative (but non-significant) relationship was found, indicating that rougher male voices were considered more attractive than smooth, clear voices. Because HNR is indicative of age (Ferrand, 2002), male voices with lower HNR may sound older, and perhaps more attractive or masculine. The lack of observed relationship in other speech types may be due to the introduction of lexical content in these stimuli: listeners have more information on which to base their attractiveness ratings, and HNR may no longer be "important" for making these assessments. Bruckert *et al.* (2010) found that smoother, clearer voices (of morphed composite vocal averages) were more attractive than the individual voices in the composite, showing that HNR and attractiveness were significantly *positively* related. The present results show mostly *negative* relationships for men and positive relationships for women. Although none of these relationships are significant, a larger number of subjects may reveal clearer effects.

Taken as a whole, no particularly strong links between measurements and attractiveness ratings were found. This could be due to the number of stimuli which were examined – if a greater number of voices had been measured, perhaps clearer patterns would have emerged. Considering our small sample size, the relationships that we did find may be considered robust, but more subtle relationships may require more statistical power to uncover. Additionally, in our spontaneous speech recordings, stimulus donors responded to the question posed using their own words. This variation in content may give away information about the speakers' social status, intelligence, and personalities, which may have affected attractiveness measurements to a greater degree than any measurable vocal trait.

Lastly, vocal recordings were captured in a specified order, with vowels first, followed by scripted speech and ending with spontaneous speech. This prescribed order may contribute to the effects detailed here, however many effects are shown differently for men and women – if order is to be held solely responsible for these effects, gender differences are unlikely to have occurred in the fashions represented here.

2.4.4 Future Directions

Based on these findings, we suggest that measures and ratings made using vowel stimuli are not necessarily representative of more naturalistic types of speech for both male and female speakers. While useful for obtaining tightly controlled recordings of voices, vowel stimuli may not be reliable proxies for obtaining results that are broadly generalisable, and researchers may thus have cause to be wary about the use of vowel stimuli. The content-richness of spontaneous speech could be considered an experimental confound, potentially giving the listener information about the speakers' social status, educational background, and personality. Although standardised text poses its own issues by eliciting patterns of intonation and phrasing that are not encountered in naturalistic encounters, the use of scripted texts may achieve a balance between laboratory controls and ecological validity, by allowing participants to speak in the way that they choose while still maintaining lexical control. Spontaneous speech recorded in a language unknown to ratings participants (in the style of Knowles 2010, Fischer *et al.* 2011 and Leongómez *et al. in press*) may be an additional and useful way to control for lexical content.

2.5 Conclusions

Vowel sounds and scripted speech are utilised by researchers to control the content of voice stimuli, and these approaches are widely used. However, our results suggest that these types of controlled recordings do not always accurately capture a speaker's various vocal measurements when compared to spontaneous speech. Additionally, attractiveness ratings were significantly higher in spontaneous speech recordings than in vowel-sound or scripted speech recordings – and in males, rated attractiveness of vowel recordings were unrelated to attractiveness ratings of spontaneous speech. We also found mixed support for previous research which links acoustic measures to ratings of vocal attractiveness, suggesting that ecological validity may be an important factor in the relationship between acoustic measures and attractiveness. Researchers may thus have cause to be wary about generalising the results of studies based on vowel sounds of male speakers and for extracting acoustic measures from such vocal samples. The reading of scripted texts may be a valid compromise between laboratory controls and ecological validity, allowing for control over lexical content while capturing a more realistic type of stimulus to present to listeners.

Because of the lack of consistent relationships observed between sustained vowels and more naturalistic stimuli, throughout the remainder of this thesis, the use of vowel stimuli will be abandoned in favour of scripted speech stimuli.

CHAPTER 3: The Time Course of Vocal Attribution: Masculinity, Attractiveness, and Trustworthiness Perceptions at Varied Exposure Times

Chapter 2 focused on differences in measurements and perception across types of stimuli, but stimuli of differing lengths may also elicit variable trait attribution. Studies using facial stimuli have demonstrated that reliable trait attributions can be made at exposure times of $\leq$100ms. To our knowledge, the time course of vocal trait attribution has not yet been investigated. We obtained masculinity, attractiveness, and trustworthiness ratings of voices after one, five, and 22+seconds of exposure time (Study 1), and after one, two, five, and ten seconds of exposure (Study 2). While we show that voice fundamental frequency (F_0) changes over the course of an utterance, results of both studies suggest that masculinity ratings are assessed after as little as one second of exposure time. Attractiveness ratings seem to take longer to develop, with stimuli of 5 seconds or longer eliciting judgements similar to those made after unconstrained exposure. Trustworthiness ratings were the slowest to develop, with high variability in shortened exposure lengths. These results may inform the way future voice perception research is designed with respect to stimulus length and content. We recommend that researchers interested in more subjective vocal attributions, such as attractiveness, trustworthiness, or other social traits, use stimuli at least five seconds in duration.

3.1 Introduction

It is clear that voice stimuli elicit a variety of trait attributions. Stimuli used in studies of voice perception, however, reflect varying content. Many studies use the isolated vowel sounds *a-e-i-o-u* (e.g Bruckert, Liénard, Lacroix, Kreutzer and Leboucher, 2006; Collins, 2000; Feinberg *et al.*, 2006; Feinberg, DeBruine, Jones and Perrett, 2008; Feinberg, Jones, Little, Burt and Perrett, 2005; Little, Connely, Feinberg, Jones and Roberts, 2011), while others use memorised or scripted texts of varying length. For example, Lander (2008) used a short memorised sentence of ~2 seconds in length, while Puts, Gaulin and Verdolini (2006) used a much longer scripted text, resulting in stimuli which averaged ~18 seconds in length. Singular spoken words have also been used (Apicella *et al.*, 2007; Apicella & Feinberg, 2009). Overall, stimuli used in experiments of voice perception vary both in length and ecological validity, though results of these experiments remain generally consistent. However, findings by Jones, Feinberg, Debruine, Little and Vukovic, (2008) suggest that the content of what is said can interact with the physical properties of voices in the perception of attractiveness. This provides additional cause to question whether the stimuli used may have important consequences for the perception of vocal qualities.

A growing body of research shows clear, strong relationships between voice pitch (fundamental frequency, or F_0) and listener perceptions. In men, lower F_0 reliably predicts ratings of attractiveness (Collins, 2000; Feinberg, Jones, Little, et al., 2005; Feinberg, 2008; Feinberg et al., 2006; Saxton et al., 2006) and ratings of masculinity (Collins, 2000). In these cases, voices follow the same preference pattern as faces – women prefer both masculine voices and faces in men (Feinberg, DeBruine, Jones and Little, 2008), and men prefer feminine voices and faces in women (Fraccaro *et al.*, 2010). This similarity has led some researchers to consider voices and faces as related cues signalling mate quality (Collins & Missing, 2003; Feinberg, DeBruine, Jones and Little, 2008; Feinberg, 2008; Feinberg, Jones, DeBruine, *et al.*, 2005). F_0 in men has also been related to other aspects of physical appearance such as height and hirsuteness (Collins, 2000). Alongside physical perception and attractiveness, personality traits can also be inferred from the voice, including extraversion, truthfulness, empathy and nervousness (Apple *et al.*, 1979; Scherer, 1978).

Because people form trait impressions from voices, and because it has been suggested that these impressions influence social behaviour, including male competition and mate choice (Feinberg, 2008; Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2011; Wolff & Puts, 2010), the speed with which these impressions are formed is of scientific interest. It is still unknown whether trait inferences about voices are fast and unreflective, or if they are slower, more deliberative responses. The consistent pattern of results, even with extremely short, non-word stimuli, suggests that these impressions may indeed be reflexive responses.

While perceptions based on voice pitch are highly replicable, it should be noted that the vocal qualities of one person may not be the same between utterances, or even within a singular utterance. In other words, an individual's voice is changeable across time to some extent. According to the source-filter theory of speech production (Fant, 1960), the anatomical nature of speech production plays a role in this. Speech requires a coordination of physical processes - respiratory muscles control breathing, while the larynx, tongue and mouth turn expiration into words. Air pressure in the lungs is at its highest immediately following inspiration, when the lungs are full, and this corresponds with the beginning of the speech utterance. Because F_0 is affected by subglottal air pressure, there is a natural pattern of intonation in human speech, with the first words of a sentence having a higher F_0 than the rest of the utterance (Fant, 1960; Lieberman, 1967). Following this initial peak in F_0, subglottal air pressure gradually decreases over the remaining utterance, which is finished with an abrupt drop in F_0 as remaining air in the lungs is expelled, in preparation for another inspiration (Fant, 1960; Lieberman, 1967; Vaissière, 2005). Thus, very short stimuli may not accurately represent subject's vocal characteristics, particularly if they over-represent the beginning of an utterance. This may include stimuli that consist of one spoken word (*e.g.* Apicella *et al.*, 2007; Apicella & Feinberg, 2009) or short sentences (*e.g.* Lander, 2008).

Since F_0 is at its highest at the beginning of an utterance, and considering how trait attributions vary based on F_0, the speed with which humans form stable (i.e. unchangeable) perceptual judgements about a voice could have impact on current research. Previous experiments have shown that facial attractiveness is assessed extremely rapidly, after exposures of only 100ms (Locher, Unger, Sociedade and Wahl, 1993; Willis & Todorov, 2006), or even in as little as 13ms (Olson & Marshuetz, 2005). Personality traits including likeability, trustworthiness, competence, and aggressiveness are also reliably assessed after exposure to facial stimuli 33ms to 100ms in length (Ballew & Todorov, 2007; Todorov *et al.*, 2009; Willis & Todorov, 2006). It is possible that humans may make trait judgements about voices in a similarly rapid fashion, and that these initial judgements may remain unchanged even after F_0 has changed over the course of the vocalisation. Conversely, F_0disparity could significantly impact listener judgements about a voice, leaving room for initial judgements to change with additional information available from longer utterances.

The aim of the current experiment is to determine how listener perceptions of physical, aesthetic, and personality traits (using masculinity, attractiveness and trustworthiness) relate to stimulus length. These judgements are included in this study because they are commonly researched traits – perceptions of all three traits have been routinely linked to vocal qualities. Furthermore, masculinity, attractiveness and trustworthiness are commonly-studied facial attributes, allowing for links between voice and face perception to be examined. While these three traits may not be purely orthogonal, differences between how these traits are perceived (should they be uncovered) could be revealing, and their prevalence in existing literature makes them relevant candidates for examination.

Additionally, we measured F_0 of the voices to examine how F_0 is related to each perceived trait, and also to explore how stable this relationship remains over time. We employ a method used in the study of the time course of face perception, in which the reliability of the ratings over a time course is investigated by comparing ratings made at shortened-exposure presentations to those made when viewing times are unconstrained. In Study 1, twenty male and twenty female voices were rated for the perceived traits of masculinity, attractiveness, and trustworthiness. The experiment was between-subjects for stimulus length and within-subjects for rated traits: listeners rated either the first one second of the recording, the first five seconds, or the full recording (22+ seconds) for all three traits. In Study 2, the experiment was within-subjects for stimulus length and between-subjects for rated traits: listeners rated each time segment of the voices for one trait (masculinity, attractiveness, or trustworthiness). Here, recordings were divided into the first one second of the recording, the first two seconds, the first five seconds, or the first 10 seconds.

3.2 Study 1

3.2.1 Methods

3.2.1.1 Stimuli

20 male and 20 female participants aged 18-44 (M = 21.7, SD 3.8) were recorded reading standardised text which described the weather in the UK, which was taken from a newspaper. This script was chosen for its length and general neutrality of context, and was the same as recorded in the scripted speech stimuli in Chapter 2. All participants were psychology students at the University of Stirling. Participation was voluntary and students were given course credit for their participation.

Recordings were obtained using a Canon XM2 3CCD digital video camera with a PCM dynamic stereo microphone. This was done in a quiet room at a distance of 110cm. Video and audio tracks were recorded directly to hard disk and encoded using 16-bit quantization with a 48kHz sampling rate using Windows Movie Maker v.2.1.4027.0. Audio tracks were isolated using BeeThink 2.0 software and converted from stereo to mono using Audacity v.1.2.6. These recordings were used to create three stimulus sets: the utterance in its entirety (mean length 26.2s ± 2.1s, range 22.6s – 29.1s), the first 5 seconds of the utterance, and the first 1 second of the utterance. These time intervals were rigidly adhered to, with all stimuli being trimmed to ±10ms of 1 and 5 seconds. Although this means that some words were cut off in the process of being uttered, and that differing speech rates led some of the stimuli to contain more words than others, we were strict with time intervals because our question involved the stability of attributions over time, not the amount of verbal content. Mean pitch measurements were obtained using Praat v.5.2.01 (Boersma & Weenink, 2009) for each stimulus length. Pitch floors and ceilings were set as per the programmers' specifications, with a range of 75-300Hz for male voices, and 100-500Hz for female voices.

3.2.1.2 Subjects

Participants in the full-recording stimulus condition (*N* = 16) were 8 males and 8 females aged 17-20(*M* = 18.6 years, *SD*1.1). Participants in the 5-second condition (*N* = 19) were 5 males and 14 females aged 18-46 (mean age = 23.7, *SD* = 7.4). Participants in the 1-second condition (*N* = 29) were 5 males and 24 females aged 18-42 (mean age = 19.8, *SD* = 4.4).Similar values for *N* have been used in previous experiments using facial and vocal stimuli (Jones, Main, DeBruine, Little and Welling, 2010; Little *et al.*, 2011; Willis & Todorov, 2006). All participants either took part online or in the lab in the Psychology department at the University of Stirling. Previous experiments have shown that similar results are obtained whether testing in the lab or via the internet when testing judgements of faces and voices (Buchanan, 2000; Feinberg, Jones, Little, *et al.*, 2005; Jones, Perrett, *et al.*, 2005; Wilson & Daly, 2004), and such testing is now commonplace within psychology research (Kraut *et al.*, 2004). Participants in all conditions included psychology students at the University of Stirling, though members of the general public also had access to the study via the lab webpage. Students who participated did so voluntarily and received course credit as compensation; they were recruited via the University of Stirling online participant database. Members of the public who took part did so voluntarily and without compensation. All phases of the experiment were approved by the ethics committee at the University of Stirling.

3.2.1.3 Procedure

The three stimulus sets were used as three different conditions of the same experiment. Within each condition, recordings were fully randomised and study participants were asked to rate each voice for attractiveness, masculinity and trustworthiness on a 7-point Likert scale. While the voice recordings were timed at one second, five seconds, or 22+ seconds in length, there was no time limit between trials, allowing listeners to make their ratings in their own time. All three trait ratings scales appeared together on the same page and were rated on the same screen, with attractiveness appearing first, followed by masculinity, and trustworthiness last. Participants were unable to select a rating until the stimulus had played in its entirety. After rating all three questions, a button moved the participant on to the next trial.

3.2.2 Results

Ratings of each voice were combined across listeners to result in a mean rating for each voice stimulus. In reporting degrees of freedom, Greenhouse-Geisser corrections are reported in ANOVAs which failed Mauchly's test of sphericity. The ANOVA models used time course as the within-subjects factor (three levels) and sex of speaker as a between-subjects factor.

Factor analysis on the three trait ratings (in all time steps) indicates that for male voices, all three attributes load together positively on the same dimension (see Table 3.1). For female voices, attractiveness and masculinity load positively together, while masculinity loads negatively on the same factor. While this shows that the traits are not strictly orthogonal, it does not imply that the time-course of perception of these traits are the same –

any difference in patterns of perception are certainly not excluded by these shared relationships.

Table 3.1 *Factor loadings, Eigenvalues and percentage variance explained using factor analysis (principal components, Study 1).*

	Male			Female		
	1s	5s	30s	1s	5s	30s
Masculinity	.902	.831	.848	-.895	-.814	-.887
Attractiveness	.913	.938	.892	.951	.933	.923
Trustworthiness	.887	.919	.860	.857	.799	.868
Eigenvalue	2.43	2.41	2.26	2.44	2.17	2.39
% Variance	81.1%	80.5%	75.2%	81.4%	72.4%	79.7%

3.2.2.1 F_0 over time

Pitch measurements correlated strongly and significantly across the time course. F_0 in the 1-second and 5-second stimulus groups correlated significantly, $r(40) = .954$, $p < .001$. Likewise, F_0 in the 5-second stimulus group correlated significantly with the 22+second stimulus group, $r(40) = .995$, $p < .001$, and 1-second F_0 significantly correlated with F_0 in the 22+second stimuli, $r(40) = .939$, $p < .001$. Despite these robust correlations, a repeated-measures ANOVA revealed significant differences in F_0 between stimulus sets, with F_0 decreasing over time, $F(1.09,41.28) = 69.69$, $p < .001$,$\eta_p^2 = .647$. See Figure 3.1. Post-hoc contrasts showed that F_0 fell between the 1-second and 5-second recordings, $F(1.09,41.28) = 70.22$, $p < .001$, $\eta_p^2 = .649$, and again fell between the 5-second and 22+second recordings, $F(1.09,41.28) = 19.14$, $p < .001$, $\eta_p^2 = .335$. No significant interaction between stimulus length and sex of speaker was observed, $F(1.09,41.28) = 2.25$, $p = .13$, $\eta_p^2 = .056$, indicating that both men and women follow the same natural intonation curve within an utterance. Means and standard deviations are reported in Table 3.2.

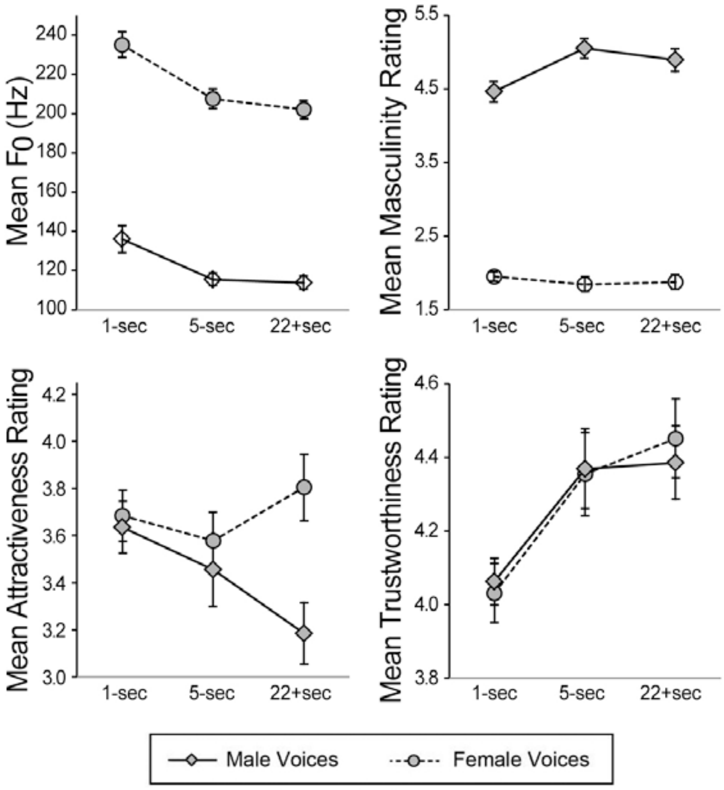

Figure 3.1
Mean F_0 *(a) and ratings of masculinity (b), attractiveness (c), and trustworthiness (d) plotted in all time steps (Study 1). Error bars represent standard error.*

Table 3.2 *Means and standard deviations of pitch measurements and ratings judgements (Study 1).*

Sex of stimulus	Stimulus length	Pitch(Hz)		Masculinity		Attractiveness		Trustworthiness	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Female (n = 20)	1-sec	235.2	28.9	1.95	0.28	3.68	0.49	3.83	0.36
	5-sec	207.7	22.0	1.84	0.44	3.58	0.53	4.16	0.51
	22+sec	202.1	21.2	1.89	0.43	3.81	0.63	4.25	0.48
Male (n = 20)	1-sec	136.1	31.2	4.47	0.61	3.63	0.49	3.86	0.28
	5-sec	115.4	15.7	5.05	0.61	3.46	0.69	4.17	0.48
	22+sec	113.7	14.3	4.89	0.70	3.19	0.59	4.19	0.45

3.2.2.2 Perceived Masculinity

In both female and male voices, masculinity ratings correlated across stimulus sets. 1-second stimuli were rated similar in masculinity to 5-second stimuli, $r(40) = .978$, $p < .001$, and the 22+second stimuli $r(40) = .981$, $p < .001$. 5-second masculinity likewise correlated significantly with 22+second masculinity, $r(40) = .979$, $p < .001$. Again, despite these correlations, in a repeated-measures ANOVA, a significant main effect of stimulus length on masculinity judgements was observed, $F(2,76) = 11.68$, $p < .001$, $\eta_p^2 = .235$, indicating that masculinity ratings increased with longer exposure times. There was also a significant interaction between speaker sex and stimulus length, $F(2,76) = 24.30$, $p < .001$, $\eta_p^2 = .390$, with men being rated as consistently more masculine than women with respect to time course.

To parse the interaction between masculinity and sex of speaker, further analysis was divided by stimulus sex. No significant differences were revealed across the time course for women, $F(2,38) = 1.75$, $p = .19$, $\eta_p^2 = .084$, that is, women were rated at a consistent masculinity level regardless of exposure length. The difference in masculinity ratings across the time course was driven by the male voices, $F(2,38) = 25.88$, $p < .001$, $\eta_p^2 = .577$. Post-hoc contrasts revealed this effect was most pronounced between 1- and 5-second groups, with male voices rated as significantly more masculine after five seconds than after one second, $F(1,19) = 51.65$, $p < .001$, $\eta_p^2 = .731$. No significant differences were revealed between the 5- and 22+second groups, $F(1,19) = 3.16$,$p = .09$, $\eta_p^2 = .143$. Considering that masculinity ratings of male voices were found to be significantly greater in the 5-second condition than in the 1-second condition, and that F_0 dropped significantly between these two time intervals, the relationship between these changes was examined. The amount of change in F_0 was not significantly correlated with the degree of change in masculinity ratings, $r(20) = -.24$, $p = .30$.

3.2.2.3 Perceived Attractiveness

In both male and female voices, attractiveness ratings in all stimulus groups correlated significantly across the time course. 1-second attractiveness correlated significantly with 5-second attractiveness, $r(40) = .709$, $p < .001$, and with 22+second attractiveness, $r(40) = .613$, $p < .001$. 5-second attractiveness also correlated significantly with 22+second attractiveness, $r(40) = .715$, $p < .001$. There was a significant main effect of stimulus length on attractiveness ratings in a repeated measures ANOVA, $F(2,76) = 3.30$, $p = .04$, $\eta_p^2 = .080$, and a significant effect of speaker sex, $F(2,76) = 9.87$, $p < .001$, $\eta_p^2 = .206$. Female voices increased in attractiveness as the time course progressed, $F(2,38) = 3.40$, $p = .04$, $\eta_p^2 = .152$. Post-hoc contrasts revealed no significant differences in attractiveness ratings for women's voices between the 1- and 5-second exposure times, $F(1,19) = 1.99$, $p = .18$, $\eta_p^2 = .095$, however, women's voices were rated as more attractive after 22+seconds than in the 5-second stimulus group, $F(1,19) = 7.56$, $p = .01$, $\eta_p^2 = .285$. Male voices, conversely, decreased in attractiveness as the time course progressed, $F(2,38) = 8.57$, $p = .001$, $\eta_p^2 = .311$. This decrease was not significant between the 1- and 5-second exposure times, $F(1,19) = 2.42$, $p = .14$, $\eta_p^2 = .113$, but was significant between the 5- and 22+second exposure times, $F(1,19) = 6.39$, $p = .02$, $\eta_p^2 = .252$.

3.2.2.4 Perceived Trustworthiness

1-second trustworthiness correlated significantly with 5-second trustworthiness, $r(40) = .495$, $p = .001$, and with 22+second trustworthiness, $r(40) = .379$, $p = .016$. 5-second trustworthiness significantly correlated with 22+second trustworthiness, $r(40) = .581$, $p < .001$. We observed a significant main effect of stimulus length on trustworthiness ratings in a repeated-measures ANOVA for voice trustworthiness, $F(2,76) = 16.35$, $p < .001$, $\eta_p^2 = .301$. No effect of speaker sex was observed, $F(2,76) = 0.27$, $p = .76$, $\eta_p^2 = .007$. Post-hoc contrasts showed that higher trustworthiness ratings were given after the 5-second exposure than after the 1-second exposure, $F(1,38) = 20.68$, $p < .001$, $\eta_p^2 = .352$. No significant difference was revealed between the 5-second and 22+second groups, $F(1,38) = 0.67$, $p = .42$, $\eta_p^2 = .017$.

3.2.2.5 Variance Explained

The amount of variance (r^2) was calculated for both F_0 and ratings judgements explained by shortened-stimulus sets, using perceptions formed without time constraints (i.e. the longest stimuli presentation times) as a baseline. These values were obtained by using the correlation coefficients calculated across the time course for each trait. See Figure 3.2.

For female voices, r^2 for F_0 over time increased as exposure times grew longer (1-sec: .696, 5-sec: .912). r^2 values for attractiveness and trustworthiness judgements also increased with exposure time (attractiveness: 1-sec: .490, 5-sec: .659; trustworthiness: 1-sec: .180, 5-sec: .411). r^2 values for masculinity judgements did not increase over the time course, and remained somewhat consistent across both shortened lengths (1-sec: .679, 5-sec: .650).

For male voices, r^2 for F_0 over time again increased as exposure times grew longer (1-sec: .438, 5-sec: .929). r^2 values for attractiveness and trustworthiness judgements also increased with exposure time (attractiveness: 1-sec: .398, 5-sec: .537; trustworthiness: 1-sec: .110, 5-sec: .269). Similar to the results for female voices, r^2 values for masculinity judgements did not increase over the time course (1-sec: .731, 5-sec: .669). See Figure 3.2.

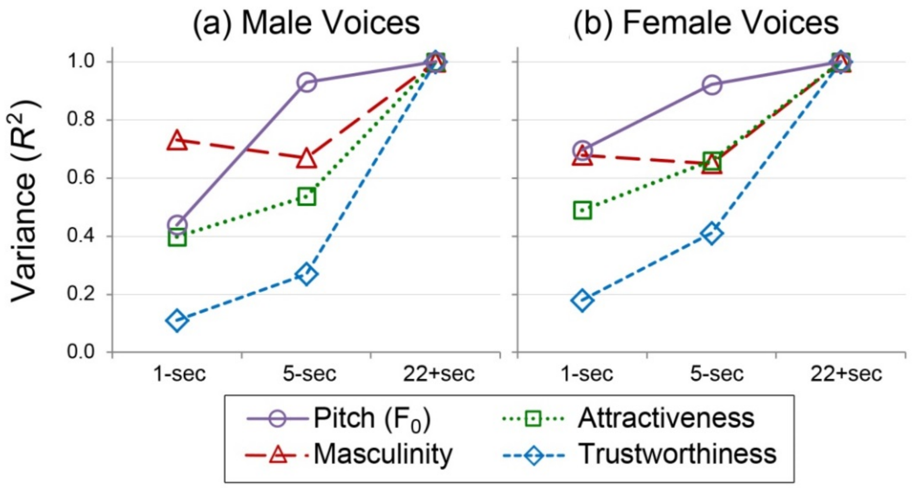

Figure 3.2
Variance explained (Study 1). Percentage of variance explained in full-recording ratings by pitch and ratings in shortened stimuli for female voices (a) and male voices (b).

3.2.2.6 Relationship Between F_0 and Perceived Traits

The relationships between F_0 and perceived traits for male and female voices were analysed separately in each stimulus set, to determine whether these correlations were influenced by presentation time. See Table 3.2. One-tailed *p*-values are reported due to the predicted directional effect of F_0 on perceptions of the three traits on voices of both sexes, based on well-established findings (Collins & Missing, 2003; Collins, 2000; Feinberg, DeBruine, Jones and Little, 2008; Feinberg, 2008; Klofstad *et al.*, 2012; Tigue *et al.*, 2012).

Table 3.3 *Correlations (Pearson r) between* F_0 *measurements and perceived traits.*

Time Step	Masculinity		Attractiveness		Trustworthiness	
	Male	Female	Male	Female	Male	Female
1-sec	-0.28	-0.17	-0.33	0.07	-0.09	-0.18
5-sec	-0.52**	-0.53**	-.060	0.22	-0.34	0.21
22+sec	-0.42*	-0.32	-0.49*	0.15	-0.31	0.34

Note: one-tailed p-values are reported for masculinity and attractiveness ratings. Trustworthiness ratings are reported using two-tailed p-values.
* p < .05, ** p < .01

Masculinity ratings were significantly and negatively correlated with F_0 in male voices in the 5-second and 22+second conditions, with a lower F_0 associated with higher masculinity ratings (see Table 3.2). A non-significant negative relationship was observed in the 1-second condition. A similar directionality was present in female voices, with lower F_0 significantly correlating with higher masculinity ratings in the 5-second condition. This relationship was not significant in the 22+second condition, although a negative trend was observed. No significant relationship was observed in the 1-second condition. Attractiveness ratings were only significantly related to F_0 in male voices in the 5- and 22+second conditions, with lower-pitched voices being rated as more attractive than higher-pitched voices. The relationship between F_0 and attractiveness was not significant in the 1-second condition. We found no significant relationship between F_0 and attractiveness in female voices in any of the three conditions. Trustworthiness ratings were not significantly correlated with F_0 in either sex, in any stimulus condition.

3.3 Study 2

The time course of vocal perception was further examined by repeating Study 1 using stimuli which had been trimmed to different lengths – whereas ratings were obtained in Study 1 after exposure times of one, five, and 22+ seconds, Study 2 was designed to obtain ratings after exposure times of one, two, five, and 10 seconds. Additionally, a between-subjects design with respect to stimulus length was used for Study 1, with separate groups of raters responding to each time-step. In Study 2, a within-subjects design was used with respect to stimulus length. Listeners rated each of the time steps for one vocal trait only; separate groups of raters were used for the traits of masculinity, attractiveness and trustworthiness. Additionally, separate groups of raters were used for male and female voices within each trait.

3.3.1 Methods

3.3.1.1 Stimuli

Stimuli were the same as in Study 1, trimmed to one, two, five, and ten seconds of exposure time. All methods of stimuli manufacture and measurement were identical to Study 1. So as not to fatigue participants, stimuli *N* for Study 2 were scaled back to 15 male and 15 female voices, randomly selected from the stimuli used in Study 1.

3.3.1.2 Subjects

Participants rating male masculinity (*N* = 21) were 10 males and 11 females aged 17-27 (*M* = 18.9 years, *SD* 2.1 years). Female masculinity (*N* = 21) was assessed by 10 male listeners and 11 female listeners aged 17-23 (*M* = 18.6 years, *SD* 1.3 years). Male attractiveness (*N* = 19) was assessed by female listeners aged 17-23 (*M* = 18.9 years, *SD* 1.4 years). Female attractiveness (*N* = 22) was assessed by male listeners aged 18-21 (*M* = 19.1 years, *SD* 0.9 years). Male trustworthiness (*N* = 28) was assessed by 3 men and 25 women aged 18-43 (*M* = 20.4 years, *SD* 4.8 years). Female trustworthiness (*N* = 26) was assessed by 7 male listeners 19 female listeners aged 17-24 (*M* = 19.2 years, *SD* 1.5 years). Participants in all conditions included psychology students at The University of Stirling. Students who participated did so voluntarily and received course credit as compensation; they were recruited via the University of Stirling online participant database.

3.3.1.3 Procedure

In each of the six conditions, participants rated the voices for one trait (masculinity, attractiveness, or trustworthiness) on a 7-point Likert scale. Participants rated the stimuli in four blocks, presented in order of presentation time: one, two, five, and ten seconds. Stimuli within each block were randomised, however, the order of the blocks was fixed. The blocks were presented in this order to reflect the time course of a naturalistic interaction. Participants were instructed to “rate the voice for [trait] once the clip is finished,” and selected their response on a 7-point Likert scale. Once the rating was selected, participants were automatically moved to the next trial. Participants were unable to select a rating until the stimulus had played in its entirety.

3.3.2 Results

As in Study 1, ratings of each voice were combined across listeners to result in a mean rating for each voice stimulus. For masculinity and trustworthiness ratings, this was done separately for male and female listeners, creating a mean female-rated value and mean male-rated value, in addition to an overall mean value. Only opposite-sex ratings were used to calculate attractiveness judgements, thus, mean values reported are from opposite-sex listeners only. In reporting degrees of freedom, Greenhouse-Geisser corrections are reported in ANOVAs which failed Mauchly’s test of sphericity. The ANOVA models used time course (three levels) and sex of listener (two levels) as the within-subjects factor, and sex of speaker as a between-subjects factor. F_0 and attractiveness models do not include listener sex as a within-subjects factor.

Factor analysis of the three assessed traits revealed a similar pattern of results to Study 1: male voices show all three traits loading together positively on one factor, and female voices load attractiveness and trustworthiness together, while masculinity loads negatively on the same factor. See Table 3.4.

Table 3.4 *Factor loadings, Eigenvalues and percentage variance explained using factor analysis (principal components, Study 2).*

	Male				Female			
	1s	2s	5s	10s	1s	2s	5s	10s
Masculinity	.862	.923	.917	.881	-.879	-.885	-.933	-.884
Attractiveness	.924	.908	.936	.917	.959	.974	.982	.973
Trustworthiness	.904	.947	.960	.936	.826	.839	.878	.822
Eigenvalue	2.42	2.57	2.64	2.50	2.37	2.43	2.61	2.40
% Variance	80.5%	85.8%	88.0%	83.2%	70.1%	81.1%	86.9%	80.1%

3.3.2.1 F_0 Over Time

As in Study 1, mean F_0 was strongly and significantly correlated across stimulus groups. 1-second F_0 correlated with 2-second F_0, $r(30) = .992$, $p < .001$, with 5-second F_0, $r(30) = .986$, $p < .001$, and with 10-second F_0, $r(30) = .984$, $p < .001$. 2-second F_0 correlated significantly with 5-second F_0, $r(30) = .993$, $p < .001$, and with 10-second F_0, $r(30) = .994$, $p < .001$. 5-second F_0 also correlated with 10-second F_0, $r(30) = .999$, $p < .001$. A repeated-measures ANOVA revealed a significant main effect of stimulus length on F_0, $F(3,84) = 114.18$, $p < .001$, $\eta_p^2 = .803$. Planned contrasts showed a significant lowering of F_0 between all time steps: 1-second F_0 was significantly higher than 2-second F_0, $F(1,28) = 86.89$, $p < .001$, $\eta_p^2 = .756$; 2-second F_0 was significantly higher than 5-second F_0, $F(1,28) = 45.90$, $p < .001$, $\eta_p^2 = .621$; 5-second F_0 was significantly higher than 10-second F_0, $F(1,28) = 47.00$, $p < .001$, $\eta_p^2 = .627$. A significant interaction between sex of speaker and stimulus length was also observed, $F(1.507,42.204) = 8.67$, $p = .002$, $\eta_p^2 = .237$. Planned contrasts indicated that female F_0 decreased to a greater degree, $F(1.618,22.654) = 63.81$, $p < .001$, $\eta_p^2 = .820$, than did male F_0, $F(1.146,16.041) = 55.08$, $p < .001$, $\eta_p^2 = .797$. See Figure 3.3. Means and standard deviations are reported in Table 3.3.

Table 3.5 *Means and standard deviations of pitch measurements and ratings judgements (Study 2).*

Sex of stimulus	Stimulus length	Pitch(Hz)		Masculinity		Attractiveness		Trustworthiness	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Female (n = 15)	1-sec	239.5	24.6	3.37	1.05	3.45	0.92	4.30	0.61
	2-sec	223.5	23.9	3.49	1.17	3.75	0.89	4.27	0.59
	5-sec	212.8	22.5	3.48	1.13	3.79	0.97	4.34	0.54
	10-sec	209.3	22.2	3.43	1.12	3.95	1.00	4.28	0.62
Male (n=15)	1-sec	132.2	20.1	4.07	1.19	3.49	0.82	4.20	0.57
	2-sec	122.2	16.7	4.18	1.26	3.40	1.03	4.17	0.66
	5-sec	117.0	15.8	4.18	1.30	3.34	1.11	4.10	0.74
	10-sec	114.9	15.0	4.32	1.16	3.41	1.16	4.22	0.73

3.3.2.2 Perceived Masculinity

Masculinity ratings of male and female voices were significantly correlated across all time steps. 1-second masculinity correlated with 2-second masculinity, $r(30) = .977$, $p < .001$, with 5-second masculinity, $r(30) = .965$, $p < .001$, and with 10-second masculinity, $r(30) = .953$, $p < .001$. 2-second masculinity correlated significantly with 5-second masculinity, $r(30) = .971$, $p < .001$, and with 10-second masculinity, $r(30) = .960$, $p < .001$. 5-second masculinity also correlated with 10-second masculinity, $r(30) = .969$, $p < .001$.

In a repeated-measures ANOVA for perceived masculinity, there was a close-to-significant effect of time course, $F(3,84) = 2.56$, $p = .06$, $\eta_p^2 = .084$, and a significant interaction between sex of listener and sex of voice, $F(1,28) = 6.89$, $p = .01$, $\eta_p^2 = .197$. Post-hoc contrasts showed that women gave higher masculinity ratings to female voices than male listeners did, $F(1,14) = 6.26$, $p = .025$, $\eta_p^2 = .309$. No other significant effects or interactions were observed for female voices (all $F < 0.95$, all $p > .43$, all $\eta_p^2 < .063$). For male voices, there was a significant main effect of time course, indicating that male voices were considered more masculine as exposure times increased, $F(3,42) = 3.74$, $p = .02$, $\eta_p^2 = .211$. See Figure 3.3. No other significant effects or interactions were observed for male voices (all $F < 0.87$, all $p > .37$, all $\eta_p^2 < .059$). Post-hoc contrasts revealed that male voices were perceived as more masculine after the 2-sec exposure compared to the 1-sec exposure, $F(1,14) = 4.46$, $p = .05$, $\eta_p^2 = .242$. See Table 3.3. No other significant changes in masculinity ratings were observed across the time course (all $F < 3.11$, all $p > .10$, all $\eta_p^2 < .182$). Taking into consideration the difference in F_0 between the 1- and 2-second trials, the relationship between the degree of change in F_0 and the degree of change in perceived masculinity was examined. The amount of change in F_0 was not significantly correlated with the degree of change in masculinity ratings, $r(15) = .32$, $p = .24$.

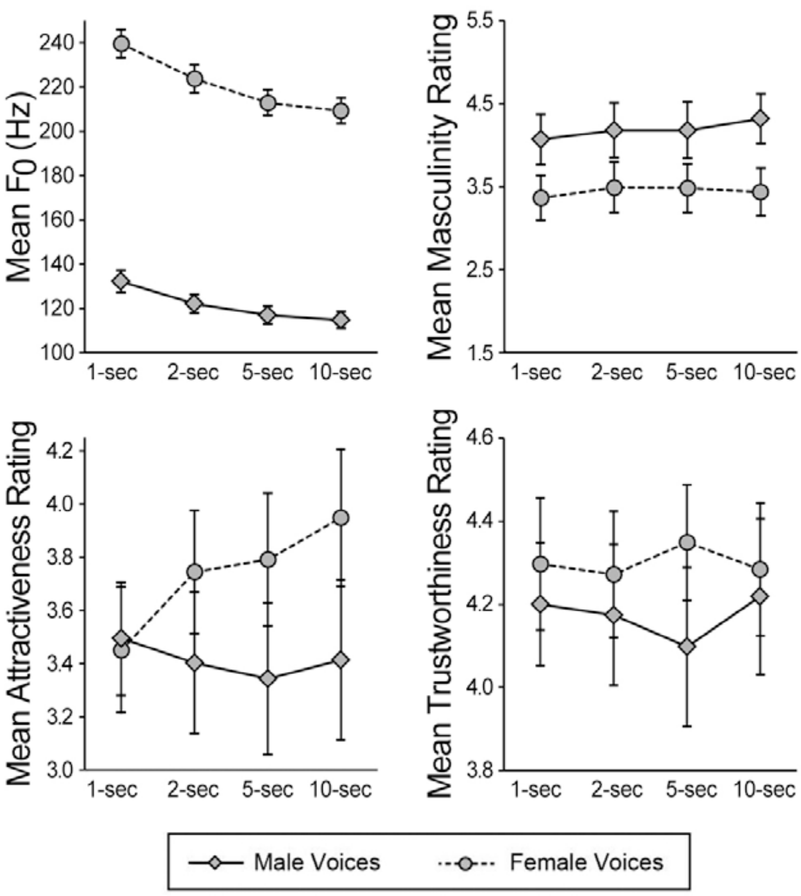

Figure 3.3
Mean F_0 *(a) and ratings of masculinity (b), attractiveness (c), and trustworthiness (d) plotted in all time steps (Study 2). Error bars represent standard error.*

3.3.2.3 Perceived Attractiveness

Attractiveness ratings of male and female voices were also significantly correlated across all time steps. 1-second attractiveness correlated with 2-second attractiveness, $r(30) = .868$, $p < .001$, with 5-second attractiveness, $r(30) = .836$, $p < .001$, and with 10-second attractiveness, $r(30) = .834$, $p < .001$. 2-second attractiveness correlated significantly with 5-second attractiveness, $r(30) = .952$, $p < .001$, and with 10-second attractiveness, $r(30) = .951$, $p < .001$. 5-second attractiveness also correlated with 10-second attractiveness, $r(30) = .970$, $p < .001$. ANOVA results did not reveal a main effect of stimulus length on attractiveness judgements, $F(1.961,54.912) = 2.44$, $p = .10$, $\eta_p^2 = .080$, however, a significant interaction between sex of speaker and stimulus length was found, $F(1.961,54.912) = 5.45$, $p = .002$, $\eta_p^2 = .163$. Further analysing the results by sex of stimulus, attractiveness ratings of female voices increased over the time course, $F(1.992,27.883) = 9.15$, $p < .001$, $\eta_p^2 = .395$. Post-hoc contrasts showed that female voice attractiveness increased between the 1-sec and 2-sec exposures, $F(1,14) = 5.74$, $p = .031$, $\eta_p^2 = .291$, and between the 5-sec and 10-sec exposures, $F(1,14) = 9.92$, $p = .007$, $\eta_p^2 = .415$. No significant difference was observed between the 2-sec and 5-sec exposures, $F(1,14) = 0.28$, $p = .61$, $\eta_p^2 = .019$. See Table 3.3. There was no significant change over time for male voices, $F(1.717,24.033) = 0.53$, $p = .57$, $\eta_p^2 = .037$. See Figure 3.3.

3.3.2.4 Perceived Trustworthiness

Trustworthiness ratings of male and female voices were significantly correlated across all time steps. 1-second trustworthiness correlated with 2-second trustworthiness, $r(30) = .862$, $p < .001$, with 5-second trustworthiness, $r(30) = .748$, $p < .001$, and with 10-second trustworthiness, $r(30) = .769$, $p < .001$. 2-second trustworthiness correlated significantly with 5-second trustworthiness, $r(30) = .818$, $p < .001$, and with 10-second trustworthiness, $r(30) = .825$, $p < .001$. 5-second trustworthiness also correlated with 10-second trustworthiness, $r(30) = .886$, $p < .001$.

For trustworthiness ratings, there was no significant effect of time course, $F(3,84) = 1.01$, $p = .39$, $\eta_p^2 = .035$, or sex of rater, $F(1,28) = 0.002$, $p = .97$, $\eta_p^2 = .000$, indicating that mean trustworthiness ratings of both male and female voices did not differ across time steps. See Figure 3.3. There were significant interactions between speaker sex and listener sex, $F(1,28) = 4.49$, $p = .04$, $\eta_p^2 = .138$, and between time course and sex of listener, $F(3,84) = 5.01$, $p = .003$, $\eta_p^2 = .152$. When the ratings were analysed independently by sex of speaker, there were no significant main effects or interactions observed for female voices (all $F < 1.80$, all $p > .20$, all $\eta_p^2 < .114$); in male voices, there was a significant interaction between time course and sex of listener, $F(3,42) = 5.08$, $p = .004$, $\eta_p^2 = .266$, indicating that female listeners found male voices more trustworthy than male listeners did with respect to exposure time. The effect of the interaction was significant between the 5- and 10-second exposure times, $F(1,14) = 11.74$, $p = .004$, $\eta_p^2 = .456$, but not between the shorter stimuli (all $F < 2.70$, all $p > .12$, all $\eta_p^2 < .162$).

3.3.2.5 Variance Explained

Using the 10-second condition as our baseline measurement, r^2 was calculated in order to reflect the variance explained by the 1-, 2-, and 5-second ratings of all three attributes (masculinity, attractiveness, trustworthiness) and for F_0. These values were calculated using the correlation coefficients for the ratings and measurements across the time course. For male voices, r^2 for F_0 over time increased steadily as exposure times grew longer (1-sec: .912; 2-sec: .974; 5-sec: .994). This was also true for attractiveness and trustworthiness ratings (attractiveness: 1-sec: .714, 2-sec: .885, 5-sec: .920; trustworthiness: 1-sec: .557, 2-sec: .682, 5-sec: .861). Masculinity ratings did not increase across the time course, but were generally consistent (1-sec: .947; 2-sec: .933; 5-sec: .951). See Figure 3.4a.

For female voices, r^2 for F_0 also increased through the time course (1-sec: .762; 2-sec: .889, 5-sec: .986). This was also true for all ratings judgements (masculinity: 1-sec: .850, 2-sec: .910, 5-sec: .935; attractiveness: 1-sec: .819, 2-sec: .931, 5-sec: .962; trustworthiness: 1-sec: .643, 2-sec: .677, 5-sec: .717). See Figure 3.4b.

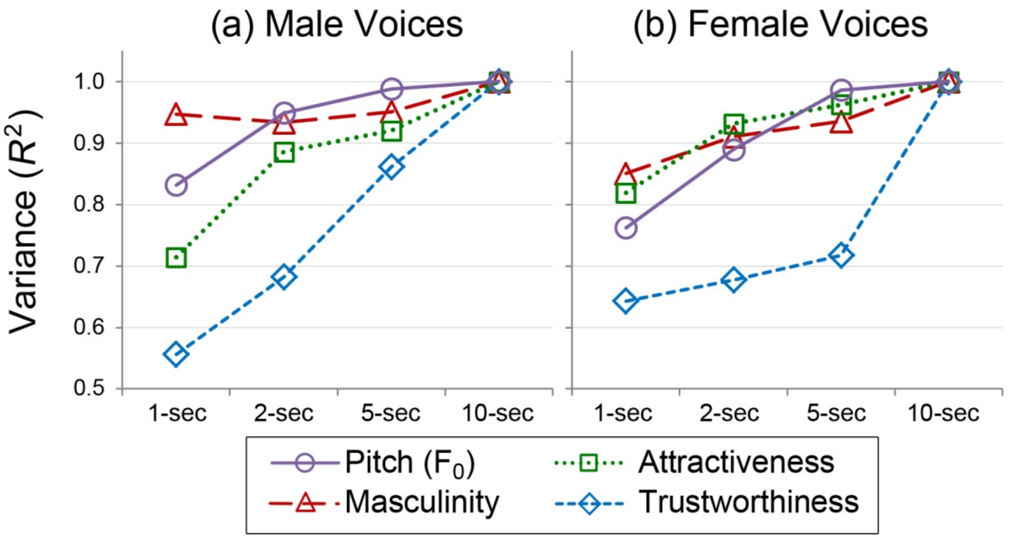

Figure 3.4
Variance explained (Study 2). Percentage of variance explained in full-recording ratings by pitch and ratings in shortened stimuli.

3.3.2.6 Relationship Between F_0 and Perceived Traits

Again the relationships between measured voice F_0 and the perceptions of all three rated traits were examined. Low male voice F_0 was significantly associated with masculinity judgements in all four time steps (see Table 3.4). Similarly, low female voice F_0 was associated with masculinity ratings, but this reached significance in only the 5- and 10-second trials. There were similar, significant negative relationships between male voice F_0 and attractiveness ratings in all time steps. Female voice F_0 was not significantly related to attractiveness ratings. Low voice F_0 also was significantly associated with trustworthiness ratings of male voices, but only in the 1- and 2-second trials; female voice F_0 was not significantly related to trustworthiness judgements. See Table 3.4.

Table 3.6 *Correlations between* F_0 *measurements and perceived traits (Study 2).*

	Masculinity		Attractiveness		Trustworthiness	
	Male	Female	Male	Female	Male	Female
1-sec	-0.511*	-0.16	-0.63**	0.09	-0.47*	0.16
2-sec	-0.48*	-0.31	-0.57*	0.25	-0.48*	0.12
5-sec	-0.47*	-0.46*	-0.61**	0.30	-0.36	0.15
10-sec	-0.57*	-0.47*	-0.56*	0.32	-0.39	0.12

Note: one-tailed *p*-values are reported for masculinity and attractiveness ratings. Trustworthiness ratings are reported using two-tailed p-values.
* $p < .05$, ** $p < .01$

3.4 General Discussion

The overall pattern of results suggests that the perception of vocal traits is subject to change according to a time course of attribution. Significant correlations were found across the time course for F_0 and for all trait perceptions, in addition to significant differences across time, indicating that the differences we observed over time changed in similar degree and directionality across voices. Of the three traits examined, masculinity judgements were the most stable over time, with a low degree of change across the time course; responses made at very short exposure times seem to reflect responses made using longer exposure lengths. Attractiveness judgements appear to change more over time, and trustworthiness judgements required the longest shortened exposure times to elicit responses that were similar to those made with unrestricted exposure time. Because stimuli of various types and lengths are used in voice perception studies, these results may inform the way in which these traits are examined in future research.

Both Study 1 and Study 2 demonstrated that voice pitch changes over time, with voice F_0 decreasing across all time steps in both studies. This is concordant with the source-filter theory of speech production, which states that a within-utterance decrease in voice pitch is, in part, a function of subglottal air pressure (Fant, 1960; Lieberman, 1967). In both studies, the change in F_0 was the most exaggerated in earlier time steps, suggesting that the largest changes in F_0 occur near the beginning of an utterance, and that F_0 stabilises in longer time frames. Because voice F_0 has been shown to influence listener attributions of traits like masculinity, attractiveness and trustworthiness (*e.g.* Collins, 2000; Feinberg, Jones, Little, *et al.*, 2005; Klofstad *et al.*, 2012), differences in F_0 over time could impact listeners' judgements of these qualities based on the length and/or type of stimulus used, particularly if the stimuli over-represent the beginning of an utterance (and the corresponding atypical F_0) by using very short sentences or singular spoken words (*e.g.* (Apicella *et al.*, 2007; Apicella & Feinberg, 2009; Lander, 2008; Puts, Apicella, *et al.*, 2012). Stimuli which use vowel sound extractions (*e.g.* Bruckert *et al.*, 2006; Collins, 2000; Feinberg, Jones, Little, *et al.*, 2005), or those which consist of longer scripted utterances (*e.g.* Puts *et al.*, 2011, 2006) are less likely to be affected by this change over time.

3.4.1 Masculinity

Significant effects of exposure length on ratings of masculinity were found in both Study 1 and Study 2, for male voices only. In Study 1, male voices were rated as more masculine after 5 seconds of exposure than they were after 1 second of exposure. Similar results were found in Study 2, with male voices being rated as more masculine after 2 seconds of exposure than after 1 second of exposure. Masculinity ratings did not change with additional exposure time in either study, indicating that male voice masculinity may be assessed with a minimum exposure length which is greater than 1 second, but less than or equal to 2 seconds of exposure – after this, perceptions of masculinity don't seem to change with longer exposure times.

Changes in masculinity ratings over the time course appeared to mirror the decrease in voice pitch between the 1- and 5-second measurements in Study 1, and similarly between the 1- and 2-second measurements in Study 2. However, when the degree of change in F_0 was examined this was not significantly associated with the degree of change in masculinity ratings in either study, *i.e.* if F_0 dropped to a greater degree, these voices were not necessarily perceived as more masculine. This may indicate that the variation in masculinity ratings with respect to time may not be due to the changes in stimuli, but rather to the cognitive process of the listener. Again, this suggests that masculinity judgements become reflective of judgements made at longer exposure lengths after 1 second and less than or equal to 2 seconds of exposure time.

Results from both studies also indicate that masculinity ratings were the least variable over time of the three traits examined, with variance explained by shortened exposure times at approximately 90% in all conditions of both studies. This lends further support to the idea that masculinity judgements can be made after very short exposure times, and do not change much with longer exposure. It may benefit listeners to assess masculinity quickly, considering that more masculine men may provide a greater threat potential both to men, due to intra-sexual competition (Archer, 1991; Mazur & Booth, 1998; Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2007; Wolff & Puts, 2010) and to women, due to the risk of violence and sexual coercion (Archer, Birring and Wu, 1998; Archer, 1991; Studer, Aylwin and Reddon, 2005). These results are similar to time-course judgements of aggressiveness and threat potential which use facial stimuli, which are likewise assessed very rapidly (Bar, Neta and Linz, 2006; Willis & Todorov, 2006).

While female voice F_0 changed in the same way as male F_0, female voices were rated equally masculine across all exposure lengths in both Study 1 and Study 2. Raters appeared to judge vocal masculinity in women after an exposure length of 1 second in both Study 1 and Study 2. It is possible that listeners do not generally attend to variation in masculinity in women’s voices in the same way that they do for men. This may be reflective of the established finding that, due to evolutionary pressures that arise from sex differences in reproductive costs, variance in sexual dimorphism is more important for men than for women (Trivers & Willard, 1973; Trivers, 1972). Also, if masculinity is a cue for threat potential in male voices, it could be that female voices simply do not sound as threatening as male voices in general. Since listeners are able to accurately assess speaker sex by the voice alone, it could be that the perception of gender might underlie the unchanging masculinity judgements in female voices (Bennett & Montero-Diaz, 1982; Childers & Wu, 1991). There may also be an issue of semantics – if "femininity" had been questioned for female voices rather than masculinity, there may have been more variation in the listeners' responses to female voices compared to male voices.

3.4.2 Attractiveness

Attractiveness ratings also changed over time, particularly for female voices. Both Study 1 and Study 2 suggest that female voices are rated as more attractive as exposure length increases. Jones *et al.*, (2008) found that the length of an utterance did not influence perceived attractiveness of female voices, however the difference in length between stimuli used (one word difference) was considerably shorther than in the present study. In male voices, attractiveness ratings decreased with exposure time in Study 1, however, Study 2 showed no significant differences in attractiveness across the time course. In Study 2, listeners were asked to rate voice stimuli for one trait only, and this added attention to a single trait may be reflected in more controlled ratings judgements.

The results also indicate that attractiveness judgements took slightly longer exposure times to reflect ratings made without time constraints. 5-second exposure lengths elicited ratings which explained over 50% of variance in unconstrained exposure ratings in Study 1, and over 90% of variance in 10-second exposures was attained after 2 seconds of exposure in Study 2. This suggests that extremely short stimuli (of less than two seconds) may not be long enough to elicit reliable attractiveness responses. The increased variance explained in

Study 2 may also be due to the added attention to voice attractiveness in Study 2 based on rating only a single trait compared to Study 1.

F_0 was negatively correlated with attractiveness ratings in men. This is in agreement with a large amount of voice preference literature which has shown that, among male voices, low F_0 is attractive (Collins, 2000; Feinberg, Jones, Little, *et al.*, 2005; Feinberg *et al.*, 2006; Little *et al.*, 2011). However, there was no significant relationship between attractiveness and F_0 in female voices in either study, in any of the presentation times. Current literature on the topic demonstrates generalised preferences for high F_0 for female voices (Collins & Missing, 2003; Feinberg, DeBruine, Jones and Perrett, 2008; Fraccaro *et al.*, 2010; Jones *et al.*, 2008). However, similar to the results of the present study, Smith *et al.* (2012) found no relationship between F_0 and attractiveness in female voice stimuli, suggesting that the proposed positive link between F_0 and attractiveness in female voices may not be as robust as previously thought.

3.4.3 Trustworthiness

Trustworthiness ratings in Study 1 increased between 1 and 5 seconds of exposure length, however, Study 2 showed no significant differences in mean trustworthiness ratings across the time course. Again, this different pattern of results in Study 2 compared to Study 1 may be reflective of the focus on a singular trait judgement in Study 2. Of the three traits we examined, trustworthiness ratings made after shortened exposures were the least reflective of those made after the longest exposure times in both studies. Similarity of ratings consistently increased with exposure time, but did not reach a point of stasis in either study. This supports the idea that complex personality traits such as trustworthiness may not be accurately assessed reflexively. This particular trait may require reflection on part of the listener, and longer exposure times may be necessary in order to capture consistent assessments. Researchers interested in social attributions like trustworthiness may wish to use stimuli of at least 5- to 10 seconds in duration, in order to capture more reliable responses.

F_0 did not significantly correlate with trustworthiness ratings of male voices in any of the three conditions in Study 1; however, there was a significant negative relationship in the 1- and 2- second conditions in Study 2. Similar findings have been reported by both Klofstad *et al.* (2012) and Tigue *et al.* (2012) using voice stimuli which had been raised and lowered in pitch. This may be explained as an influence of vocal attractiveness, since lower-pitched male voices were also found attractive and masculine in the current data. There were no significant relationships between F_0 and trustworthiness ratings among female voices. This result contrasts with Klofstad *et al.* (2012), who found that lower-pitched female voices were more trustworthy than higher-pitched female voices. However, Klofstad *et al*'s results were obtained by asking questions relating to voting and leadership, and may thus have a different conceptual meaning (*e.g.* dominance) from the general trustworthiness prompt (in this case, "rate the voice for trustworthiness once the clip is finished").

3.5 Conclusions

In summary, ratings of the three traits tended to correlate across the different time frames in both Study 1 and Study 2. Masculinity ratings after one second of presentation time predicted ratings given when participants heard longer extracts of the same speech, indicating rapid perception of this trait in voices. Attractiveness appeared to take longer to assess, and stimuli under two seconds in length may not elicit ratings which are reflective of those made at longer exposure lengths. Trustworthiness appeared to be the most deliberative attribute of the three we tested, with ratings of shortened stimuli showing low similarity to those made using longer stimuli. This may be of particular interest and value to researchers examining the perceptions of these traits, particularly attractiveness and trustworthiness judgements, which appear to be influenced by exposure length. Future research may consider this time course when developing stimuli used for trait perception. While physical traits seem to be quickly and reflexively assessed by listeners, researchers interested in more subjective vocal attributions, such as attractiveness, trustworthiness, or other social traits, may wish to use stimuli at least five seconds in duration.

CHAPTER 4: Vocal Fundamental and Formant Frequencies Affect Perceptions of Speaker Cooperativeness

In recent years, the perception of social traits in faces and voices has received much attention. Facial and vocal masculinity is linked to perceptions of trustworthiness, however, while feminine faces are generally considered to be trustworthy, vocal trustworthiness is associated with masculinized vocal features. Vocal traits such as pitch and formants have previously been associated with perceived social traits such as trustworthiness and dominance, but the link between these measurements and perceptions of cooperativeness have yet to be examined. In Study 1, cooperativeness ratings of male and female voices were examined against four vocal measurements: fundamental frequency (F_0), pitch variation (F_0-*SD*), formant dispersion (D_f) and formant position (P_f). Feminine pitch traits (F_0 and F_0-*SD*) were associated with higher cooperativeness ratings in both male and female voice stimuli. Additionally, voices with masculine formant traits (D_f and P_f) were considered more cooperative than voices with feminine formant traits. In Study 2, manipulated voices with feminized F_0 were found more cooperative than voices with masculinized F_0 among both male and female speakers, confirming our results from Study 1. Feminine pitch qualities may indicate an individual who is friendly and non-threatening, while masculine formant qualities may reflect an individual that is socially dominant or prestigious, and the perception of these associated traits may influence the perceived cooperativeness of the speakers.

4.1 Introduction

Previous research has shown that a variety of personality attributions are made based on facial appearance, including trustworthiness, competence, aggressiveness and dominance (Little, Roberts, Jones and Debruine, 2012; Oosterhof & Todorov, 2008; Todorov et al., 2009; Todorov, Said, et al., 2008). Further, these attributions are linked to morphological aspects of facial appearance. Pro-social traits tend to be associated with faces that are feminine and babyish, while negative and anti-social traits are associated with masculine, mature faces (Montepare & Zebrowitz, 1998; Oosterhof & Todorov, 2008; Perrett *et al*., 1998).

Similarly, voices also elicit personality attributions. Voice pitch (F_0) influences the perception of personality traits such as truthfulness, persuasiveness, nervousness and friendliness (Apple et al., 1979; Kramer, 1977). More recently, voice pitch has been associated with perceptions of trustworthiness (Klofstad et al., 2012; Tigue et al., 2012) and sexual infidelity (O'Connor, Re and Feinberg, 2011). Voice pitch is sexually dimorphic in humans, and lower pitch in men is commonly associated with masculinity and attractiveness (Feinberg, 2008). As with masculine facial traits, masculine vocal traits are also associated with negative traits, such as physical dominance and threat potential (Puts, Apicella, et al., 2012; Wolff & Puts, 2010). However, whether feminine voices or masculine voices are associated with pro-social personality traits remains somewhat unclear. While feminine faces are generally found to be more trustworthy than masculine faces, a number of studies have shown that masculine voices are more trustworthy than feminine voices (Apple et al., 1979; Klofstad et al., 2012; Oosterhof & Todorov, 2008; Tigue et al., 2012; Todorov, Baron, et al., 2008). This seemingly contradictory pattern of results suggests that further examination of the factors influencing perceptions of vocal prosociality is warranted.

4.1.1 Trustworthiness

Voice pitch has been associated with perceptions of trustworthiness. Voices with low F_0 are considered more truthful and trustworthy than voices with high F_0, in both male and female voices (Apple *et al*., 1979; Klofstad *et al*., 2012; Tigue *et al*., 2012). Masculine (i.e. low-pitch) voices are considered attractive in men, and feminine (i.e. high-pitch) voices are considered attractive in women (Feinberg, 2008; Feinberg, Jones, DeBruine, *et al*., 2005). Attractiveness is often associated with positive personality attributions via a "halo" effect (Eagly et al., 1991; Zuckerman et al., 1995). The halo effect may explain why low-pitched male voices are considered trustworthy (Klofstad *et al*., 2012; Tigue *et al*., 2012), however, masculinised pitch also makes men seem likely to engage in sexual infidelity (O'Connor *et al*., 2011), which is not in line with a straightforward halo effect. These seemingly paradoxical findings suggest that perceptions of prosociality may have a more complex link with vocal masculinity. Additionally, Tigue *et al.* (2012) found that lower-pitched voices were considered more trustworthy than their higher-pitched counterparts in voices of both sexes, suggesting that trustworthiness may not be exclusively related to vocal attractiveness, at least among female speakers. Rather, there may be a generalised effect of vocal masculinity being considered trustworthy in voices of both men and women. Additionally, Klofstad *et al*. (2012), Tigue *et al*. (2012) and O'Connor *et al.* (2011) each used manipulated versions of stimuli (raised and lowered F_0). Because listeners chose between very masculine and very feminine male voices, as opposed to measuring impressions based on normal variation in a naturalistic sample, this may have led to choosing masculine voices because the feminised voices sounded too high-pitched by direct comparison to masculinised voices. Furthermore, these studies did not examine acoustic traits other than F_0.

F_0 variation is another vocal attribute which may affect the perception of prosociality. In contrast to jitter or F_0 tremor, which are perceived as voice roughness, F_0 variation is captured by measuring the standard deviation in voice pitch throughout an utterance. As such, the pitch variation (F_0-SD) captures the amount of within-utterance variation in pitch, and low values of F_0-SD are perceived as monotony. A high variation in F_0 (F_0-*SD*) is considered a pleasant vocal attribute (Apple *et al*., 1979; Scherer, 1974) and its presence in both play behaviour in non-human primates and in human child-directed speech suggests that it may be used as a signal of affiliation (Goedeking, 1988; Trainor et al., 2000). Variation in F_0 may then also be related to perceptions of prosociality. Formant measures (formant dispersion, D_f, and formant position, P_f) may also influence listeners' attributions of prosociality, due to their relationships with dominance and intrasexual competition (Puts, Apicella, et al., 2012; Puts et al., 2007).

4.1.2 Dominance

The link between low voice pitch and trustworthiness is a surprising one, due to the association between masculinity, anti-social behaviour and dominance (Mazur & Booth, 1998). Masculine-sounding male voices are considered to be cues to dominance which could aid intra-sexual competition (Puts, Apicella, et al., 2012; Puts et al., 2007; Wolff & Puts, 2010), and low F_0 is associated with dominance both cross-culturally in humans and within non-human species (Morton, 1977; Ohala, 1983, 1984). Thus, if a speaker wishes to sound submissive, they may wish to affect higher-pitched vocalisations, with the goal of sounding small and nonthreatening (Ohala, 1984). Low F_0, D_f and P_f are related to body size, and it has been suggested that these traits serve as cues to threat potential (Feinberg, Jones, Little, et al., 2005; Puts, Apicella, et al., 2012; Puts et al., 2007). Speakers with naturally higher measurements of these vocal traits may be perceived as submissive, which could give the impression to listeners as being naturally more prosocial and cooperative.

A high F_0 may be related to increased perceptions of submissiveness because nervousness (such as that brought about by lying or fear) has an impact on vocal fold tension. An autonomic nervous response via vagus nerve stimulation tightens the vocal folds, which increases F_0 (Charous et al., 2001). F_0 variation may also be influenced by emotional arousal, and may reveal emotional traits of the speaker, such as whether they feel confident or threatened (Banse & Scherer, 1996; Hodges-Simeon et al., 2011), and a low F_0 variation has been suggested as a means of intimidation (Hodges-Simeon et al., 2010). Because low F_0 and F_0-SD are related to threat potential and intimidation, and because high measures of these traits may be related to nervousness and fear, individuals with naturally higher pitch and pitch variation may accordingly be perceived as submissive, which in turn could positively influence perceptions of prosociality. Low measures of these traits may negatively influence ratings of cooperativeness, as dominant individuals may use threat or physical strength to get their way, while cooperation requires working in tandem to a common, mutually-beneficial end. Thus, voices that sound masculine and dominant may be considered attractive, or even trustworthy, but an inverse relationship between masculinity and cooperativeness may be expected because masculine individuals may behave in a more selfish way or be less likely to acquiesce to the needs of others (Booth & Osgood, 1993; Dabbs & Morris, 1990).

A listener’s own dominance may additionally influence the way they attribute prosociality to others. Watkins, Jones and DeBruine's (2010) finding that dominant men are less sensitive to facial dominance cues in other men lends support to the idea that social trait attribution may be modulated in part by the individual differences of the listeners. This is also supported by research showing that taller (i.e. more dominant) men are less sensitive to dominance cues in masculinised faces and voices than shorter men (Watkins, Fraccaro, *et al*., 2010), however this study also found that height was not associated with self-rated dominance, nor was self-rated dominance associated with dominance attributions in faces and voices. A possible explanation for the differing results presented in the two aforementioned studies may be in the way dominance was measured. Watkins, Jones, *et al*. (2010) measured dominance as a personality trait using an 11-item questionnaire (Goldberg, 1999) while Watkins, Fraccaro, *et al*. (2010) utilised a single scaled question about the participants' dominance, which may be more reflective of the participants' conceptions of their own physical dominance rather than capturing dominant personality characteristics. Research by Wolff and Puts (2010) did not find that self-rated physical dominance, physical aggressiveness, or morphometric measures of strength predicted dominance attributions of others, however the measures taken by these researchers focus on traits which reflect physical formidability rather than dominant personality traits such as those measured by Watkins, Jones, *et al*. (2010). It may thus be reasonable to suspect that individual differences in dominance as a personality characteristic may interact with the way social traits are perceived in the others.

4.1.3 The Present Research

In the present study, ratings of cooperativeness were examined for male and female voices based on a naturalistic sample (Study 1). Here, measurements of pitch (F_0) and pitch variation (F_0-*SD*), as well as two measures of formants (formant dispersion, D_f, and formant position, P_f) were investigated. In Study 2, we examined the effect of manipulated F_0 on ratings of cooperativeness. In both Study 1 and Study 2, the dominance of the subjects who rated the stimuli was also measured, in order to determine if this factor affected how cooperative they found the voices of others.

4.2 Study 1

4.2.1 Methods

4.2.1.1 Stimuli

16 men and 16 women were recruited as stimulus donors (male ages 18-30, mean age 20.4 years, *SD* 2.73 years; female ages 18-23, mean age 19.4 years, *SD* 1.46 years). All were undergraduate psychology students at the University of Stirling. Recordings were obtained using an Audio-Technica AT-4041 microphone with a cardioid pickup pattern, at a distance of approximately 65cm using a preamp (M-Audio Audiobuddy). Audio was recorded directly to hard disk as .wma files using Windows Movie Maker v.2.1.4027.0, with a 48kHz sampling rate and 16-bit quantisation. The room was quiet and partially soundproofed with 1.5-inch thick sound-dampening foam. Participants were recorded while reading a scripted text. This text was selected due to its neutrality of content (see Cowan & Little, 2013). For the purposes of this experiment, 5 seconds of speech was extracted from this scripted recording: “October frequently brings the first frost of the season over the greater part of the UK.” Extraction was completed using Audacity (v.2.0.2). Participants whose first language was not English, and those who exhibited difficulties reading from a script (*e.g.* omitting words, stuttering, long pauses, or repeating words), were excluded. Additionally, participants over the age of 30 were excluded from the stimulus set so that perceived age would not play a role in participants' ratings (Linville & Fisher, 1985; Mulac & Giles, 1996).

For analysis and playback, audio files were converted to single-channel .mp3 at 320kbps/48kHz using Switch v.2.04. All voice measurements were obtained using Praat v.5.3.03 (Boersma & Weenink, 2013). F_0 was measured using Praat’s autocorrelation algorithm. Pitch was searched for between 65-300Hz for male voices, and between 100-600Hz for female voices, according to the manufacturer’s recommendations (Boersma &

Weenink, 2013). Measurements of the first four formants were taken ($F_1 – F_4$) using Linear Predictive Coding with the BURG algorithm, using 10 poles and pre-emphasis. Maximum frequencies were set at 5500Hz for female voices and 5000Hz for male voices, again per manufacturer recommendations. These formant measures were used to calculate both formant dispersion (D_f, see Fitch, 1997), which is the average distance between the four formants in Hz, and formant position (P_f, see Puts, Apicella and Cárdenas, 2012), which is obtained by assigning each formant a *z*-score and taking the mean of these four standardised measures. F_0-*SD* (the within-utterance standard deviation of F_0) was also recorded. All measurements were obtained using the entire utterance as the measurement window length. The use of scripted speech should normalise the unvoiced segments across individual speakers, and formant measures taken in this manner may more accurately reflect the perceptions of the listener. These four measurements were chosen in order to capture vocal masculinity, due to the sexual dimorphism exhibited by each of these measures. Additionally, these four measures have all been related to attributions of social traits in previous studies.

4.2.1.2 Subjects

Participants (*N* = 79) were psychology undergraduates at the University of Stirling. Females (*n* = 57) were aged 18-35 (*M* = 19.56 years, *SD* 2.8 years); males (*n* = 22) were aged 18-30 (*M* = 19.95 years, *SD* 2.5 years). All took part in the study to fulfill a course requirement. All phases of this experiment were approved by the University of Stirling Ethics Committee.

4.2.1.3 Procedure

Following Havlicek, Roberts and Flegr (2005) and Watkins, Jones and DeBruine (2010), participants completed the 11-item dominance subscale of the IPIP (Goldberg, 1999, *ipip.ori.org*) resulting in a range of scores from 18 to 43, with a mean score of 31.9 (*SD* 5.2). This questionnaire was administered prior to stimuli exposure. Male and female voices were presented in separate blocks, and randomised within each block. Participants were asked to rate the voices for how cooperative they thought the person sounded. For the purposes of this study, cooperativeness was defined as "a measure of how likely you think a person might be to work with you toward a mutually beneficial goal - *e.g.* writing a presentation or contributing to group work. In these situations, cooperative people will do their fair share of the work required. A person who is uncooperative is not likely to contribute their fair share of work or resources, but will still enjoy the rewards of effort provided by others." This definition stresses mutual-benefit cooperation and highlights the possible existence of defectors/free-riders. Voices were rated on a 7-point Likert scale, with 1 indicating low cooperativeness, and 7 indicating high cooperativeness.

4.2.1.4 Data Treatment

For ANOVA analyses, the 16 voice stimuli for each sex were placed into high- or low- F_0 groups (separated evenly into two quantiles of the 8 highest- and 8 lowest- F_0 in the sample). This median split was performed in order to maximise statistical power due to the low number of voices sampled (N = 16 for each sex), and to make the results more comparable to experiments which use manipulated stimuli. This same method was used to create high and low quantiles based on the other traits measured (F_0-*SD*, D_f, P_f). Voices which fell into one high group did not necessarily fall into high groups of other measurements, *e.g.* voices with high F_0 were not entirely the same as voices with high D_f, etc. While this method of collapsing the data into two separate groups is not without its disadvantages, similar methods have been usefully applied by previous researchers to examine differences between groups based on high and low measures of other traits (*e.g.* Cowan & Little, 2013; Little, Burt, Penton-Voak and Perrett, 2001; Penton-Voak & Chen, 2004; Penton-Voak *et al.*, 2003; Stanton, Liening and Schultheiss, 2011). Linear mixed effects models were performed in R (R Core Team, 2014) with the *lmerTest* package (Kuznetsova, Brockhoff and Christensen, 2014).

4.2.2 Results

For tests using the stimulus as the unit of analysis, mean cooperativeness ratings were calculated by averaging the ratings of all listeners. Mean ratings were also calculated separately for male and female listeners in order to determine whether male and female listeners use different cues for perceiving cooperativeness. For tests using the listener as the unit of analysis, each participant's mean cooperativeness rating given to all stimuli and separate mean ratings of male and female voices were calculated. Additionally, mean ratings were calculated (for each participant) for high- F_0 and low- F_0 voices separately; the same method was used to calculate ratings based on high/low F_0-*SD*, D_f, and P_f.

The mean acoustic measurements of all parameters are within the averages of those examined in previous research, barring mean female voice F_0, which is lower than the population-level average. While the minimum and mean F_0 are lower, the upper limit is on par with those measured by other researchers (*e.g.* Feinberg, DeBruine, Jones and Perrett, 2008; Puts *et al.*, 2012), which indicates that a wider range of female F_0 has been captured, and not an unrepresentative sample. Based on the types of analyses used, there is no reason to suspect that these lower-than-average female voices should elicit a pattern of results that would differ in directionality from a more restricted stimulus set, as a wider range of F_0 has been measured than that which is typically utilised.

Table 4.1 *Sexual dimorphism of male and female voice measurements in Study 1. Means, standard deviations, and t-values (independent-samples t-tests) are reported.*

	Male Voices		Female Voices			95% CI	
Measure	Mean	*SD*	Mean	*SD*	*t*(30)	LL	UL
F_0	110.3Hz	17.1Hz	185.5Hz	42.7Hz	6.55***	99.3	51.3
F_0-*SD*	11.6Hz	5.0Hz	38.6Hz	19.0Hz	5.49***	37.4	16.6
Df	1030Hz	43Hz	1151Hz	31Hz	9.03***	148	93
Pf	-0.66	0.52	0.66	0.60	6.60***	1.7	0.9

Note. Degrees of freedom for t = 30 in all cases barring F_0 (df = 19.713) and F_0-*SD* (df = 17.051). These cases did not pass Levene's test for equality of variance, and thus, we report corrected confidence intervals.
*** $p < .001$

4.2.2.1 Vocal Measurements and Cooperativeness Ratings

All four vocal measurements obtained (F_0, F_0-*SD*, D_f, P_f) were sexually dimorphic, with all measures significantly lower for male voices than for female voices (independent-samples t-tests; all $t > 6.55$, all $p < .001$). See Table 4.1. Of all measurements obtained, F_0 was the only metric which was directly correlated to cooperativeness ratings (Table 4.2). The effect was present only in male voices, when rated by other men, $r(16) = .526$, $p = .04$. This positive relationship was also present when male voices were rated by women, however, this was not significant, $r(16) = .279$, $p = .30$. Female voice F_0 was not related to cooperativeness when rated by men, $r(16) = -.01$, $p = .96$, or by women, $r(16) = -.08$, $p = .76$

Table 4.2 *Correlated vocal measurements and cooperativeness ratings in Study 1.*

	Male Voices		Female Voices	
	Rated by Men	Rated by Women	Rated by Men	Rated by Women
Measure	*r*	*r*	*r*	*r*
F_0	0.53*	0.28	-0.01	-0.08
F_0-*SD*	0.21	-0.18	-0.26	-0.20
Df	0.19	0.08	0.18	0.31
Pf	-0.15	-0.01	-0.36	0.23

* $p < .05$

A 2x2 ANOVA (sex of voice; high/low F_0; sex of listener as a between-subjects factor) revealed a significant main effect of F_0 only, $F(1, 77) = 14.81$, $p < .001$, with high-pitched voices being found more cooperative than low-pitched voices, in both male and female stimuli, when rated by both male and female listeners. See Table 4.3. No other significant effects or interactions were observed (all $F < 0.73$, all $p > .40$). See Figure 4.1a.

For the measure F_0-*SD*, the 2x2 ANOVA (sex of voice; high/low F_0-*SD*; sex of listener as a between-subjects factor), a significant main effect of F_0-*SD* was revealed, $F(1, 77) = 26.94$, $p < .001$, indicating that a high F_0-*SD* was found more cooperative than low F_0-*SD* in both male and female voices. There was also a significant interaction between F_0-*SD* and sex of speaker, $F(1, 77) = 10.33$, $p = .002$, and a marginally-significant three-way interaction between F_0-*SD*, sex of voice, and sex of listener, $F(1, 77) = 3.93$, $p = .051$. This appears to be driven by men listening to female voices, who were not significantly influenced by high- or low- F_0-*SD*, $t(21) = 0.61$, $p = .55$, 95% CI [-0.30, 0.16]. However, men did rate other men as more cooperative if they had a high F_0-*SD*, $t(21) = 3.40$, p = .003, 95% CI [0.23, 0.93]. Female listeners also gave significantly higher cooperativeness ratings to both male and female voices with high F_0-*SD* (male voices: $t(56) = 3.98$, $p < .001$, 95% CI [0.17, 0.51]; female voices: $t(56) = 2.32$, $p = .024$ CI [0.03, 0.34]).

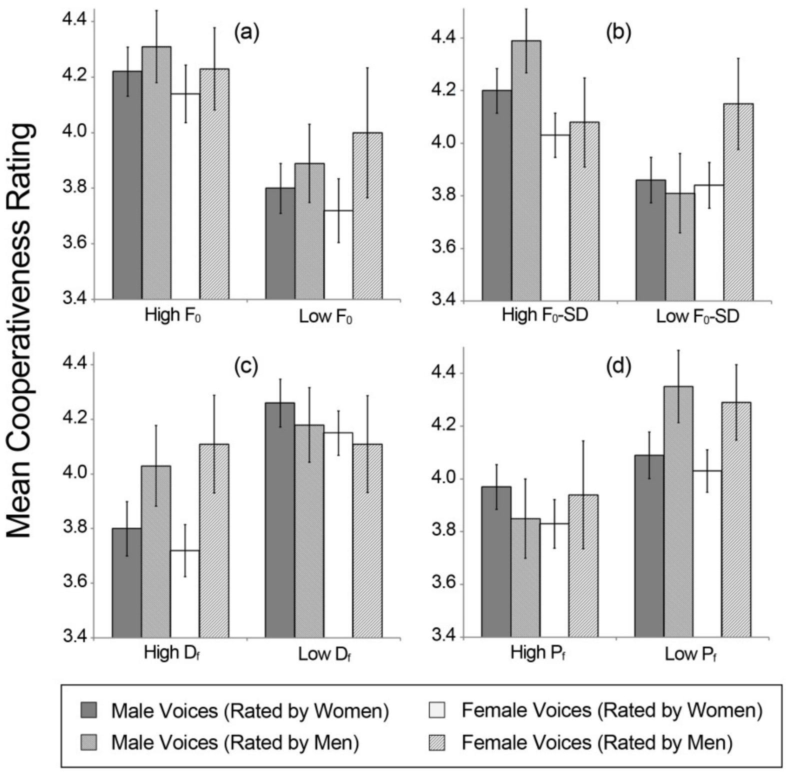

Figure 4.1
Cooperativeness ratings (Study 1) by high/low F_0 *(a),* F_0*-SD (b),* D_f *(c, and* P_f *(d). Error bars represent the standard error of the mean.*

Similarly, a 2x2 ANOVA for D_f (sex of voice; high/low D_f; sex of listener as a between-subjects factor) revealed a significant main effect of D_f, $F(1, 77) = 13.55$, $p < .001$, with low D_f being found more cooperative in voices of both men and women. There was also a significant interaction between D_f and sex of listener, $F(1, 77) = 6.98$, $p = .01$, indicating that female listeners were influenced by D_f to a greater degree than were male listeners. Post-hoc contrasts revealed that only female listeners gave higher cooperativeness ratings to voices with low D_f (female listeners/male voices: $t(56) = 4.11$, $p < .001$, 95% CI [-0.69, -0.24]; female listeners/female voices: $t(56) = 4.79$, $p < .001$, 95% CI [-0.62, -0.25]; male listeners/male voices: $t(21) = 0.00$, $p = 1.0$, 95% CI [-0.31, 0.31]; male listeners/female voices: $t(21) = 0.79$, $p = .44$, 95% CI [-0.54, 0.24]).

The 2x2 ANOVA for P_f (sex of voice; high/low P_f; sex of listener as a between-subjects factor) returned a significant main effect of P_f, $F(1, 77) = 19.99$, $p < .001$, with low

P_f found more cooperative in voices of both sexes. A marginally-significant interaction between P_f and sex of listener was observed, $F(1, 77) = 3.83$, $p = .054$. This appears to be driven by female ratings of male voices, which were not rated significantly differently based on high- or low- P_f, $t(56) = -1.47$, $p = .15$, 95% CI [-0.30, 0.04], while ratings of female voices were, $t(56) = -2.51$, $p = .015$, 95% CI [-0.37, -0.04]. Both male and female voice cooperativeness ratings were similarly affected by P_f amongst male listeners (male voices: $t(21) = 2.62$, $p = .016$, 95% CI [-0.90, -.10]; female voices: $t(21) = 2.54$, $p = .019$, 95% CI [-0.64, -0.06]).

Table 4.3 *Categorical measurements and mean ratings (Study 1). Voices were categorized by high or low* F_0, F_0-*SD*, D_f *and* P_f *for ANOVA analyses.*

				Female Listeners		Male Listeners	
Category	Range	Mean	*SD*	Mean Coop. Rating	*SD*	Mean Coop. Rating	*SD*
High F_0 (Male Voices)	112.3 - 139.2 Hz	123.9	9.1	4.22	0.66	4.31	0.61
Low F_0 (Male Voices)	80.3 - 111.6 Hz	96.6	10.8	3.80	0.68	3.89	0.66
High F_0 (Female Voices)	193.3 - 290.2 Hz	215.2	33.3	4.14	0.78	4.23	0.69
Low F_0 (Female Voices)	117.1 - 190.2 Hz	155.9	28.0	3.72	0.87	4.00	1.10
High F_0-*SD* (Male Voices)	10.9 - 21.0 Hz	15.4	3.9	4.20	0.64	4.39	0.57
Low F_0-*SD* (Male Voices)	4.8 - 10.6 Hz	7.7	2.0	3.86	0.65	3.81	0.71
High F_0-*SD* (Female Voices)	36.5 - 84.6 Hz	52.3	16.5	4.03	0.63	4.08	0.79
Low F_0-*SD* (Female Voices)	6.6 - 34.2 Hz	25.0	8.8	3.84	0.66	4.15	0.81
High D_f(Male Voices)	1036 – 1100 Hz	1065	26.0	3.80	0.75	4.03	0.69
Low D_f(Male Voices)	964 – 1033 Hz	996	25.2	4.26	0.66	4.18	0.64
High D_f(Female Voices)	1154 – 1219 Hz	1174	21.5	3.72	0.72	4.11	0.84
Low D_f(Female Voices)	1079 – 1146 Hz	1128	21.6	4.15	0.61	4.11	0.83
High P_f (Male Voices)	-0.63 - 0.77	-0.29	0.48	3.97	0.64	3.85	0.70
Low P_f (Male Voices)	-1.27 - -0.72	-1.03	0.22	4.09	0.66	4.35	0.64
High P_f (Female Voices)	0.53 - 1.97	1.08	0.58	3.83	0.69	3.94	0.96
Low P_f (Female Voices)	-0.03 - 0.52	0.24	0.20	4.03	0.60	4.29	0.67

4.2.2.2 Mixed Effects Models

Generalised linear mixed effects regression (GLMER) models were also performed to determine what vocal traits influenced ratings of cooperativeness. Random effects in all of the models were the listener and voice stimulus. For male listeners of male stimuli, the model was significant, $\chi^2(4) = 15.74$, $p = .003$. There was a near-significant fixed effect of F_0, $F = 3.52$, $p = .06$, and significant fixed effects of F_0-*SD*, $F = 6.01$, $p = .03$, and listener dominance, $F = 5.50$, $p = .02$. There was also a significant interaction between F_0 and dominance, $F = 4.27$, $p = .04$. The model shows that men with higher voice pitch and high pitch variation were rated by other men as more cooperative, and that male listeners who were low in dominance gave higher cooperativeness ratings than high-dominance listeners. The interaction between F_0 and dominance suggests that high-dominance men found high F_0 to be cooperative, while low-dominance men favoured voices which were low in F_0. See Figure 4.2.

The model for female listeners was also significant, $\chi^2(2) = 11.66$, $p = .003$. Women rated male stimuli as cooperative based on two fixed factors. F_0 was positively associated with cooperativeness, $F = 8.47$, $p = .004$. There was also a significant effect of women's age on cooperativeness ratings, such that older women gave higher cooperativeness ratings overall, $F = 7.48$, $p = .006$. There was also a significant interaction between listener age and F_0, $F = 5.07$, $p = .02$. The interaction between women's age and F_0 indicates that older women found low-pitched voices cooperative, while younger women found high-pitched voices cooperative.

There was also a significant model for male listeners of female stimuli, $\chi^2(3) = 16.12$, $p = .001$. There was a significant fixed effect of F_0 on ratings of cooperativeness, $F = 9.10$, $p = .003$, such that higher pitch was associated with higher cooperativeness ratings. There was also a significant effect of listener dominance, $F = 15.69$, $p < .001$, suggesting that high-dominance men found women to be less cooperative than low-dominance men. As with men listening to male voices, a similar significant interaction between dominance and F_0 was found, $F = 10.20$, $p = .002$. See Figure 4.2. This interaction suggests that high-dominance men found high-pitched female voices to be more cooperative, and low-dominance men favoured lower-pitched female voices.

For women listening to the voices of other women, a model was found which approached significance, $\chi^2(1) = 3.27$, $p = .07$, indicating that a low D_f was found more cooperative than high D_f, $F = 3.20$, $p = .10$. While this was not a significant predictor, the model does indicate that a directionality is present. In these four models presented, the addition of further factors, including vocal measurements and listener age and dominance, and interactions between these, did not significantly improve the models beyond the results presented.

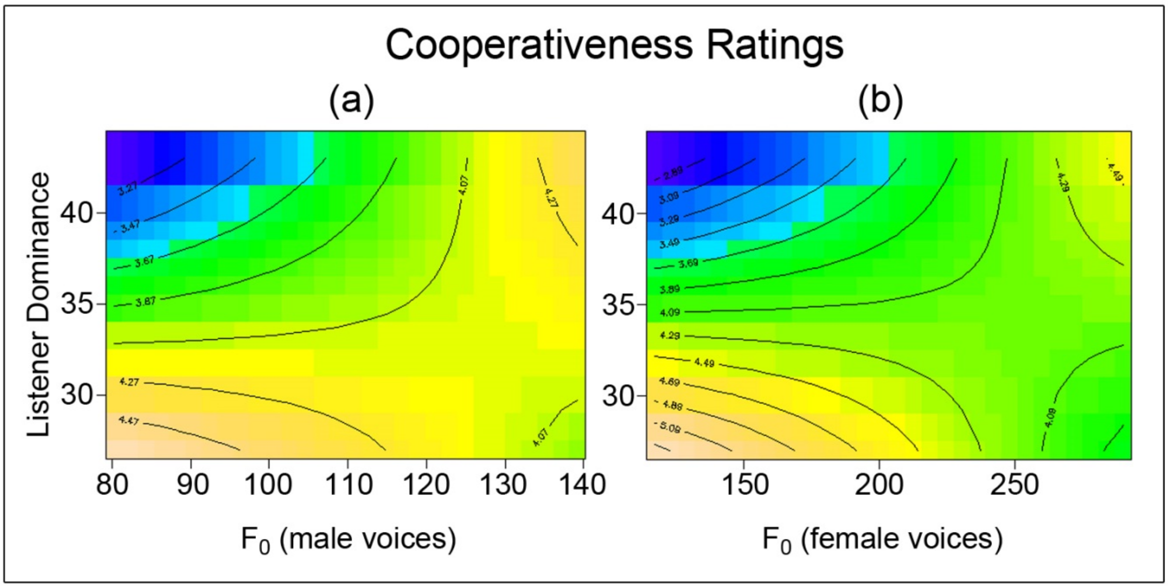

Figure 4.2
3D plot of cooperativeness ratings as a function of the interaction between F_0 (x-axis) and the dominance scores of male listeners (y-axis).A heat map colour scheme is shown, with blue areas reflecting low cooperativeness ratings and orange areas reflecting high cooperativeness ratings. Ratings of male voices (a) and female voices (b) are shown separately.

4.2.2.3 Individual Differences

Among female listeners, age was positively correlated with cooperativeness ratings of male voices, such that older listeners gave higher cooperativeness scores to men, *r*(57) = .327, *p* = .013, but not to women, *r*(57) = .185, *p* = .17. Age was unrelated to cooperativeness ratings among male listeners when rating men, *r*(22) = .271, *p* = .22 and when rating women, *r*(22) = .240, *p* = .28.

Among male listeners, scores on the dominance questionnaire were negatively correlated with mean cooperativeness ratings, such that low dominance was related to higher ratings of cooperativeness in others, *r*(22) = -0.47, *p* = .026. This was mainly true for voices of women, *r*(22) = -.506, *p* = .016, though a similar directionality was present for male voices as well, *r*(22) = -.346, *p* = .115. Dominance scores for female listeners were not related to mean cooperativeness ratings in voices of either sex (male *r*(57) = .097, *p* = .47; female *r*(22) = .085, *p* = .53). These relationships are shown in Figure 4.3.

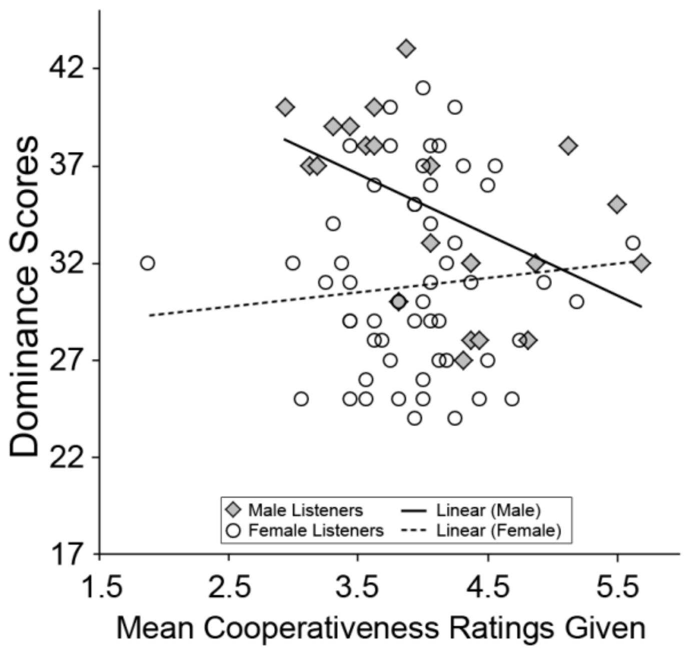

Figure 4.3
Correlations between mean cooperativeness ratings given and listeners' dominance scores.

4.3 Study 2

The results from Study 1 indicated that of all the measures examined, F_0 was most strongly linked to cooperativeness judgments, with statistically significant results with relatively low degrees of freedom (using the vocal stimulus as the unit of analysis, $N = 16$). While other vocal measurements did influence cooperativeness judgments with more degrees of freedom (using the rater as the unit of analysis, $N = 79$), F_0 was the only vocal measurement which was related to cooperativeness measurements in both methods of analysis as well as in the mixed effects models. Additionally, F_0 has been manipulated in numerous experiments by other researchers examining subjective traits such as attractiveness, dominance, and trustworthiness (*e.g.* Feinberg, Jones, Little, Burt and Perrett, 2005; Puts *et al*., 2007; Tigue *et al*., 2012). Thus, the relationship between F_0 and cooperation was further examined by repeating Study 1 using a stimulus set consisting of voices with manipulated F_0.

4.3.1 Methods

4.3.1.1 Stimuli

8 male and 8 female voices were randomly selected from the stimuli used in Study 1 (male ages 19-30 years, $M = 21.0$ years, SD 3.74 years; female ages 18-23 years, $M = 19.6$ years, SD 1.92 years). Pitch manipulations were made using Praat v.5.3.56 (Boersma & Weenink, 2013). Using Praat's pitch-synchronous overlap add (PSOLA) method, each voice was manipulated in Hz by +/- 0.5 equivalent rectangular bandwidths (ERBs), which is perceptually equivalent to a manipulation of +/- 20Hz (Traunmüller, 1990). This created a raised and lowered version of each voice, resulting in a total of 16 male and 16 female voices. The PSOLA method alters the pitch of the voice, while leaving other aspects (*e.g.*

formants) unchanged. Numerous other experiments have successfully used the PSOLA method in experiments examining perceived attractiveness, dominance and trustworthiness (*e.g.* Feinberg *et al*., 2006; Jones, Feinberg, Debruine, Little and Vukovic, 2008; Jones, Feinberg, DeBruine, Little and Vukovic, 2010; Klofstad *et al*., 2012; Puts, 2005; Tigue *et al*., 2012; Vukovic *et al*., 2008), allowing this experiment to be directly comparable to a large amount of previously published literature. Amplitude was scaled to create a constant presentation volume using RMS (root-mean-squared) method.

4.3.1.2 Subjects

Participants (*N* = 101) were psychology undergraduates at the University of Stirling. Females (*n* = 70) were aged 16-40 years (*M* = 20.3 years, *SD* 4.66 years); males (*n* = 31) were aged 17-49 years (*M* = 20.7 years, *SD* 5.81 years). All took part in the study to fulfil a course requirement.

4.3.2.3 Procedure

Apart from the stimuli, the procedure for Study 2 was identical to Study 1. Stimuli were again presented in separate blocks of male/female voices. Each block consisted of 8 voices, which had been both raised and lowered in F_0, resulting in 16 voice stimuli per block. Within each block, the order of presentation was randomised.

4.3.2 Results

As in Study 1, each participant's mean cooperativeness rating given to all stimuli was calculated, as well as the separate mean ratings of male and female voices. Additionally, each participant's mean rating of high- F_0 and low- F_0 voices was calculated for male and female voices separately.

Listeners rated both male and female voices which had been raised in pitch as significantly more cooperative than voices which had been lowered (male voices: raised M = 4.26, SD = 0.64, lowered M = 3.97, SD = 0.70, $t(100)$ = 4.71, p < .001, 95% CI [0.17, 0.42]; female voices: raised M = 4.22, SD = 0.71, lowered M = 3.96, SD = 0.70, $t(100)$ = 4.36, p < .001, 95% CI [0.14, 0.37]). See Figure 4.4. A 2x2 ANOVA (pitch, sex of voice, sex of rater as a between-subjects factor) revealed a significant main effect of F_0 on cooperativeness ratings, $F(1,99)$ = 38.07, p < .001. No other significant main effects or interactions were observed (all other $F \leq 2.66$, all other $p \geq .11$).

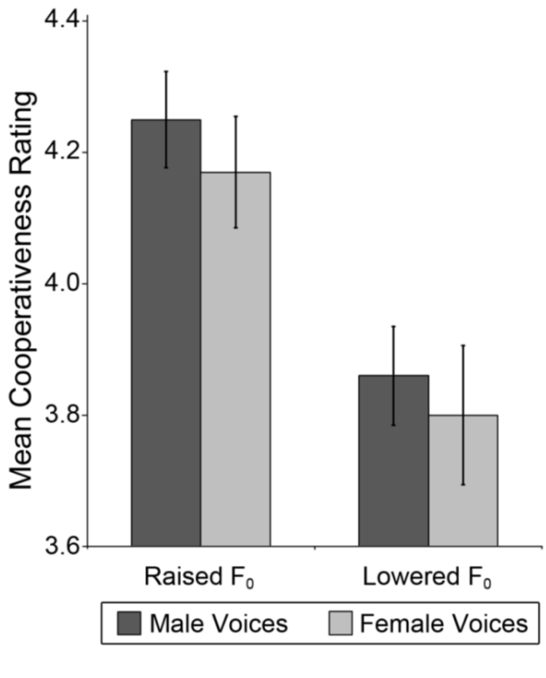

Figure 4.4
Cooperativeness ratings by pitch condition (Study 2). Error bars represent the standard error of the mean.

Age was positively correlated with cooperativeness ratings for female raters, $r(70) = .26$, $p = .028$. When further examined by sex of speaker, the correlation remained for male voices, $r(70) = .27$, $p = .023$, but not for female voices, $r(70) = .20$, $p = .104$, though the directionality of the effect is the same. There was no significant correlation found between age and cooperativeness ratings by male listeners, $r(31) = -.18$, $p = .33$. No significant correlations between dominance scores and cooperativeness ratings were found among male listeners (rating women: $r(31) = -.08$, $p = .66$; rating men: $r(31) = .03$, $p = .86$).

Using ANCOVA, it was investigated whether self-measures of dominance were related to listeners' sensitivity to masculinity cues when judging the pro-sociality of others (within-subjects factor: mean cooperativeness rating [masculinised, feminised]; between-subjects factor: sex of listener; covariates: age, dominance score). No significant main effect of listener dominance was found for men listening to voices of other men, $F(1, 67) = 0.17$, $p = .68$), suggesting that listeners' own dominance did not affect men's sensitivity to dominance cues of other men while assessing cooperativeness. There was a non-significant interaction between listener age and masculinity cues, $F(1, 67) = 3.46$, $p = .074$, which suggests that older men may have been more sensitive to the pitch manipulation than younger men, and rated masculinised voices as less cooperative than younger men. There was no effect of age or dominance on men's sensitivity to the pitch manipulation in female voices, all $F < 1.30$, all $p > .26$. We also found no significant effects of age or dominance on women's sensitivity to manipulated voice pitch when assessing the cooperativeness of other women, all $F < 2.69$, all $p > .11$, or of other men, all $F < 0.92$, all $p > .34$.

While there was no relationship between dominance scores and cooperativeness judgments among female listeners in Study 1, a significant negative correlation was found in Study 2, irrespective of F_0 manipulation. Here, dominance was negatively correlated with cooperativeness ratings of women rating female voices, $r(70) = -.26$, $p = .03$, but not when women rated male voices, $r(70) = -.08$, $p = .52$.

4.4 General Discussion

Results from Study 1 indicated that feminine pitch traits (high F_0 and high F_0-*SD*) and masculine formant traits (low D_f and P_f) were considered more cooperative-sounding in male voices than those with masculine pitch traits and feminine formant traits. In female voices, feminine pitch traits (high F_0 and high F_0-*SD*) were also considered more cooperative than masculine pitch traits. A masculine P_f was also considered more cooperative than a feminine P_f. The mixed models suggest that a masculine D_f may be associated with judgments of female voice cooperativeness. Study 2 confirmed the findings from Study 1 regarding a positive association between F_0 and cooperativeness ratings, with feminised voice pitch being found more cooperative in the voices of both men and women. Individual differences of the listeners also influenced cooperativeness ratings. Among female listeners, age was positively correlated with cooperativeness ratings given to other women in both Study 1 and Study 2. In Study 1 only, dominance in male listeners was negatively correlated with cooperativeness ratings given to women, and dominance seemed to interact with the way men respond to F_0 as a cue to prosociality.

4.4.1 What makes a voice sound cooperative?

Both Study 1 and Study 2 illustrate that the vocal factor which influences cooperativeness to the greatest extent is a high F_0. Male listeners also displayed a tendency to rate men's voices with a high F_0-*SD* as more cooperative than voices with a low F_0-*SD*. Additionally, voices with low formant measures were rated as more cooperative than voices with high formant measures (both D_f and P_f for men, and P_f for women); however, the mixed model analyses confirmed this relationship between a low D_f and increased ratings of cooperativeness for women listening to other women's voices only. It is important to note that while D_f is a measure of the spacing between the formants, P_f is a measurement of the mean frequency of the formants. So, while cooperative voices had little space between the formants (a low D_f), those formants also have a low mean measured value (a low P_f), according to our ANOVA results.

Feminine pitch and pitch variation, combined with masculine formants, appear to be vocal traits that influence perceptions of cooperativeness. The results regarding the relationship between a high F_0 and ratings of cooperativeness were the most robust, and the mixed model analyses also suggest that a high F_0-*SD* and a low D_f may also positively influence ratings of cooperativeness. While the results concerning P_f were not conclusive based on the mixed effects models presented in Study 1, the relationships which were uncovered using the median-split technique are intriguing, and worth examining further. Future research may wish to examine the relative importance of pitch and formant traits more thoroughly.

4.4.1.1 Pitch and Pitch Variation

High F_0 positively influenced perceptions of a speaker's cooperativeness in voices of both sexes, in both a naturalistic (Study 1) and manipulated sample (Study 2). In both studies, a higher voice pitch elicited increased ratings of cooperativeness in male and female voices. Low F_0 is associated with masculinity (Feinberg, DeBruine, Jones and Little, 2008; Feinberg, 2008; Pisanski, Mishra and Rendall, 2012; Pisanski & Rendall, 2011; Puts, Apicella, et al., 2012) and is related to high testosterone levels among men (Dabbs & Mallinger, 1999). Given that men with high testosterone are prone to numerous antisocial and risk-taking behaviours (Apicella et al., 2008; Archer et al., 1998; Booth & Osgood, 1993; Coates & Herbert, 2008; Dabbs & Morris, 1990; Mazur & Booth, 1998; D. B. O'Connor, Archer and Wu, 2004; Rowe, Maughan, Worthman, Costello and Angold, 2004; Stanton et al., 2011; Studer et al., 2005) , the perception of masculinity may suggest a general air of uncooperativeness or lack of prosociality. This finding underscores the importance of F_0 as a male intra-sexual signal, with low F_0 indicating physical dominance (Hodges-Simeon *et al*., 2010; Vukovic *et al*., 2011; Wolff & Puts, 2010). Given that physically dominant persons may use their physical strength to get what they want, and less dominant individuals may be less likely to physically challenge others, persons with lower perceived masculinity and dominance may be considered desirable as potential cooperators. Additionally, vocal femininity may be associated with a certain degree of compliance, which may also be captured by the construct of "cooperativeness."

Voices with high F_0-*SD* (i.e. more dynamic, less monotone voices) were rated as more cooperative than voices with low F_0-*SD*. Hodges-Simeon *et al*. (2010) found that low F_0-*SD* predicted higher ratings of physical dominance, which may be an undesirable trait in potential cooperators. F_0-*SD* is sexually-dimorphic, with high F_0-*SD* being a feminine trait. This trait in men, then, may sound friendlier and less dominant than monotone voices. Given that a high F_0-*SD* was also associated with cooperativeness in female voices (in ANOVA results only) lends further credence to F_0-*SD*'s inverse relationship with perceived dominance in voices of both sexes. These results support the suggestion that variation in F_0 may serve to elicit and maintain positive emotional states in the listener (Traunmüller & Eriksson, 1995). Hodges-Simeon *et al*. (2010) liken F_0 variation to a "smile," positing high F_0-*SD* as a submissive social gesture. These data support such an association.

4.4.1.2 Formants

Low D_f is generally considered a masculine trait (Feinberg, Jones, Little, *et al*., 2005; Wolff & Puts, 2010) and is associated with a larger body size (Feinberg, Jones, Little, *et al*., 2005) due to the allometric relationship between body height and vocal tract length (Evans, Neave and Wakelin, 2006; Fitch, 1997). While a masculine trait may not be overtly expected to be associated with cooperativeness, it has previously been suggested that formants have a greater effect on dominance judgments than on judgments of masculinity

(Hodges-Simeon *et al*., 2010; Pisanski & Rendall, 2011), suggesting a link between the formants and perceptions of social traits. When listening to both male and female voices, women exhibited a stronger response to differences in D_f than men did, rating men and women with low D_f as significantly more cooperative than voices with high D_f., although the mixed model analyses only confirmed a relationship between low D_f and cooperativeness for women listening to the voices of other women. If the formants are a somewhat reliable indicator of speaker size or height, women may find taller individuals to be more cooperative, possibly because height could confer prestige. While voices with low D_f are considered dominant, Puts *et al*. (2007) suggest that this could also be related to social dominance or prestige. This generalised association between low D_f and cooperativeness may be linked to the social benefits of height. Taller individuals have greater social status (Cavelaars *et al*., 2000; Power, Manor and Li, 2002), higher levels of education (Power *et al*., 2002; Silventoinen, Lahelma and Rahkonen, 1999) and greater earnings than shorter individuals (Judge & Cable, 2004; Loh, 1993), which may make them more favourable as potential cooperators. It could be that female listeners are more attuned to cues of social status, while male listeners are more attuned to cues of physical dominance. Indeed, while female listeners seem to be influenced by D_f (and men were not), male listeners seem to be influenced by P_f (and women were to a lesser degree). Although this was not confirmed in the mixed model results, this relationship is an intriguing one. P_f is associated with masculinity and dominance (Puts, Apicella, et al., 2012), and was sexually dimorphic in the present data. It is also negatively related to height, weight, arm strength, and physical aggression, suggesting that this trait may be generally indicative of threat potential (Puts, Apicella, et al., 2012). Why a trait thus linked to threat potential would be positively associated with cooperativeness is unclear, though it may be that male listeners associate physical threat potential with social dominance, while female listeners may assess the trait in purely social terms.

Further research may give more attention to the inter-relationships between dominance, prestige, and pro-sociality. The median split used in Study 1 to divide the stimuli into groups based on high and low measures of vocal traits does unfortunately contribute to some loss in variation amongst these traits, and is dependent upon the voices in the particular stimulus set. The nature of the data required that differing median splits be made for each of the four measured vocal traits, such that different vocal stimuli fell on either side of this split dependent upon the trait under analysis. This calls attention to the variability and complexity of vocal characteristics, and serves as testament to their respective importance when examining perceptions based on these traits. Additionally, here formants were measured over the entire utterance and, while scripted speech should help normalise the proportion of unvoiced segments, future studies could focus only on the voiced parts of recordings.

4.4.2 Individual Differences

While no effects of listener age or self-measures of dominance were found to be associated with sensitivity to dominant vocal cues in Study 2, generalised effects of age and dominance were observed on mean ratings of cooperativeness given across both studies. Women's age was positively correlated with cooperativeness ratings of male voices in both Study 1 and Study 2, and the mixed models in Study 1 confirm that older women tended to give higher cooperativeness scores to men's voices than younger women. Women's age was associated with their sensitivity to F_0 as a cue to cooperativeness, such that older women found low-pitched voices cooperative, while younger women tended to favour higher-pitched male voices. Age can be judged by listening to vocal stimuli with reasonable accuracy (Mulac & Giles, 1996; Ramig, Scherer and Titze, 1984), and older women generally prefer older men as potential mates (Buss & Schmitt, 1993; Buss, 1989; Mathes, Brennan, Haugen and Rice, 1985). With age, women may become more confident and socially able, and thus, may find young men less intimidating (the mean ages of stimuli in Study 1 and Study 2 were 20.4 years and 21.0 years, respectively, while the age range of female listeners was 18-35 years and 16-40 years respectively). Women's ratings of young female voices did not produce the same pattern of results, which seem to be unaffected by the age of the listener. There was no effect of age amongst male listeners in either Study 1 or Study 2.

Self-rated dominance appeared to play a role in how male listeners attribute vocal cooperativeness. Low-dominance men gave higher mean cooperativeness ratings to male and female voices than high-dominance men in Study 1. The dominance of the listener also interacted with men's sensitivity to F_0 as a cue to cooperativeness. Low-dominance men found low-pitched male and female voices more cooperative, and high-dominance men tended to favour high-pitched male and female voices. It may be that dominant men feel more socially favourable toward voices which exhibit signs of a small physical stature and low masculinity, as these voices may not pose any threats to the listener's perceived dominance status.

While this result was not replicated in Study 2, a similar effect of dominance was found on women's cooperativeness ratings of female voices (which was not found in Study 1). Like men rating other men in Study 1, high dominance in women was associated with lower overall cooperativeness ratings given to the voices of other women. Watkins, Jones and DeBruine (2010) demonstrated that dominance (measured as a personality characteristic) influences the way men perceived dominance in other men, and it may also affect the way listeners perceive other social traits, such as cooperativeness. Dominant personality traits may "interfere" with the way both men and women perceive cooperativeness, such that those who are dominant may tend to view same-sex individuals as uncooperative, and favour low masculinity in a cooperative partner, possibly due to an enhanced sense of intra-sexual competition among high-dominance individuals. Additionally, low-dominance individuals may have a generally more positive view of others, as they are less likely to engage in aggressive and antisocial behaviour (Ehrenkranz, Bliss and Sheard, 1974; Rowe *et al*., 2004).

4.5 Conclusions

While it was expected that vocal qualities which indicated smaller, shorter, feminine individuals would be found more cooperative than those indicating larger, taller, more masculine individuals, the results here were mixed. Feminine F_0 and F_0-*SD* positively influenced cooperativeness ratings, with higher F_0 being most robustly associated with higher perceived cooperativeness. F_0's link to perceptions of speaker masculinity is reflected here – high pitched and dynamic voices sounded more cooperative than low-pitched and monotone voices. Formants also appeared to play a role in cooperativeness judgments, with more masculine formant measures (D_f and P_f) being found more cooperative than feminine formants in both male and female voices. These seemingly dichotomous results further illustrate the relative importance of F_0 and the formants on person perception and the perception of social traits.

Self-rated dominance had some effect on how male and female listeners perceived vocal cooperativeness. High-dominance men seemed to be more sensitive to F_0, as they found high-pitched voices of male and female speakers as cooperative, while low-dominance men found low-pitched voices to be more cooperative. However, the results regarding the negative link between dominance and the attribution of prosociality in others which were found in Study 1 were not replicated in Study 2. Older women tended to rate low-pitched male voices as cooperative, while younger women favoured higher-pitched male voices, perhaps because masculine voices may sound intimidating to younger women. Further experiments may usefully examine the individual differences of listeners and how these affect their perceptions of social traits.

In summary, these results demonstrated that different vocal traits can work in synchrony to create complex interpersonal judgments. While it was found that pitch alone was a consistent factor influencing listeners' cooperativeness ratings of both male and female voices, pitch variation and formants also play an important role in the perception of cooperativeness.

CHAPTER 5: Effects of Vocal Characteristics on Memory for Voice Identities and Verbal Content

This chapter will examine what makes a voice memorable. Recently, studies have shown that pairing images of objects with attractive faces and voices enhances memory for those objects. However, while viewers preferentially attend to attractive faces, they do not preferentially remember those faces in recognition tests. In a series of four experiments, I examine whether manipulations of voice pitch influences listeners' recall of voices which are heard in exposure phases, and whether the content spoken by attractive voices is preferentially recalled using a crossmodal visual recognition test. Experimental manipulation of voice pitch did not influence memory for voice identities, however attractive male voices were better remembered than unattractive voices, independent of pitch manipulation. Performance on content recognition tests was enhanced by raising the voice pitch of both male and female speakers during the exposure phase. We link this pattern of results to perceptions of vocal attractiveness and dominance, and how these perceptions might impact how memories are formed and retained.

5.1 Introduction

Physical attractiveness has been associated with enhanced attention and memory because observers preferentially attend to attractive individuals (Maner *et al.*, 2003). Voice pitch (fundamental frequency, F_0) influences perceptions of attractiveness for both male and female speakers (for a review, see Feinberg, 2008). Low male voice pitch (fundamental frequency, or F_0) is attractive, both in naturalistic and manipulated stimuli (Collins, 2000; Feinberg, Jones, Little, *et al.*, 2005). Following the sexually dimorphic nature of the trait, high voice pitch is found attractive in female voices (Anderson *et al.*, 2010; Collins &

Missing, 2003; Feinberg, DeBruine, Jones and Perrett, 2008). But are attractive voices more likely to be remembered than unattractive voices? In studies using face stimuli, attractive female faces are more likely to be remembered, but while women selectively attend to attractive male faces, they do not remember them later (Becker, Kenrick, Guerin and Maner, 2005; Maner *et al.*, 2003). Recently, it has been found that women's memory for visual images was greater when the images were accompanied by masculinised (i.e. lowered), attractive male voices (Smith *et al.*, 2012). It has been suggested that this enhanced memory for items associated with attractive male voices found by Smith *et al.* (2012) is mate-choice relevant, due to women's preferential recall of content accompanied by masculinised opposite-sex voices, but not same-sex voices. This may be adaptively beneficial because women could evaluate the mate quality of men based on the objects they are associated with. In addition to attractiveness, the voice also influences a variety of other physical and social attributions about the speaker, including traits like dominance, threat potential, trustworthiness and honesty (Helfrich & Weidenbecher, 2011; Puts, Apicella, *et al.*, 2012; Tigue *et al.*, 2012; Vukovic *et al.*, 2011; Wolff & Puts, 2010), all factors which may unconsciously influence a listener's encoding of the voices and/or spoken content.

Similar attractiveness based biases in memory have also been found using facial images rather than voices. Images paired with masculinised male faces were more likely to be remembered by women, but only if the participants found masculine male faces to be more attractive than feminine male faces (Allan *et al.*, 2012). Again, women's preferential recall of objects associated with attractive male faces may be associated with women's evaluation of potential mates. This memory bias also may be due to added visual attention given to attractive male faces and the proximity of objects to those faces. Whatever the cause, the effect of masculinity preference on memory for objects associated with masculine and feminine faces found by Allan *et al.* (2012) may also apply to voice stimuli. If women are more likely to remember masculinised male voices, memory for manipulated voices may be associated with the degree to which women perceive the voice stimuli as being attractive.

Both Allan *et al.* (2012) and Smith *et al.* (2012) show how object recognition can be enhanced by attractive vocal and facial stimuli. Recognition of spoken content may also be affected by voice attractiveness. Helfrich and Weidenbecher (2011) found that short-term memory for spoken content in the absence of auditory or visual stimuli (a "retention" test consisting of multiple-choice questions about the content) was higher for lowered and unmanipulated male voices than for raised male voices (Helfrich & Weidenbecher, 2011). Further, these lowered and unmanipulated stimuli were found more attractive than voices with raised F_0, lending additional support to the idea that perceptions of voice attractiveness may, like facial attractiveness, be associated with enhanced listener memory.

Listeners are able to remember individual voices which they have previously heard, and are also able to recall discrete word-voice pairs – that is, recognising a specific voice speaking a specific word or phrase. The origins of this occurrence are in episodic memory (Craik & Kirsner, 1974; Goldinger, 1996; Papcun *et al.*, 1989; Schacter & Church, 1992), and remembering voice-word pairs may be similar to remembering a past event. It has also been demonstrated that listeners have implicit memory for F_0 and voice intonation, with a reduction in recall accuracy for words when voices are subsequently presented with manipulated F_0 and intonation compared to the initial exposure (Church & Schacter, 1994). Memory for spoken or text-presented words is strongest in the congruent modality, where the recognition test phase is in the same sensory modality as the exposure phase, and a shift in sensory modality (crossmodal memory testing), for example from spoken-to-textual or textual-to-spoken, reduces or eliminates priming effects on recall and recognition (Ellis, 1982; Gibson & Bahrey, 2005; Jackson & Morton, 1984). Further to this, it has also been shown that simply imagining the speaker's voice in the recognition phase enhances participants' crossmodal visual word recognition (Geiselman & Glenny, 1977). These results lend support to the idea that recall of discrete voice-word pairs is a part of episodic memory, and recall of spoken content may involve a different type of memory encoding. To our knowledge, possible effects of voice attractiveness on crossmodal memory have not yet been examined. One aim of this study is to examine whether listeners' perceptions of vocal attractiveness mediate the reduction in recognition accuracy across sensory modalities.

There are three reasons that one may expect to see some effect of voice attractiveness and/or pitch manipulation on crossmodal memory. Firstly, viewing attractive faces is associated with reward-circuitry activation – specifically, the orbitofrontal cortex (OFC) (Aharon *et al.*, 2001; O'Doherty *et al.*, 2003; Winston, O'Doherty, Kilner, Perrett and Dolan, 2007). This region is associated with the representation of reward from a variety of sources, including gustatory, olfactory and somatosensory stimuli, as well as abstract (*e.g.* monetary) reward (for a review, see O'Doherty, 2004). OFC activation is also present when listening to pleasant music, indicating that auditory stimuli can also activate reward circuitry (Blood, Zatorre, Bermudez and Evans, 1999; Blood & Zatorre, 2001). Activation of the OFC has been shown to positively influence memory for attractive faces via a direct neural connectivity to the medial temporal lobe (MTL), a region involved in memory encoding (Marzi & Viggiano, 2010; Tsukiura & Cabeza, 2011). Because attractive faces are rewarding, and their association with reward helps them to be better remembered, we may see similar effects of reward on memory for attractive voices.

Secondly, emotional voice intonation has been shown to affect memory. Kitayama (1996) found that messages delivered in emotional voices, both positively- and negatively-valenced, capture the attention of the listener. This attention is necessarily divided between the emotion and the verbal content. Attractive and/or masculine voices may inspire autonomic arousal on the part of the listeners, due to mate-choice relevant emotions and the perception of dominance and threat potential, which is associated with masculine male voices (Puts, Apicella, *et al.*, 2012). Attractive voices may thus distract from content memory by activating neural processes which are not related to memory, detracting cognitive resources away from the encoding process (Kitayama, 1996), such as mate evaluation and threat evaluation. Indeed, while a number of studies have shown that observers have attentional biases for angry or threatening face stimuli (Hansen & Hansen, 1988; Van Honk, Tuiten, de Haan, van den Hout and Stam, 2001; Öhman, Lundqvist and Esteves, 2001; Pinkham, Griffin, Baron, Sasson and Gur, 2010) which is adaptively beneficial for the reduction of danger and physical harm (Öhman & Mineka, 2001), they are not better remembered than happy or pleasant faces (Bradley, Greenwald, Petry and Lang, 1992; D'Argembeau, Comblain and Etienne, 2003; Foa, Golboa-Schechtman, Amir and Freshman, 2000). Furthermore, it has been shown that women selectively attend to attractive male faces, but do not preferentially remember them in recognition tests (Anderson *et al.*, 2010), lending additional support to the idea that attractiveness draws attention away from memory encoding.

Thirdly, attractiveness may enhance crossmodal memory due to purely mate-choice relevant motivations. Memory should, in part, be tuned to aid in the performance of actions that enhance our reproductive fitness (Nairne & Pandeirada, 2008). As suggested by Allan *et al.* (2012) and Smith *et al.* (2012), attractive stimuli may signal information that is important to remember for the evaluation of potential mates. While Allan *et al.* (2012) and Smith *et al.* (2012) tested recognition memory for objects associated with attractive voices and faces, we test for retention of content words spoken by voices of variable attractiveness. Because retention memory can confer survival advantages when the content recalled is adaptively beneficial, for example remembering survival-related words, sources of food, or indeed, remembering the appearance of potential mates and rivals (Nairne & Pandeirada, 2008; Nairne, Thompson and Pandeirada, 2007), we might expect that words spoken by attractive-sounding voices might receive greater memory prioritisation than those spoken by unattractive voices.

While it has been demonstrated that listener memory is affected by vocal attributes and sensory modality, it has not yet been examined whether listeners preferentially recall attractive and/or manipulated-F_0 voices, or whether voice attractiveness mediates the reduction in recall across modalities. The present study attempts to parse the effects of voice attractiveness on these two types of memory in a series of four experiments. Experiments 1 and 2 tested memory for voices with manipulated pitch using manipulated auditory stimuli in both the exposure and test phases (Experiment 1: male voices; Experiment 2: female voices). Because manipulated voice pitch affects perceived attractiveness in both male and female voices, and observers have been shown to preferentially attend to attractive faces, enhanced sexual dimorphism of voices (*i.e.* lowered pitch in male voices and raised pitch in female voices) may enhance listeners' attention to the auditory stimuli. This attention may (or may not, as in Becker et al., 2005) translate to memory. Experiments 3 and 4 tested crossmodal memory with manipulated auditory stimuli in the exposure phase and visual (written word) stimuli in the test phase (Experiment 3: male voices; Experiment 4: female voices). Again, because attractive faces receive preferential attention (but not memory), it is expected that manipulated/attractive voices may enhance recall of spoken content, perhaps to a greater degree than when paired with voices from the exposure trials. While previous research has found that changing modalities in recall trials negatively impacts accuracy (Ellis, 1982; Gibson & Bahrey, 2005; Jackson & Morton, 1984), this effect may be either ameliorated or enhanced by pitch manipulations and/or the attractiveness of the voices in the learning phase. I also examine whether masculinity/femininity preferences or rated voice attractiveness have an effect on listener performance in recognition trials.

5.2 Methods

5.2.1 Stimuli

Four men (mean age = 23.8 years, *SD* = 3.8 years) and four women (mean age = 20.0 years, *SD* 0.0 years) were recorded in a quiet, partially-soundproofed room using an Audio-Technica AT-4041 microphone with a cardioid pickup pattern, at a distance of between 25-35cm from the microphone. Sound was recorded direct to hard disk in stereo using Windows Movie Maker v. 2.1.4027.0, with a sampling rate of 48kHz and 16-bit quantisation; stimuli were converted from stereo to mono using Audacity v.2.0.3. Numerous previous experiments have used stimuli from 4-6 donors to measure responses to manipulated pitch (*e.g.* Feinberg *et al.*, 2006; Jones, Feinberg, DeBruine, Little and Vukovic, 2010; O'Connor *et al.*, 2012; Vukovic *et al.*, 2008). Participants read a list of words obtained from the Academic Word List (AWL) (Coxhead, 2000). Fifty-eight words from Sublist 1 (the full list) and six words from Sublist 2 were used to create a total set of sixty-four words. These words were chosen due to their level of complexity and commonality – they are not so common as to be easily forgotten, and are words that should be easily recognised and understood by university students.

Each voice donor read 16 words from the full list of 64 words. Of these 16 words, eight were raised and eight were lowered in pitch. Pitch manipulation was ±20Hz, achieved using the pitch-synchronous-overlap-add algorithm (PSOLA) using Praat v. 5.3.22 (Boersma & Weenink, 2012). The PSOLA algorithm isolates the pitch of a voice for manipulation, while leaving other aspects of the voice intact. Four raised and four lowered words from each voice donor were used in the exposure trials, to become target words in the recognition trials. The remaining four raised and four lowered words appeared in the recognition trials only as distractor stimuli, such that each speaker contributed four raised and four lowered target word stimuli, as well as four raised and four lowered distractor stimuli in the spoken word recognition trials. Identical manipulation and counterbalancing was used for male and female voice stimulus sets. In the word recognition trials, test stimuli were typographic images of the sixty-four words in a standard, non-stylised font.

5.2.2 Participants

Participants in the male spoken word trials (*N* = 93) were 47 men (mean age = 20.5, *SD* 2.4 years) and 46 women (mean age = 20.7, *SD* 4.7 years). Participants in the male textual word recognition trials (*N* = 46) were 22 men (mean age = 21.0, *SD* 4.2 years) and 24 women (mean age 20.38, *SD* 3.3 years). Participants in the female spoken word trials (*N* = 74) were 21 men (mean age = 26.7, *SD* 8.7 years) and 53 women (mean age = 26.4, *SD* 11.4 years). Participants in the female textual word recognition trials (*N* = 33) were 10 men (mean age = 19.2, *SD* 1.4 years) and 23 women (mean age = 21.6, *SD* 6.4 years). All participants were either students from the University of Stirling or members of the general public. Participants received either course credit or entry into a £25 prize draw as compensation. Ethics permission was approved by the University of Stirling Ethics Committee.

5.2.3 Procedure

To disguise the true nature of the experiment, study subjects were informed that they were participating in an experiment about voice perception and maths performance. Participants listened to thirty-two fully-randomised raised and lowered target stimulus words, and were asked to rate them for attractiveness on a 7-point Likert scale. In addition to gathering attractiveness data, this also assured that the participants attended to the voices without explicitly being told they were to be given a memory task. Each listener heard voices from one gender only. Following the attractiveness ratings, participants were given an “easy” skill-level Sudoku puzzle distractor task, which was timed for three minutes. After exactly three minutes, the subjects were moved to the testing/recall phase of the experiment, in which they either listened to the full set of sixty-four words (consisting of thirty-two target and thirty-two distractor stimuli), or were presented with textual/typographic images of the sixty-four words, and were asked whether they had heard each word in the previous phase of the experiment. “Yes” and “no” were the only options available. The recognition trials were also fully-randomised. Following this task, participants were debriefed as to the true nature of the experiment.

5.2.4 Data Treatment

Correct responses were scored with the number “1,” and all incorrect answers with the number “0.” It was thus possible to calculate accuracy scores as percentages of correct responses, and this was done in four categories of stimulus: target/raised, target/lowered, distractor/raised, and distractor/lowered.

5.2.5 Preference Data

In each experiment, each participant’s mean attractiveness rating was calculated separately for raised voice stimuli and lowered voice stimuli. The lowered mean was then subtracted from the raised mean, resulting in a scale in which negative values indicated a preference for lowered voices, and positive values indicated a preference for raised voices, with chance (no preference) being at 0. This variable was used as a covariate in ANOVA analyses.

Preferences for voice masculinity and femininity were examined using one-sample *t*-tests. In the male spoken word trials, women’s mean preference was for masculinised voices, but this was not significantly different from chance (female preference: M = -0.23, $t(45) = -1.57$, $p = .12$). Male listeners had no substantial degree of preference (male preference: M = -0.01, $t(46) = -0.13$, $p = .90$). Women’s mean preference in the textual word recognition trials indicated a preference for feminised voices, however this was also not statistically different from chance (*M* = 0.15, $t(31) = 1.24$, $p = .22$). Male preference in the male textual word recognition trials indicated that men found masculinised male voices more attractive than feminised voices, but this also did not differ significantly from chance (*M* = -0.42, $t(11) = -2.06$, $p = .06$).

Both male and female participants exhibited preferences for raised female voices in the spoken word trials at rates above chance (male preference: *M* = 0.39, $t(20) = 2.92$, $p = .008$; female preference: *M* = 0.32, $t(52) = 3.57$, $p < .001$). Textual word recognition trials revealed similar preferences for raised female voices, however these were not significant (male preference: *M* = 0.33, $t(9) = 1.63$, $p = .13$; female preference: *M* = 0.24, $t(22) = 1.79$, $p = .09$).

5.3 Results

Mean scores were first calculated using the stimulus as the unit of measurement. Attractiveness ratings and accuracy scores were averaged for each stimulus in order to directly compare results across sensory modalities (crossmodally). On average, participants performed significantly worse on textual word recognition trials than in the spoken word recognition trials for target stimuli (voice trial *M* = 70.3%, *SD* 10%; word trial *M* = 50.1%, *SD* 14%; $t(63) = 10.43$, $p < .001$) and for distractor stimuli (voice trial *M* = 70.6%, *SD* 12%; word trial *M* = 53.0%, *SD* 19%; $t(63) = 6.33$, $p < .001$). Stimuli were also divided within each separate experiment into groups of high- and low- attractiveness based on mean attractiveness ratings by all participants.

Mean accuracy scores were further calculated for each participant in each experiment, separately for raised and lowered voices, yielding mean accuracy scores separately for raised-F_0 target words and lowered-F_0 target words. In the spoken word recognition trials, mean accuracy was also obtained for raised-F_0 distractor words and lowered-F_0 distractor words. No such data was available from the textual word recognition trials because the test stimuli were presented with only visual images of the words which had been spoken in the learning phase.

5.3.1 Male Spoken Word Recognition Trials – Experiment 1

On average, women correctly recalled hearing 72.7% of lowered-F_0 target words and 74.7% of raised-F_0 target words. Both lowered- and raised-voice accuracy was significantly above chance (lowered: $t(45) = 8.65$, $p < .001$; raised $t(45) = 9.74$, $p < .001$). Male listeners exhibited similar accuracy levels, correctly responding to 70.6% of lowered-F_0 target words (above chance: $t(46) = 8.50$, $p < .001$) and 67.7% of raised-F_0 target words (above chance: $t(46) = 7.51$, $p < .001$).

Women were better than men at recognising target voices, but the difference was not significant (male *M* = 69.1%, *SD* 13%; female *M* = 73.7, *SD* 16%; $t(91) = 1.52$, $p = .13$). There was no sex difference in performance for distractor trials (male *M* = 73.5%, *SD* 17%; female *M* = 70.9%, *SD* 15%; $t(91) = -0.76$, $p = .45$). Women did not have a better memory for lowered-F_0 target words compared to men (male *M* = 70.6%, *SD* 17%; female *M* = 72.7%, *SD* 18%; $t(91) = 0.58$, $p = .56$), but did have a better memory for raised-F_0 target words compared to men (male *M* = 67.7%, *SD* 16%; female *M* = 74.7%, *SD* 17%; $t(91) = 2.04$, $p = .04$).

ANOVA analyses (within-subjects factors: accuracy by target F_0 manipulation (lowered, raised); between-subjects factor: sex of listener; covariate: masculinity preference) revealed no effect of manipulated pitch on target word accuracy ($F(1,90) = 0.06$, $p = .80$). There was also no interaction between pitch manipulation and listener sex ($F(1,90) = 1.74$, $p = .19$) and no interaction between pitch manipulation and preference on accuracy for masculinised vs. feminised target stimuli (F(1,90) = 0.01, p = .91). Accuracy for target stimuli across all trials is shown in Figure 5.1.

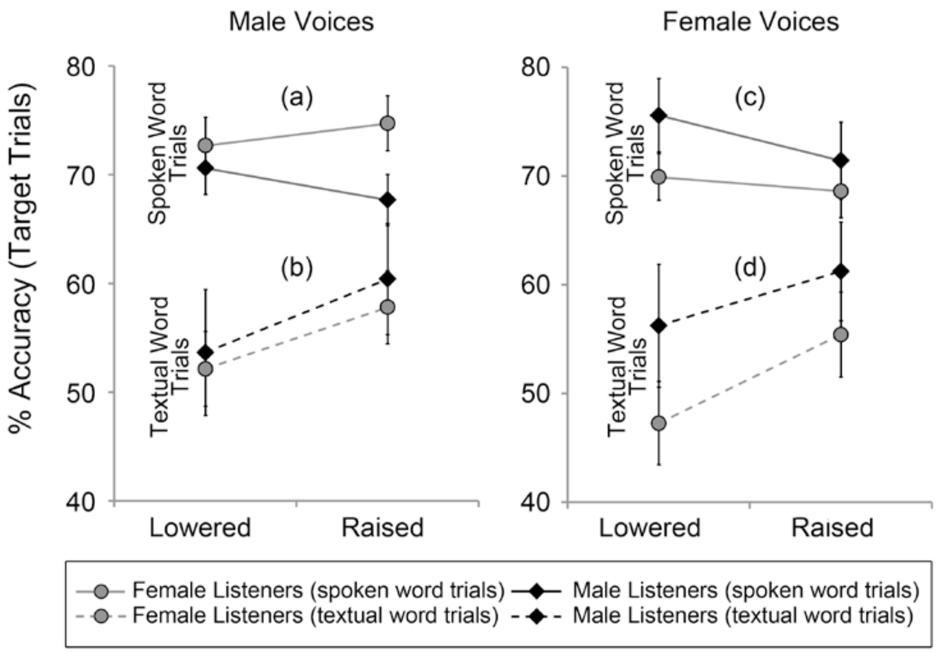

Figure 5.1
Listener memory by pitch manipulation in spoken word and visual word trials. Accuracy on manipulated-F_0 *target stimuli in the test phase for male spoken word trials* (*a*)*, male textual word trials* (*b*)*, female spoken word trials* (*c*) *and female textual word trials* (*d*)*. Error bars represent standard error of the mean.*

The extent to which women preferred masculinised male voices was not correlated with their recall of masculinised vs. feminised voice stimuli (r(46) = .001, p = .99). Male-rated differences in attractiveness between raised- and lowered-F_0 target stimuli was also unassociated with their degree of accuracy on test stimuli (r(47) = -.026, p = .86).

Pertaining to the accuracy of distractor-stimuli trials, ANOVA results for distractor trials (within-subjects factors: distractor accuracy by F_0 manipulation (lowered, raised); between-subjects factor: sex of listener; covariate: masculinity preference) revealed a main effect of manipulated pitch on accuracy (F(1,90) = 31.38, p < .001), indicating that listeners were more likely to misattribute memory to voices which had been lowered in pitch (masculinised distractor accuracy M = 67.7%, SD 17%; feminised accuracy M = 76.7%, SD 19%). There was no observed effect of listener sex (F(1,90) = 1.39, p = .24) or masculinity preference ($F(1,90) = 0.04$, $p = .84$). Correct responses for distractor stimuli across both spoken word trials are shown in Figure 5.2.

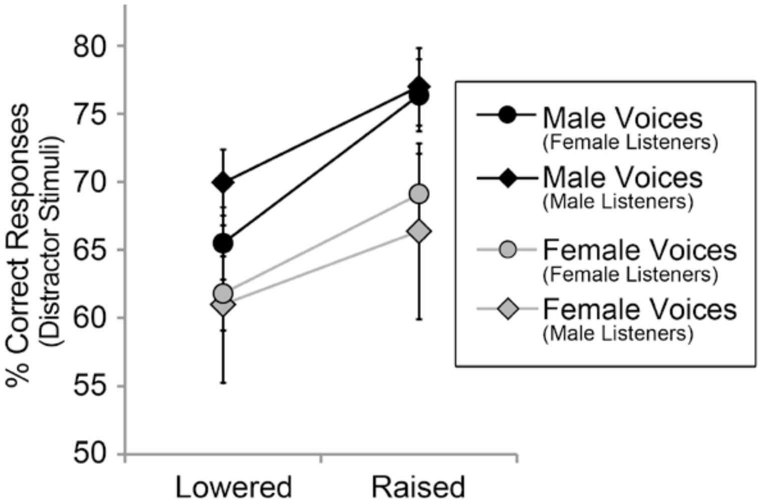

Figure 5.2
Percent correct responses for manipulated-F_0 *distractor stimuli in the test phase* (*spoken word trials only).*

5.3.2 Female Spoken Word Recognition Trials – Experiment 2

On average, men correctly recalled hearing 71.4% of raised-F_0 target words (above chance: $t(20) = 6.04$, $p < .001$) and 75.6% of lowered-F_0 target words (above chance: $t(20) = 7.60$, $p < .001$). Female listeners correctly identified 68.6% of raised-F_0 target words (above chance: $t(52) = 7.63$, $p < .001$) and 69.9% of lowered-F_0 target words (above chance: $t(52) = 9.36$, $p < .001$)

Men were better than women at recognising target voices, but the difference was not significant (male $M = 73.5\%$, SD 14%; female $M = 69.3\%$, SD 15%; $t(72) = -1.12$, $p = .27$). There was no sex difference in performance for distractor trials (male $M = 63.7\%$, SD 27%; female $M = 65.4\%$, SD 19%; $t(72) = 0.10$, $p = .75$. Men did not have a better memory for raised-F_0 target words compared to women (male *M* = 71.4%, *SD* 16%; female *M* = 68.6, *SD* 18%; *t*(72) = -0.62, *p* = .53) or for lowered-F_0 target words compared to women (male *M* = 75.6%, *SD* 15%; female *M* = 69.9%, *SD* 16%; *t*(72) = -1.42, *p* = .16).

ANOVA analyses (within-subjects factors: accuracy by target F_0 manipulation (lowered, raised); between-subjects factor: sex of listener; covariate: masculinity preference) revealed a non-significant main effect of manipulated pitch on target word accuracy (*F*(1,71) = 3.18, *p* = .08). There were no interactions of sex of listener (*F*(1,71) = 0.66, *p* = .42) or femininity preference (*F*(1,71) = 1.20, *p* = .28). Men's degree of preference was not associated with their degree of recognition of feminised vs. masculinised target stimuli (*r*(21) = -.153, *p* = .51). Female-rated differences in attractiveness between raised- and lowered-F_0 target stimuli was positively but non-significantly associated with their degree of recognition for raised- vs lowered-F_0 stimuli (*r*(53) = .220, *p* = .11), such that women were better at recognising voices which were manipulated in the direction they found attractive.

Manipulated F_0 significantly influenced the degree of correct responses in distractor trials. ANOVA results (within-subjects factors: accuracy by distractor F_0 manipulation (lowered, raised); between-subjects factor: sex of listener; covariate: masculinity preference) revealed a significant main effect of manipulated pitch on accuracy (*F*(1,71) = 6.08, *p* = .016), indicating that feminised female voices received a lower number of accurate responses than masculinised female voices (feminised accuracy *M* = 68.3%, *SD* 24%; masculinised accuracy *M* = 61.6%, *SD* = 22%). No other effects or interactions were observed (all F < 0.22; all *p* > .64).

5.3.3 Male Textual Word Recognition Trials – Experiment 3

On average, women correctly recalled hearing 52.1% of lowered-F_0 target words and 57.8% of raised-F_0 target words. Accuracy for lowered-F_0 target words was not significantly different from chance, *t*(31) = 0.62, *p* = .54; accuracy for raised-F_0 target words significantly better than chance, *t*(31) = 2.32, *p* = .03. Male listeners exhibited similar accuracy levels, correctly responding to 53.6% of lowered-F_0 target words (n.s. different from chance: *t*(11) = 0.63, *p* = .54) and 60.4% of raised-F_0 target words (close to significantly better than chance: *t*(11) = 2.64, *p* = .07).

Men and women performed equally well on the textual word recognition trials, showing no significant differences in memory for target words (female *M* = 55.0%, *SD* 17%; male *M* = 57.0%, *SD* 17%; *t*(42) = 0.35, *p* = .73), or in identification of distractor words (female *M* = 69.1%, *SD* 15%; male *M* = 70.6%, *SD* 9%; *t*(31.86) = 0.39, *p* = .70). Women did not have a better memory for lowered-F_0 target words compared to men (female *M* = 52.1%, *SD* 19%; male *M* = 53.6%, *SD* 20%; *t*(42) = 0.91, *p* = .82), or for raised-F_0 target words compared to men (female *M* = 57.8%, *SD* 19%; male *M* = 60.4%, *SD* 18%; *t*(42) = 0.41, *p* = .68).

ANOVA analyses (within-subjects factors: accuracy by target F_0 manipulation (lowered, raised); between-subjects factor: sex of listener; covariate: masculinity preference) revealed a main effect of manipulated pitch on target word accuracy (*F*(1,41) = 4.33, *p* = .04), with higher accuracy for feminised target stimuli (*M* = 58.5%, *SD* 19%) than for masculinised target stimuli (*M* = 52.6%, *SD* 19%). There was no observed effect of masculinity preference on accuracy (*F*(1,41) = 0.01, *p* = .92), nor was there an observed effect of listener sex (*F*(1,41) = 0.21, *p* = .89). The extent to which women preferred masculinised male voices was not correlated with their recall of words which had been spoken by masculinised vs. feminised voice stimuli (*r*(32) = -.10, *p* = .58). There was also no association between male preference scores and differential accuracy for male listeners (*r*(12) = .273, *p* = .39).

5.3.4 Female Textual Word Recognition Trials – Experiment 4

On average, men correctly recalled hearing 61.3% of raised-F_0 target words (significantly above chance: $t(9) = 2.48$, $p = .035$) and 56.3% of lowered-F_0 target words (n.s. different from chance: $t(9) = 1.10$, $p = .30$). Female listeners correctly identified 55.4% of raised-F_0 target words (n.s. different from chance: $t(22) = 1.39$, $p = .18$) and 47.3% of lowered-F_0 target words (n.s. different from chance: $t(22) = -0.71$, $p = .49$).

Men and women again performed equally well on the recognition trials, with similar accuracy for target words (male *M* = 58.8%, *SD* 15%; female *M* = 51.4%, *SD* 16%; $t(31) = -1.27$, $p = .21$) and for distractor words (male *M* = 63.7%, *SD* 17%; female *M* = 73.0%, *SD* 14%, $t(31) = 1.64$, $p = .11$). Men did not have a better memory for raised-F_0 target words compared to women (male *M* = 61.3%, *SD* 14%; female *M* = 55.4%, *SD* 19%; $t(31) = 0.87$, $p = .39$) or for lowered-F_0 target words compared to women (male *M* = 56.3%, *SD* 18%; female *M* = 47.3%, *SD* 18%; $t(31) = 1.30$, $p = .20$).

ANOVA analyses (within-subjects factors: accuracy by target F_0 manipulation (lowered, raised); between-subjects factor: sex of listener; covariate: masculinity preference) found a main effect of manipulated pitch on listener accuracy ($F(1,30) = 5.35$, $p = .028$). This is reflective of a higher rate of correct identification for target words which had been spoken by feminised voices compared to target words which were spoken by masculinised voices. There were no significant interactions of listener sex ($F(1,30) = 0.14$, $p = .71$) or femininity preference ($F(1,30) = 1.98$, $p = .17$), indicating that the extent to which listeners perceived feminised female voices as more attractive than masculinised female voices was not associated with their accuracy for feminised vs. masculinised stimuli.

Women's accuracy on feminised vs. masculinised stimuli (calculated by subtracting accuracy for lowered-F_0 stimuli from raised-F_0 stimuli) was negatively correlated with their preference for feminised vs. masculinised stimuli ($r(23) = -.42$, $p = .046$), such that women tended to remember the words spoken by voices they considered less attractive, *i.e.* if women preferred feminised voices, they displayed higher accuracy with words spoken by masculinised voices, and if they preferred masculinised voices, they had greater accuracy with words spoken by feminised voices. A similar negative relationship was found amongst male listeners, but this was not significant ($r(10) = -.347$, $p = .33$).

5.3.5 Attractiveness and Memory

In male spoken word trials, using the stimulus as the unit of analysis, Pearson correlations between voice attractiveness and target accuracy showed that mean female-rated attractiveness in the learning phase was positively correlated with women's average memory for the stimuli in the test phase ($r(32) = .517$, $p = .002$). Male-rated attractiveness was not related to men's accuracy in the test phase ($r(32) = .138$, $p = .45$). Correlations between opposite-sex rated voice attractiveness and target accuracy in both spoken word and textual word trials are shown in Figure 5.3.

When stimuli from the spoken word trials were divided by high- and low-rated attractiveness (upper and lower quantiles), attractive male voices elicited greater accuracy amongst female listeners (high-attractiveness $M = 75.8\%$, SD 9%; low-attractiveness $M = 69.4\%$, SD 8%; $t(30) = 2.20$, $p = .04$). Men's memory for male voices was similarly positively affected by voice attractiveness (high-attractiveness $M = 73.0\%$, SD 8%; low-attractiveness $M = 65.3\%$, SD 13%; $t(30) = 2.07$, $p = .047$).

In female spoken word trials, Pearson correlations between voice attractiveness and target accuracy showed that mean male-rated attractiveness in the learning phase was not significantly correlated with men's average memory for the stimuli in the test phase ($r(32) = .136$, $p = .46$). Female-rated attractiveness was also unrelated to women's accuracy in the test phase ($r(32) = .120$, $p = .51$). Memory for female spoken words was unaffected by high-or low-rated voice attractiveness for listeners of either sex (male $t(30) = 0.49$, $p = .63$; female $t(30) = 0.95$, $p = .35$).

In male textual word trials, female-rated attractiveness in the learning phase was negatively but not significantly correlated with women's average memory for the stimuli in the test phase ($r(32) = -.212$, $p = .24$). This negative directionality is significantly different to the positive directionality observed in the spoken word recognition trials (Fisher's *r*-to-*z*: $z = 3.00$, $p = .003$). Male-rated attractiveness was unrelated to accuracy in the test phase ($r(32) = -.065$, $p = .73$). High-attractiveness male voices significantly impacted women's memory for textual words in the recall phase (high-attractiveness $M = 43.2\%$, SD 7%; low-attractiveness $M = 51.6\%$, SD 13%, $t(30) = 2.26$, $p = .03$). Men's crossmodal memory was unaffected by high- or low-attractiveness stimuli (high-attractiveness $M = 41.7\%$, SD 13%; low-attractiveness $M = 49.5\%$, SD 16%, $t(30) = 1.52$, $p = .14$).

In female textual word trials, male-rated attractiveness in the learning phase was not significantly correlated with men's average memory for the content in the test phase ($r(32) = .224$, $p = .22$). Female-rated voice attractiveness was positively related to stimuli recognition in the test phase, but this was short of statistical significance ($r(32) = .309$, $p = .09$). Women's crossmodal memory for textual words was unaffected by whether the original voice stimulus was of high- or low-rated attractiveness (high-attractiveness $M = 51.3\%$, SD 18%; low-attractiveness $M = 45.4\%$, SD 18%; $t(30) = 0.92$, $p = .36$). Men's crossmodal memory for female voice content was similarly unaffected by the attractiveness of the original voice stimulus (high-attractiveness $M = 60.2\%$, SD 16%; low-attractiveness $M = 56.6\%$, SD 19%; $t(30) = 0.58$, $p = .57$).

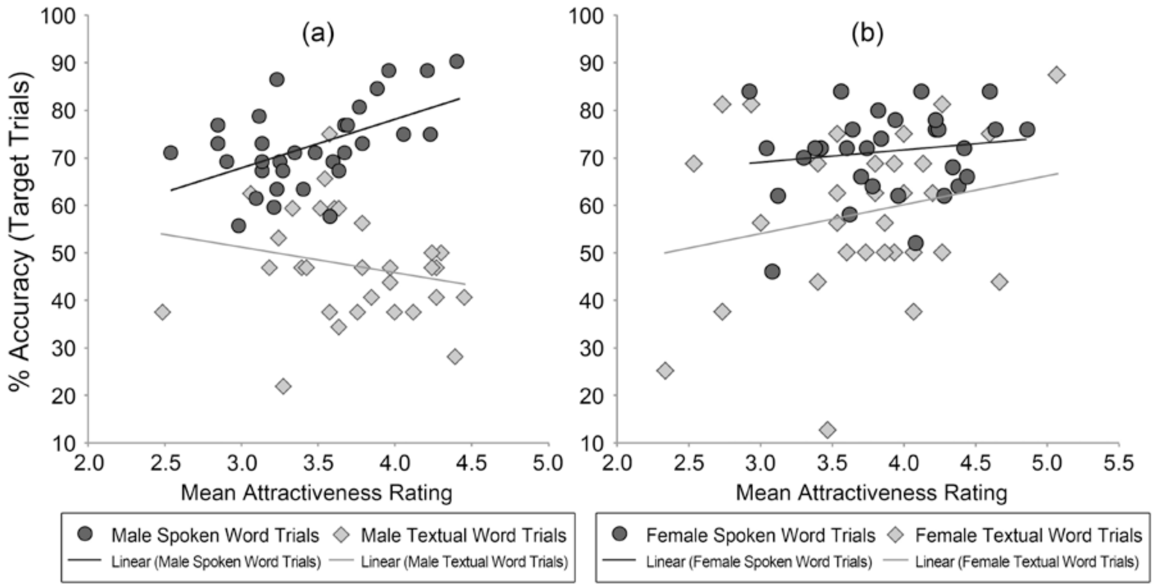

Figure 5.3
Correlations between mean opposite-sex rated voice attractiveness and mean opposite-sex target accuracy for male voices (*a*) *and female voices* (*b*).

5.3.6 Effects of Voice Attractiveness Across Modalities

Across trials and using the voice as the unit of analysis, an ANCOVA model (DV: accuracy (spoken word, textual word); covariate: rated attractiveness; fixed factor: sex of listener) revealed a significant main effect of sex of speaker ($F(1,60) = 6.68$, $p = .012$) and a significant interaction between speaker sex and voice attractiveness ($F(1,60) = 8.01$, $p = .006$) and the model was significant ($F(3,60) = 4.14$, $p = .01$). The dependent variable is the size of the difference between spoken word-trial accuracy and textual word-trial accuracy (capturing the difference between within- and cross-modality) for each stimulus. T he covariate is the mean attractiveness score taken from across both spoken word and textual word trials.

The interaction can be explained by a differing directional effect of voice attractiveness on comparative modality accuracy by speaker sex, such that attractive female voices were better at encoding cross-modal memory in the textual word trials, and attractive male voices were better at eliciting accuracy in the congruent modality (see Figure 5.4). The data was further analysed separately by sex of speaker. Near-significant main effects of voice attractiveness on the difference between voice and word recognition accuracy were found for speakers of both sexes (male voices: $F(1,30) = 4.09$, $p = .052$; female voices: $F(1,30) = 3.87$, $p = .059$). No other effects or interactions were entered into the models.

Post-hoc analysis showed that male voice attractiveness was negatively correlated with the difference in accuracy between voice and word trials ($r(32) = -.346$, $p = .05$), indicating that attractive male voices were better remembered in the voice trials, but voice attractiveness inhibited memory for content in the word trials. Conversely, female voice attractiveness was positively but non-significantly correlated with the difference in accuracy between voice and word trials ($r(32) = .338$, $p = .06$), indicating that attractive female voices were better at enhancing content recognition in the word trials, but were less likely to be remembered in the voice trials. These correlation coefficients are significantly different from each other (Fisher's r-to-z transformation; $z = -2.71$, p (two-tailed) $= .007$), indicating a significantly different directionality of effect. These relationships are shown in Figure 5.4.

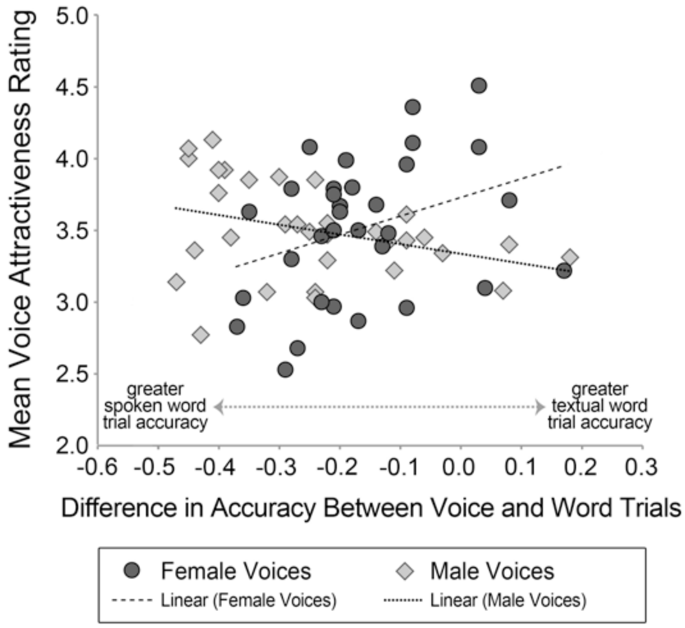

Figure 5.4
Correlations between rated voice attractiveness and differential accuracy across congruent spoken word trials and crossmodal textual word recognition trials.

An opposite-sex bias in spoken word trial accuracy was also observed using a 2x2 ANOVA (sex of voice, sex of listener) wherein men were better at recognising female target stimuli and women were better at recognising male target stimuli ($F(1,62) = 12.70$, $p < .001$). Post-hoc tests showed this was near significance for female listeners (male M = 72.6%, SD 8.9%; female M = 68.0%, SD 11.7%; $t(62) = 1.77$, $p = .08$), but no effect was found for male listeners (male M = 69.1%, SD 11.1%; female M = 71.3%, SD 9.4%; $t(62) = 0.84$, $p = .40$). No independent main effect of listener sex was found. A similar opposite-sex bias was found for textual word trials ($F(1,62) = 11.11$, $p = .001$) where men were better at recognising female target stimuli than male target stimuli (male M = 45.6%, SD 14.8%; female M = 58.4%, SD 16.9%; $t(62) = 3.23$, $p = .002$), but women were not better at recognising male over female target stimuli (male M = 47.4%, SD 11.2%; female M = 48.3%, SD 17.8%; $t(62) = 0.26$, $p = .80$). There was also a main effect of listener sex on accuracy, showing that men performed better than women on textual word trials overall ($F(1,62) = 5.41$, $p = .02$).

5.4 Discussion

5.4.1 Summary of Results

Manipulated F_0 had only limited effects on memory for voices. While pitch manipulation had no observable effects on target recognition, listeners were more likely to misattribute memory to distractor stimuli with masculinised F_0 in both male and female voices. Feminised F_0 aided listeners in the cross-modality textual content recognition task in voices of both sexes, with higher accuracy on textual word recognition trials for feminised voices of both male and female speakers. The rated attractiveness of male voices was positively related to recall accuracy in the spoken word recognition trials, independent of pitch manipulation. Conversely, attractive male voices seemed to impede content recognition in the textual word trials. No effects of female voice attractiveness were found for either spoken word or textual word recognition.

As expected, a shift in modality from auditory to visual/content recognition negatively impacted listeners' recall. Rated voice attractiveness, and not manipulated F_0, seemed to mediate this drop in performance across modalities for male stimuli: relative to spoken word recognition, lower rates of textual word/content recognition were observed for male voices which listeners rated as attractive; conversely, higher rates of textual content recognition relative to spoken word recognition were observed for attractive female voices.

5.4.2 No Effect of Manipulated F_0 in Spoken Word Trials

There was no effect of manipulated F_0 on memory in either male or female spoken word trials. This was unexpected, considering that pairing with manipulated male voices and faces affects rates of object recognition (Allan *et al.*, 2012; Smith *et al.*, 2012). Any recall of voice stimuli with this method of presentation could, however, result from listener memory for specific voices, which may activate different underlying processes relative to visual object memory. Indeed, Goldinger (1996) showed that idiosyncratic vocal attributes of spoken word stimuli are a part of episodic memory. While Allan *et al.* (2012) and Smith *et al.* (2012) both suggest mate-choice relevant mechanisms underlying the selective memory for objects associated with attractive voices, there was no effect of manipulated F_0 in the absence of visual target stimuli, and all voices were similarly likely to be remembered regardless of pitch manipulation. The lack of clear preference for masculinised vs. feminised male voices amongst study participants may have also contributed to this lack of effect. However, it should be noted that there were consistent preferences for feminised female voices, and while these results show that feminised female voices elicited higher mean accuracy than masculinised voices, this was short of statistical significance.

5.4.3 Negative Effect of Masculinised F_0 in Textual Word Trials Only

For both male and female voices, crossmodal memory for content (textual word trials) was highest for voices which had been feminised in pitch. While the preference data showed no clear preferences for either masculinised or feminised voices for male speakers, masculinised voices elicited lower rates of accuracy than feminised voices for both male and female listeners. Furthermore, there was no effect of masculinity preference on this effect. While Smith *et al.* (2012) found an effect of preference on the degree of recall of masculinised vs. feminised objects, their preference data also followed suit, with a strong preference for masculinised male voices. The present study did not reveal the same associations, although this may be because the listeners did not exhibit clear preferences.

Another possible explanation for the difference in these findings is that lowering the pitch of the voice stimuli may have resulted in voices that distracted from memory encoding. Masculinised voice stimuli are perceived as dominant and are indicative of threat potential (Puts, Apicella, *et al.*, 2012; Wolff & Puts, 2010), which may serve to distract cognitive resources away from encoding and retrieval. Listeners' attention may have thus been diverted to evaluating the voice and away from the content which was spoken. This selective allocation of attention is consistent with an adaptationist approach. Because failure to attend to stimuli which signal threat potential may result in physical harm (Öhman & Mineka, 2001), it may be prudent to divert cognitive resources to assessing the qualities of a speaker rather than the content which is spoken. Indeed, while a number of studies have found enhanced attention for angry faces (Hansen & Hansen, 1988; Van Honk *et al.*, 2001; Öhman *et al.*, 2001; Pinkham *et al.*, 2010) and threatening faces (Öhman *et al.*, 2001), angry and threatening faces are not better remembered than happy or pleasant faces (Bradley *et al.*, 1992; D'Argembeau *et al.*, 2003; Foa *et al.*, 2000). Dominant and/or threatening-sounding voice stimuli may generate the same attentional biases which could serve to detract priority from memory-related processes.

Helfrich and Weidenbecher (2011) found reduced short-term content recall for male voices which had been feminised. They found no difference in short-term content retention for unmanipulated vs. low-pitched manipulated voices. This series of experiments tested content recall based on a nonfictional text, which implies procedural or conceptual learning as opposed to basic recognition, whereas the listeners in the present study heard discrete spoken words and were simply asked whether they had heard the words in the exposure phase.

The masculinity preference data did reveal a generalised preference by both men and women for feminised female voices in textual word trials. While female voice attractiveness was not directly related to accuracy, participants of both sexes exhibited greater accuracy for words which had originally been presented in a feminised F_0 vs. a masculinised F_0. It may be that men tended to remember content spoken by feminine voices because men may receive enhanced reproductive opportunities by doing so. Additionally, women may remember content spoken by feminine female voices due to a response to intrasexual competition. These results are in agreement with Becker, Kenrick, Guerin and Maner (2005), who found enhanced episodic memory for attractive female faces. Existing literature on voices and mate preferences shows a clear relationship between vocal femininity and attractiveness in female voices (Collins & Missing, 2003; Feinberg, DeBruine, Jones and Perrett, 2008; Jones *et al.*, 2008) and feminine voice pitch is considered a cue to sexual interest, fertility status and to general mate value (Feinberg, 2008; Fraccaro et al., 2011; Puts, Jones, et al., 2012). Men may thus be unconsciously attuned to voices with traits which signal high mate value, and likewise, women may also be more attentive to these voices due to intrasexual competition. Indeed, Puts, Barndt, Welling, Dawood and Burriss (2011) found that women were particularly attuned to the cues of vocal femininity which males found the most attractive.

5.4.4 Reduced Accuracy with Crossmodal Content Recognition

Stimuli recognition was greatly decreased in textual word recognition trials compared to spoken word recognition trials, suggesting that the change in modality impeded participants' memory. This confirms existing work on the topic which also indicate that a change in modality from spoken to written has a negative impact on listener memory (Clarke & Morton, 1983; Ellis, 1982; Gibson & Bahrey, 2005; Jackson & Morton, 1984). The current study adds to this work, showing that manipulations in F_0 elicit preferential biases in this effect – that voices with high F_0 seem to facilitate greater crossmodal encoding and/or better recall by listeners.

Voice attractiveness also seems to mediate this reduction in accuracy across modalities. While attractive female voices were subject to less crossmodal memory degradation than unattractive female voices, the opposite was true for male stimuli, such that more attractive voices were subject to greater crossmodal memory degradation than unattractive voices. As previously stated, male voice attractiveness is associated with dominance and threat potential (Puts, Apicella, *et al.*, 2012; Wolff & Puts, 2010), and these perceptions again may detract attention away from encoding and toward mate evaluation (for female listeners) and threat evaluation (for male listeners).

5.4.5 Do attractive voices help us remember?

5.4.5.1 No Effect of Masculinity/Femininity Preference

Similar to the results of Allan *et al.* (2012), there were no clear generalised preferences for masculine/feminine male voices when participants' attractiveness ratings were examined, however, contrary to their results, there was no effect of masculinity preference on memory for specific voices or spoken words. While participants of both sexes exhibited a preference for femininity in female voices, there was no effect of preference on memory for feminised voices in either the spoken word or textual word trials. This is similar to the findings of Allan *et al.* (2012), who also found significant femininity preferences, but no effect of this preference on memory for objects associated with female faces.

5.4.5.2 Remembering Attractive Voices

While female participants did not show any strong preferences for masculinised vs. feminised male voices, and there were no interactions of voice preference on memory for target stimuli, the male voices which male and female listeners considered attractive in the exposure phase were better recognised in the test phase of spoken word trials, independent of pitch manipulation. This bias for remembering attractive individuals can be linked to the finding that observers selectively attend to attractive stimuli such as faces (Becker *et al.*, 2005; Maner *et al.*, 2003). However, these studies found that recognition memory was greater for attractive female faces, while memory for attractive male stimuli was attenuated. The present data show the reverse: that attractive male voices were better remembered than unattractive voices, but female voice attractiveness was unrelated to memory.

Attractive faces may be processed as reward stimuli, evidenced by numerous functional neuroimaging studies which show greater activity in a region associated with processing rewards when participants view attractive faces compared to unattractive and neutral faces (Aharon *et al.*, 2001; Bray & O'Doherty, 2007; Cloutier, Heatherton, Whalen and Kelley, 2008; Ishai, 2007; O'Doherty *et al.*, 2003; Tsukiura & Cabeza, 2011; Winston *et al.*, 2007), and the same brain region has been shown to encode reward from auditory as well as visual modalities (O'Doherty, 2004). This activation has a direct influence on memory, with stronger connectivity between the responsible brain regions when viewing attractive facial stimuli (Tsukiura & Cabeza, 2011) and larger event-related potential (ERP) components related to encoding and recognition memory when viewing attractive faces relative to unattractive faces (Marzi & Viggiano, 2010). The results of the current study suggest that attractive vocal stimuli may be similar to attractive facial stimuli in this sense, and that attractive voices may enhance stimulus-specific memory due to their neurobiological association with reward. However, only attractive male voices were better remembered; female voices did not follow this same pattern. Because attractive female faces have been shown to activate reward circuitry (Aharon *et al.*, 2001), yet we did not find a relationship between female voice attractiveness and memory, an alternative explanation may be necessary to explain this sex difference more parsimoniously.

Women may remember attractive male voices because of the “what is beautiful is good” effect (Dion, Berscheid and Walster, 1972; Eagly *et al.*, 1991), and thus memory for attractive voices may be enhanced due to an association with a positive valence. Men may also remember these voices for the same reason. However, this still does not explain the observed enhanced memory for attractive male voices, but not for attractive female voices. Invoking mate-choice relevance to this effect, it may be inferred that women remember attractive male voices because they are especially attuned to cues of mate value present in attractive-sounding male voices (Feinberg, 2008).

Also, it may be true that men will preferentially allocate memory to attractive male voices because they inspire a threat in intra-sexual competition (Puts, Apicella, *et al.*, 2012). Male mate value is not assessed in the same way as female mate value, which is less variable as males are less “choosy” when evaluating potential mates (Trivers, 1972). Variation in attractiveness is thus more costly to males than to females - an unattractive male may have no mating opportunities and subsequently, no offspring, while unattractive women are more secure in their reproductive potential (Trivers & Willard, 1973). As such, observers may not attend to variations in female attractiveness to the same degree as variations in male attractiveness. This explanation is supported by the pattern of results in Smith *et al.* (2012), who found that sexual dimorphism of male voices affected object memory, while sexual dimorphism of female voices did not.

There were also opposite-sex biases amongst the listeners, such that men performed better on female stimuli and women performed better on male stimuli. This is similar to the results of the object recognition studies of Smith *et al.* (2012) and Allan *et al.* (2012), both of which suggest mate-choice relevant memory based on sex differences in memory for objects associated with masculine male faces. Both of these studies have shown preferential recall for these items by women (though Smith *et al.* do not include male participants), and no results using female stimuli. The present results have added to this area of research by showing that listeners exhibit preferential recall of opposite-sex voices in addition to the content which is spoken by voices of the opposite sex.

As alternatives to mate-choice relevance, other explanations may be equally valid, including the aforementioned link between attractiveness and reward activation, and increased attention due to emotional arousal (Cloutier *et al.*, 2008; Kitayama, 1996). These explanations are not necessarily mutually-exclusive, and all may be applicable to a certain degree. Further research may be carried out to tease apart the relative causality of this observed effect.

5.4.5.3 Remembering Content

While it has been shown that attractive male voices are more likely to be remembered than unattractive male voices, the results concerning memory for the content of what is spoken are more mixed. Although male voice attractiveness was positively related to memory for spoken words, attractiveness was negatively related to memory for textual words (*i.e.* content). This suggests that voice attractiveness may play divergent roles in these two types of memory.

Both male and female voices with feminised F_0 were better at encoding content in the exposure phases of textual word trials. While low F_0 is generally considered attractive among male voices, low F_0 is also linked to the attribution of negative and anti-social traits including dominance and physical threat (Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2007), and it is important to note again that listeners did not show any clear preference for masculinised F_0 in their attractiveness ratings of the male stimuli. Further, a low F_0 is attributed to high levels of testosterone, which is also associated with numerous antisocial behaviours (Archer, 1991; Dabbs & Mallinger, 1999; Mazur & Booth, 1998). Thus, a feminised F_0 in both male and female voices may be associated with pro-social personality traits and low threat potential. These perceptions of pleasantness and low threat potential could better facilitate memory compared to unpleasant or threatening voices. Indeed, Schacter and Church's (1992) series of stem-completion task experiments showed that stimuli pleasantness positively influenced stem-completion rates to a greater degree than voice pitch, indicating that pleasant-sounding voices may enhance memory.

While there was a positive association between male voice attractiveness and memory in the spoken word trials, in textual word trials, the relationship was negative, indicating that memory of both men and women may be impeded for words originally spoken by attractive male voices. It may be, then, that attractive male voices distract from the content of what is spoken rather than enhance attention and facilitate memory encoding. These results agree with Maner *et al.*'s (2003) finding that both male and female observers exhibited attenuated memory for attractive male faces, and that women attend to, but do not remember, attractive men (Maner *et al.*, 2003). Attractive male voices may activate mental functions that are not related to memory, for example emotional arousal (related to perceived attractiveness and/or threat potential). Kitayama (1996) found similar results using emotional stimuli – if a stimulus was delivered in an emotional voice, it was less likely to be remembered, providing that memory load during encoding was low. The study design of the present research put no memory load on the listeners, as they were unaware during encoding that they would be asked to recall what they had heard while rating the voices for attractiveness.

The findings for female spoken words show no direct relationship between voice attractiveness and memory for content, however, content spoken by feminised female voices in the textual word trials was better recognised than content which was spoken by masculinised female voices, and listeners exhibited a clear preference for feminised voices compared to masculinised voices. This is consistent with the work of Maner *et al.* (2003) and Becker *et al.* (2005) who show that attractive female faces receive more attention from both men and women. Unlike male voices, attractive female voices are high in pitch, and as such, may not sound as threatening as attractive male voices, which are generally lower in pitch. The increased crossmodal recognition accuracy may be attributed to enhanced attention due to voice attractiveness, without the distraction of perceived physical dominance or threat potential.

5.4.6 Effect of Manipulated F_0 on Distractor Stimuli Accuracy.

Manipulated F_0 also had a significant effect on accuracy for distractor stimuli, with masculinised voices of both male and female speakers eliciting lower accuracy (or a greater number of false-positives) than feminised voices. Masculine F_0 is positively linked to attributions of dominance (Puts, Apicella, *et al.*, 2012; Puts *et al.*, 2007; Wolff & Puts, 2010). Listeners may have thus been influenced by the perceived dominance of the voices, causing them to mis-attribute memory to voices which they perceived as dominant or authoritative. They may additionally have been more prone to acquiesce to dominant-sounding stimuli than to non-dominant sounding stimuli due to an adaptive response related to the perceived threat potential of masculinised voices (Puts, Apicella, *et al.*, 2012). Future research may explore this relationship by measuring perceived dominance and attractiveness of distractor stimuli following recognition trials.

5.5 Conclusions

Manipulated F_0 did not influence memory for spoken words, however, when the sensory modality of recognition stimuli was changed from auditory to textual (crossmodal content recognition), participants of both sexes were better at recognising content spoken by raised-F_0 (feminised) voices of both male and female speakers. This may be because masculinised voices were more distracting than feminised voices, and may have diverted attention away from the encoding process, possibly via listeners' perceptions of vocal dominance and threat potential. Listeners of both sexes were more likely to mis-attribute memory to lowered-F_0 (masculinised) distractor stimuli of both male and female speakers in spoken word trials, which may be a result of an acquiescence bias for voices which sound dominant or threatening.

While attractive male voice stimuli were better recalled by both male and female listeners, male voice attractiveness attenuated recognition of spoken content. It may be beneficial for listeners of both sexes to remember attractive male voices due to mate-choice relevant motives and to intrasexual competition; however, the content spoken by attractive male voices may be less likely to be remembered because attractive voices may activate different cognitive processes which detract from memory, such as reward processing, emotional arousal and threat evaluation. Female attractiveness was not related to recognition of spoken words or textual content, but attractiveness did positively influence the recollection of textual content over spoken words, which is opposite from the directionality seen for male voice attractiveness. This sex difference in attention and memory for attractive individuals may be related to sex differences in the rewards and consequences of variance in mate value. The results of this study serve to further investigate the relationship between stimulus attractiveness and adaptive memory by examining both male and female voices using listeners of both sexes in a congruent- and cross-modal design, and may be of interest to those interested in the communication of important messages. Overall, the relationship between voice attractiveness and memory is complicated, and may be influenced by perceived vocal masculinity and dominance.

CHAPTER 6: Masculinity of Voices, Faces and Videos: A Cross-Modal Matching Task

Much prior work on vocal masculinity is predicated on the assumption that vocal and facial masculinity are dual cues that reflect underlying genetic quality. A handful of experiments have shown that while women expect masculine male voices to be accompanied by other sexually dimorphic traits, in reality masculine voices seem to stand on their own. In this chapter, I test whether vocal and facial masculinity are correlated in both men and women. I also test observers' abilities in matching voices to their respective faces, and whether concomitant masculinity assists participants in making correct matches. I examine these questions in three studies using static and moving images and videos of both male and female speakers. While I found that masculinity of faces and voices was unrelated in both men and women, participants were still able to match male (but not female) static faces to their respective voices at rates above chance, which suggests that while they did not co-occur, observers did not *expect* them to co-occur, and did not match faces and voices assortatively on the construct of masculinity. Whether vocal and facial masculinity reflect the same underlying qualities is not necessarily dependent upon their concomitance, and the developmental processes which underlie the development of these two secondary sexual characteristics is discussed.

6.1 Introduction

Facial and vocal masculinity are both secondary sexual characteristics linked to the expression of elevated testosterone levels around the age of puberty (Akcam *et al.*, 2004; Beckford *et al.*, 1985; Van Borsel, De Cuypere, Rubens and Destaerke, 2000; Dabbs & Mallinger, 1999; Harries *et al.*, 1997; Hollien, Green and Massey, 1994; Thornhill &

Gangestad, 1999). Face and voice masculinity are sexually dimorphic, with male faces having larger lower facial features and more robust brow bones relative to female faces, and men's voices are lower in pitch than women's voices (Feinberg, 2008; Perrett *et al.*, 1998; Thornhill & Gangestad, 1999). Because females do not experience the same hormonal changes during puberty as males, they do not experience changes in faces and voices in the same way (Perrett *et al.*, 1998). Femininity in women's faces and voices are linked to oestrogen and fertility (Abitbol *et al.*, 1999; Thornhill & Grammer, 1999), which are cued by the presence of feminine facial features, including smallness of the jaw and chin, enlarged lips and cheeks, and a high voice pitch (Feinberg, 2008; Feinberg, Jones, DeBruine, *et al.*, 2005; Johnston & Franklin, 1993; Puts, Jones, *et al.*, 2012; Thornhill & Gangestad, 1999). Because facial and vocal masculinity both signal the presence of testosterone, and because preferences for masculine male voices and faces are correlated, these two traits are generally considered to be dual cues to mate value. Likewise, femininity in female voices and faces are also both preferred by men, are indicative of oestrogen, and both are considered to reflect a woman's underlying genetic and reproductive quality (Collins & Missing, 2003; Feinberg, 2008; Feinberg, Jones, DeBruine, *et al.*, 2005).

Research into the concomitant presence of these dual cues to mate value has revealed divergent results based on the gender which is studied. In women, facial and vocal attractiveness are linked to femininity (Feinberg, Jones, DeBruine, *et al.*, 2005), and they are correlated with each other – women who have attractive faces also have attractive voices, and men who are asked to imagine the faces that "belong" to the voices which they find attractive result in feminine conceptualisations (Collins & Missing, 2003; Feinberg, Jones, DeBruine, *et al.*, 2005; Lander, 2008; Röder, Fink, Feinberg and Neave, 2013). This suggests not only that facial and vocal femininity in women are related, but also that they are expected to co-occur by listeners and viewers.

Conversely, masculinity in male faces and voices do not have the same kind of association. Collins (2000) showed that masculine male voice traits (low pitch and low formant spacing) did not correlate with any body characteristic she examined, however, facial masculinity was not assessed. Additionally, while women exhibited strong agreement that masculine voices belonged to men who were older, heavier, more muscular and more likely to have a hairy chest, the women were incorrect on all assessments apart from weight. There is also some evidence that men who have attractive faces also have attractive voices. While Saxton, Caryl and Roberts (2006) found that male voice and face attractiveness was correlated using isolated words as stimuli (counting from one to four), this was not replicated by Lander (2008) who found that a positive relationship between vocal and visual attractiveness was only present for moving images, but not for static face images, when using a sentence read aloud as a stimulus.

Moreover, Puts, Gaulin, Sporter and McBurney (2004) found that low voice pitch was not associated with other sexually dimorphic traits in men. Using factor analysis, they found that no other traits loaded with voice pitch. Most notably, voice pitch was unrelated to measures of physical dominance, mating success, or 2D:4D, all of which are independently associated with masculinity (Apicella *et al.*, 2007; Mazur & Booth, 1998; Mikach & Bailey, 1999; Neave, Laing, Fink and Manning, 2003; Puts *et al.*, 2004). Women do, however, exhibit preferences for both male voices and faces that are masculine, and preferences for masculinity in these two traits are correlated amongst individual observers (Feinberg, DeBruine, Jones and Little, 2008; Feinberg, Jones, Little, *et al.*, 2005; Jones, Feinberg, *et al.*, 2010; Perrett *et al.*, 1998). This does suggest that male facial and vocal masculinity are related as testosterone-dependent cues to mate value (Feinberg, 2008), but whether they can be expected by observers to co-occur is somewhat unclear. While the results of Collins (2000) show that women *expect* masculine traits to co-occur in men, the fact that they were incorrect on their assessments suggests that these traits may not be associated in a sample of naturalistic, unmanipulated faces and voices.

Studies have shown that matching voices and static images of faces is not something at which we are generally successful. Kamachi, Hill, Lander and Vatikiotis-Bateson (2003) and Lachs and Pisoni (2004) both presented tasks to participants in which they were asked to match a voice recording with the image of the correct speaker when given the option of a correct target image and an incorrect distractor image. Both of these studies report accuracy rates that were not significantly different from chance (56% and 50% accuracy, respectively). Using a similar design, Mavica and Barenholtz (2013) report accuracy rates at 57%, which was significantly above chance in this study, but still very close to the non-significant 56% reported by Kamachi *et al.* In comparison, a number of studies have shown that accuracy is above chance when voices are matched to muted (i.e. visual-only) video images of individuals speaking, both when they are speaking the same utterances and when speaking different utterances relative to the auditory stimulus (Kamachi *et al.*, 2003; Lachs & Pisoni, 2004; Lander, Hill, Kamachi and Vatikiotis-Bateson, 2007). Still, accuracy rates range from just 58-63% - while these percentages are indeed above chance, they cannot be considered high (see Table 6.1 for a full comparison). That test subjects are better at pairing voices to moving images of speakers, but not to their static faces, suggests that the movement of the lips and face ("speech reading") may be necessary for the task.

While all four of these studies use stimuli of both sexes in their matching tasks, only Mavica and Barenholtz (2013) looked for effects based on the gender of the stimuli, but report no difference in mean accuracy. It is also notable that this was the only study to find matching accuracy for static images at rates significantly above chance. Mavica and Barenholtz (2013) also examined differences in rated speaker traits across modalities. They examined listener perceptions of height, weight, and age, as well as openness, conscientiousness, extraversion, agreeableness and calmness, but listeners' accuracy was not affected by differences in any of these dimensions across types of stimuli. This suggests that participants did not make their matches using assortative judgements of any of these traits.

Table 6.1
Summary of previous studies matching voice to static face and video identities. Mean accuracy rates are listed in addition to whether or not the reported rates are above chance.

Publication	Gender of Stimuli	Stimulus Models	Stimulus Content	Static Faces	Congruent Video	Incongruent Video	Temporally Reversed
Kamachi *et al.* (2003)	both, sex diff's not examined	Japanese	sentences	no, 56%	yes, 61%	yes, 59%	no, 55%
Lachs and Pisoni (2004)	not stated, sex diff's not examined	Caucasian	single words	no, 50%	yes, 61%	-	no, 53%
Lander *et al.* (2007)	both, sex diff's not examined	Japanese	sentences	-	yes, 63%	yes, 58%	-
Mavica and Barenholtz (2013)	both, no sex diff's found	Caucasian	single words	yes, 53%	-	-	-
			sentences	yes, 57%	-	-	-

Because current literature presents mixed results on whether participants can successfully match identity across modalities, more work is needed to determine the scale of these abilities. The present research examines matching accuracy across three separate studies in which participants were asked to match voices to static faces, to muted congruent videos (where speakers are reading the same sentence as in the voice stimulus), and to muted incongruent videos (where speakers are reading a different sentence relative to the voice stimulus). We also expand upon the results of previous matching studies by looking at differential masculinity between voices and faces. If masculinity differences between voices and faces affects matching accuracy, this would lend support to the idea that voice and face masculinity are expected to "go together"; conversely, if there is no effect of differential masculinity on matching accuracy, it may be that voice and face masculinity are not expected to co-occur. Furthermore, we test this hypothesis while examining both male and female stimuli separately. Previous studies have not examined masculinity differences in a matching task, and most have not examined differences based on the sex of stimulus.

6.2 Methods

6.2.1 Stimuli

30 men and 30 women, all Caucasian, were recruited as stimuli donors. Photographs were taken using a digital camera at a standardised distance of 100cm. Facial photographs were masked such that hair and clothing were removed from the image and only the face of the individual was shown. Videos were obtained using a Canon XM2 3CCD digital video camera, and an Audio-Technica AT-4041 microphone with a cardioid pickup pattern, at a distance of approximately 65cm. Audio and video were recorded directly to hard disk using Windows Movie Maker v.2.1.4027.0, with a 48kHz sampling rate and 16-bit quantisation in a partially-soundproofed room. Participants were recorded while reading a scripted text mounted next to the video camera. This text was selected due to its neutrality of content. For the purposes of this experiment, one sentence was extracted from this video recording: "October frequently brings the first frost of the season over the greater part of the UK." Those participants whose first language was not English and those who had difficulties reading from a script (*e.g.* omitting words, stuttering, long pauses, or repeating words) were excluded from the stimulus set. Additionally, participants over the age of 30 were excluded so that perceived age would not play a role in participants' matching abilities (due to age being reliably detected from the voice). From the remaining participants, 20 men (ages 18-30; mean age 20.35, *SD* 2.87) and 20 women (ages 18-23; mean age 19.25, *SD* 1.33) were randomly selected to serve as our stimulus set.

Videos were edited (sentence extraction) using Windows Movie Maker v.6.0.6000.16937, and saved as .wmv files encoded at 1Mbps/25 frames per second. Sound was extracted from the videos using Windows Media Stream Editor (WME9). This created matched muted .wmv video files and .wma audio files. For use in flash coding (for stimuli presentation), videos were converted to .mp4 using OJOsoft Total Video converter

2.7.2.1017(A), at 24fps/H.264 codec/AVC/768kbps; audio files were converted to mp3 in mono at 320kbps using Switch v.2.04.

6.2.2 Participants

In the static face trials (Study 1), participants (*N* = 79) were 54 women (aged 14 – 50, mean age 22.1, *SD* 8.4 years) and 25 men (aged 17 – 48, mean age 23.2, *SD* 7.9 years). In the congruent video trials, participants (*N* = 75) were 62 women (aged 17 – 28, mean age 19.5, *SD* 2.1 years) and 12 men (aged 17 – 49, mean age 22.5, *SD* 8.8 years). In the incongruent video trials, participants (*N* = 74) were 61 women (aged 16 – 40, mean age 20.5, *SD* 4.7 years) and 13 men (aged 17 – 28, mean age 19.7, *SD* 3.4 years).All participants took part voluntarily; some participants received credit toward fulfilling a course requirement. All phases of this experiment were approved by the ethics committee at the University of Stirling.

6.2.3 Procedure

In the static face trials (Study 1), each target image was paired randomly with a distractor image. This created 10 pairs of faces for each sex. In the congruent video trials (Study 2), the same stimulus set of 20 male and 20 female speakers was used as in Study 1. The voices and target stimuli remained the same and distractors were re-shuffled in order to create new pairs. In the videos, speakers were reading the same sentence as in the audio stimuli. In the incongruent video trials, the same stimulus set of 20 male and 20 female speakers was used as in Studies 1 and 2. The voices, target stimuli and distractor stimuli remained the same and were not reshuffled to create new pairs relative to Study 2. In the videos, speakers were reading a different sentence as in the audio stimuli (audio: “October frequently brings the first frost of the season over the greater part of the UK;” video: “But occasionally, summer takes a last lingering look at the mature colours of the autumn countryside”).

These image or video pairs were always presented together to all study participants, such that everyone participating in the experiment responded to the same pairs. While listening to a voice stimulus, participants viewed the pair of target and distractor faces or videos which appeared below the audio player icon. Participants were instructed to choose which face or video they thought belonged to the voice. Screen location of the target and distractor images (left/right) was randomised. Male and female pairs were presented in separate blocks, and pair presentation order was randomised. For video pairs, participants played two muted videos which appeared below the audio player icon; they were required to click a “play” button in order to view the videos, as we did not want these to play concurrently with the voice, which would conflate accuracy by making the video which synched up with the audio track a clear choice to participants. After playing both videos, participants were instructed to choose which video they thought belonged to the voice they had previously heard. Screen location of the target and distractor images (left/right) was randomised.

Following the matching task, participants were then re-exposed to the voices and faces or videos they had previously matched and asked to rate them for masculinity on a 7-point Likert scale. Again, male and female stimuli were presented in separate, randomised blocks.

6.2.4 Data Treatment and Accuracy Variables

Correct responses were coded with the number “1” and incorrect responses with “0”. It was thus possible to calculate accuracy scores as percentages. Because the target and distractor faces differed in masculinity from one another, it was expected that this difference in masculinity would correlate with listeners' matching abilities, i.e. if the paired stimuli were very different in masculinity, matching would in theory be easier, and corresponding increases in accuracy for these pairs may be observed. Conversely, if the target and distractor faces were similarly masculine, matching them should be more difficult, and this should correspond to a reduction in accuracy. In order to capture this difference, the mean rated masculinity of the distractor face was subtracted from the mean rated masculinity of the target face for each pair, with negative values indicating pairs in which the distractor face was more masculine than the target face, and positive values indicating pairs in which the target face was more masculine than the distractor face (Figure 6.1a). The absolute value of this difference was used as a variable in the analyses in order to capture the absolute (non-directional) difference in masculinity between the stimuli.

Two additional variables were required to reflect the difference in masculinity between the voice and the two faces presented to participants. The difference in masculinity between the voice and target face was calculated by subtracting the mean rated masculinity of the target face from the rated masculinity of the voice, with negative values indicating that the face was more masculine than the voice, and positive values indicating that the voice was more masculine than the face (Figure 6.1b). Again, the absolute value of this difference was used in the analyses, with the prediction being that similarity would facilitate matching accuracy, and dissimilarity would impede accuracy. The same method was used to calculate a variable which reflected the difference in masculinity between the voice and the distractor face (Figure 6.1c), with the prediction being that similarity between the voice and distractor image would impede matching accuracy, and dissimilarity would facilitate accuracy. The distribution of all three masculinity difference variables we calculated passed K-S tests for normality, apart from the difference between voice and target face masculinity among males in the static face condition, which was skewed toward similarity (values approaching zero), however, there was still a distribution of difference scores with a range from 0.27 to 2.13 (*M* = 0.89, *SD* 0.69). Figure 6.1 contains a visual representation of how these variables relate to the stimuli.

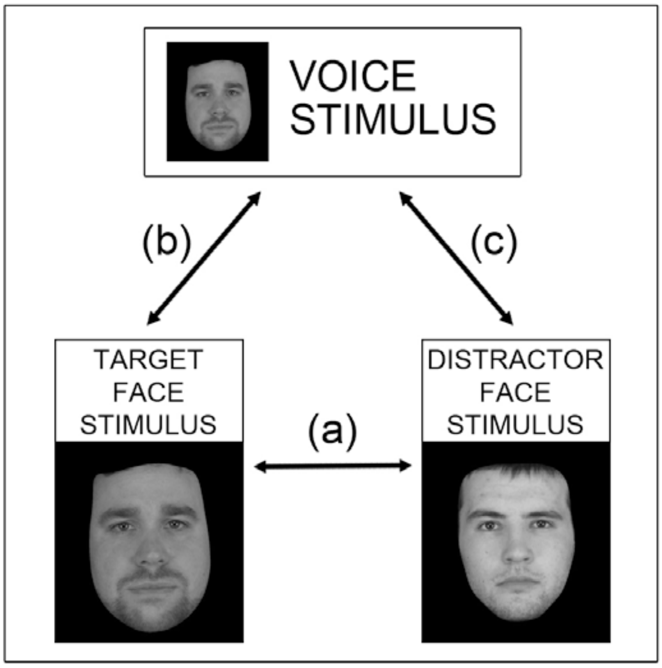

Figure 6.1
Diagram of stimuli and masculinity difference variables. The differences in masculinity between the target and distractor face (a), between the voice and target face (b), and between the voice and distractor face (c) are shown.

6.3 Results

Accuracy scores were averaged across participants to capture mean accuracy per matching pair. Mean values were also calculated for masculinity scores for each voice, target face, and distractor face. This was done separately for Studies 1, 2 and 3.

6.3.1 Study 1: Voices and Static Face Images

Participants matched male pairs at rates significantly better than chance (*M* = 67.2%, *SD* 12%; $t(78) = 12.45$, $p < .001$), but matching for female pairs was worse than chance (*M*

= 46.6%, *SD* 15%; *t*(78) = -2.02, *p* = .046). See Figure 6.2. Participants were significantly more accurate when matching male pairs compared to female pairs (*t*(78) = -10.57, *p* < .001). Taken together, mean overall accuracy was above chance (*M* = 56.9%, *SD* 11%; *t*(78) = 5.77, *p* < .001). Men and women performed equally well across trials for male pairs (*t*(77) = 0.60, *p* = .55) and for female pairs (*t*(77) = 1.71, *p* = .09).

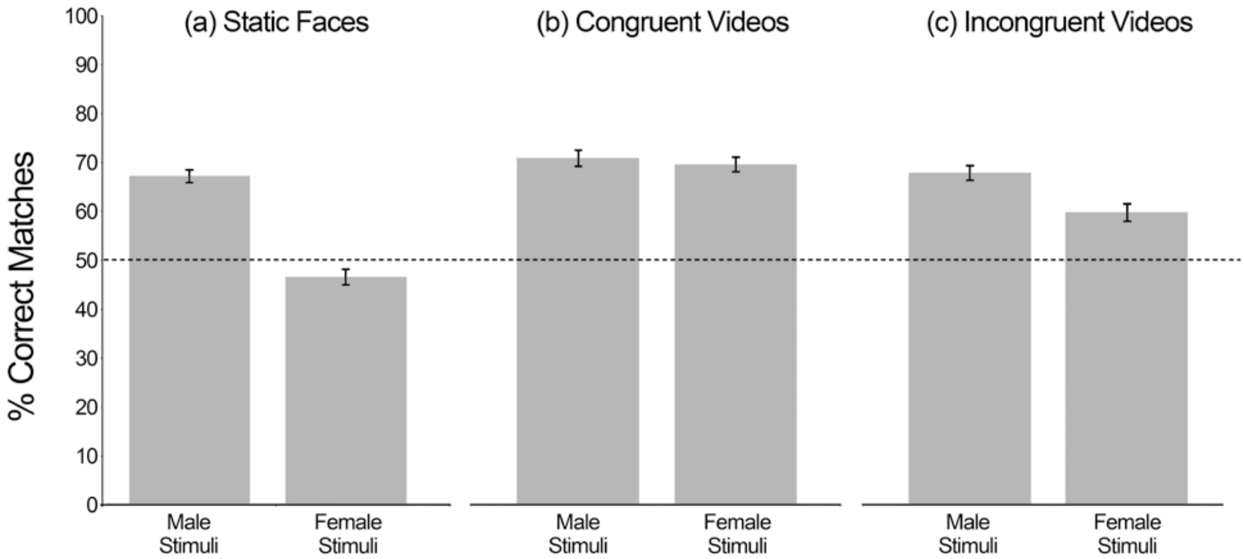

Figure 6.2
Listener performance for Studies 1, 2 and 3. Matching accuracy on male and female stimulus pairs using static faces in Study 1 (a), congruent videos in Study 2 (b), and incongruent videos in Study 3 (c). Error bars represent standard error of the mean, with chance accuracy denoted at 50%.

6.3.2 Study 2: Voices and Congruent Videos

Mean accuracy was significantly better than chance for male pairs (*M* = 70.8%, *SD* 14%; *t*(73) = 12.43, *p* < .001) and for female pairs (*M* = 69.5%, *SD* 14%; *t*(73) = 11.85, *p* < .001). Taken together, overall accuracy was 70.1% (SD 11%; above chance, *t*(73) = 15.27, *p* < .001). Mean accuracy was not significantly different based on stimulus sex (*t*(73) = 0.67, *p* = .50). Men and women performed equally well across trials for male pairs (*t*(72) = 0.65, *p* = .52) and for female pairs (*t*(72) = 1.66, *p* = .10).

6.3.3 Study 3: Voices and Incongruent Videos

Accuracy was above chance for male pairs (M = 67.8%, SD 13%; $t(73) = 11.39$, $p < .001$) and for female pairs, but to a lesser degree (M = 59.7%, SD 17%; $t(73) = 5.01$, $p < .001$). Taken together, overall accuracy was 63.8% (SD 11%; above chance, $t(73) = 10.96$, $p < .001$). Mean accuracy was significantly higher for male pairs than for female pairs ($t(73) = 3.27$, $p = .002$). Men and women performed equally well across trials for male pairs ($t(72) = 0.72$, $p = .47$) and for female pairs ($t(72) = 0.48$, $p = .63$).

Changing the video stimulus from congruent to incongruent reduced mean accuracy scores, but this did not reach statistical significance (congruent accuracy M = 70.3%, SD 12%; incongruent accuracy M = 63.8%, SD 15%; $t(19) = 1.94$, $p = .07$). When analysed separately by stimulus sex, this decrease was significant for female pairs (congruent accuracy M = 70.9%, SD 11%; incongruent accuracy M = 59.7%, SD 14%; $t(9) = 2.33$, $p = .04$), but not for male pairs (congruent accuracy M = 69.6%, SD 14%; incongruent accuracy M = 67.8%, SD 15%; $t(9) = 0.41$, $p = .69$). Accuracy between congruent and incongruent conditions was correlated but nonsignificant for male pairs ($r(10) = .55$, $p = .10$); the correlation was weaker for female pairs and did not approach significance ($r(10) = .29$, $p = .41$).

6.3.4 Masculinity Differences

Across all three studies, there were no significant correlations between masculinity differences and matching accuracy, although some moderate directionalities were observed (see Table 6.2).

Table 6.2 *Predicted and observed directionalities and correlations between masculinity difference variables and matching accuracy.*

Relationship with accuracy	Voice, Face Difference	Voice, Distractor Face Difference	Face, Distractor Face Difference
Predicted	negative	positive	positive
Voice + Face Trials, M	.30, n.s.	.51, n.s.	-.27, n.s.
Voice + Face Trials, F	.03, n.s.	.59, n.s.	-.48, n.s.
Voice + Congruent Video Trials, M	-.21, n.s.	-.24, n.s.	.59, n.s.
Voice + Congruent Video Trials, F	.38, n.s.	.03, n.s.	-.46, n.s.
Voice + Incongruent Video Trials, M	-.56, n.s.	-.39, n.s.	.44, n.s.
Voice + Incongruent Video Trials, F	.23, n.s.	.55, n.s.	.01, n.s.

6.3.4.1 Independent and Partial Correlations in Study 1 (Male Faces)

Masculinity of male voices and faces was not correlated, although a negative directionality was observed ($r(10) = -.29$, $p = .41$). Masculinity similarity between male voices and target faces was not correlated with accuracy ($r(10) = .30$ $p = .40$), although a positive directionality was present, indicating that voice-face pairs that were dissimilar in masculinity were more easily matched than voices and faces which were similarly masculine. Controlling for the difference between the voice and the distractor face and for the difference in masculinity between the two faces presented increased the correlation coefficient, but fell short of statistical significance (partial $r(6) = .54$, $p = .17$).

The difference in masculinity between the voice and the distractor face was positively but non-significantly correlated with accuracy ($r(10) = .51$, $p = .13$), indicating that pairs with greater differences in masculinity between the voice and distractor face were more easily matched. Controlling for the other two difference variables again increased the correlation coefficient, but this was still short of statistical significance (partial $r(6) = .65$, $p = .08$).

The difference in masculinity between the target and distractor faces was also positively but nonsignificantly correlated with matching accuracy ($r(10) = .23$, $p = .51$), indicating that face pairs which were similarly masculine were less easily matched. Controlling for the other two difference variables reversed the directionality of the effect, but this again was nonsignificant (partial $r(6) = -.37$, $p = .37$).

6.3.4.2 Independent and Partial Correlations in Study 1 (Female Faces)

Masculinity ratings of female voice and face stimuli were not correlated ($r(10) = .14$, $p = .69$). Masculinity similarity between female voices and target faces was not independently correlated with accuracy ($r(10) = .03$, $p = .93$). Controlling for the difference between the voice and the distractor face and for the difference in masculinity between the two faces presented did not substantially increase the effect size or significance (partial $r(6) = .13$, $p = .76$).

The difference in masculinity between the voice and the distractor face was positively but non-significantly correlated with accuracy ($r(10) = .59$, $p = .07$), indicating that pairs with greater differences in masculinity between the voice and distractor face were more easily matched. Controlling for the other two difference variables reduced the correlation coefficient and the significance level (partial $r(6) = .43$, $p = .28$).

The difference in masculinity between the target and distractor faces was also positively but nonsignificantly correlated with matching accuracy ($r(10) = .43$, $p = .21$), indicating that face pairs which were similarly masculine were less easily matched. Controlling for the other two difference variables reduced the effect size and statistical significance (partial $r(6) = .10$, $p = .82$).

6.3.4.3 Mixed Effects Model for Study 1

Because similar directionalities were observed in the correlation analyses with respect to stimulus sex, the data were analysed for male and female stimuli together in a linear mixed effects model. All mixed model analyses were performed in R (R Core Team, 2014) using the library lmerTest (Kuznetsova et al., 2014). The model which best fit the data was comprised of two fixed effects (controlling for subject and stimulus pair as random effects). The fixed effects are as follows: (1) the difference in masculinity between the voice and the distractor face (F = 8.12, p = .004) and (2) the sex of the stimulus (F = 9.13, p = .007), and the model was significant ($\chi^2(2)$ = 15.48, p < .001). The accuracy data was further examined separately by stimulus sex to determine the directionality of this effect. For male stimuli, no fixed factors were found which predicted matching accuracy. For female pairs, the difference in masculinity between the voice and the distractor face affected matching accuracy (F = 12.79, p < .001) and the model was significant ($\chi^2(1)$ = 14.63, p < .001), with larger differences in masculinity eliciting greater accuracy. Including the other difference variables and listener sex did not improve the model.

6.3.4.4 Independent and Partial Correlations in Study 2 (Male Congruent Videos)

Masculinity of male voices and videos was not correlated, although a negative directionality was observed (r(10) = -.28, p = .53). Masculinity dissimilarity between male voices and target videos was not correlated with accuracy (r(10) = -.21 p = .57), although a negative directionality was present, indicating that voice-video pairs that were similar in masculinity (low difference scores) were more easily matched than voices and faces which were dissimilarly masculine. Controlling for the difference in masculinity between the voice and the distractor video and for the difference in masculinity between the two videos presented did not increase the correlation coefficient (partial r(6) = -.15, p = .71).

The difference in masculinity between the voice and the distractor face was not correlated with accuracy ($r(10) = -.24$, $p = .51$), although again a negative directionality was observed, indicating that pairs with greater similarity in masculinity between the voice and distractor video were more easily matched. Controlling for the other two difference variables increased the effect size of the correlation, but this was still short of statistical significance (partial $r(6) = -.52$, $p = .19$).

The difference in masculinity between the target and distractor faces was positively but nonsignificantly correlated with matching accuracy ($r(10) = .59$, $p = .07$), indicating that video pairs which were similarly masculine were less easily matched. Controlling for the other two difference variables increased the correlation coefficient and was statistically significant (partial $r(6) = .75$, $p = .03$).

6.3.4.5 Independent and Partial Correlations in Study 2 (Female Congruent Videos)

Masculinity ratings of female voice and video stimuli were not correlated ($r(10) = -.23$, $p = .53$), although a negative directionality was observed. Masculinity similarity between female voices and target videos was not independently correlated with accuracy ($r(10) = .38$, $p = .28$), but a positive directionality was observed, indicating that voices and videos with greater differences in accuracy were more easily matched. Controlling for the difference between the voice and the distractor face and for the difference in masculinity between the two faces presented decreased the correlation coefficient and the significance level (partial $r(6) = .29$, $p = .49$).

The difference in masculinity between the voice and the distractor video not correlated with accuracy ($r(10) = .03$, $p = .93$). Controlling for the other two difference variables increased the correlation coefficient but was still not statistically significant (partial $r(6) = .25$, $p = .54$).

The difference in masculinity between the target and distractor videos was negatively but nonsignificantly correlated with matching accuracy ($r(10) = -.46$, $p = .18$), indicating that video pairs which were similarly masculine were more easily matched. Controlling for the other two difference variables increased the effect size but still fell short of statistical significance (partial $r(6) = -.53$, $p = .18$).

6.3.4.6 Mixed Effects Model for Study 2

Because differing directionalities were observed for some correlations concerning the masculinity difference variables with respect to stimulus sex, male and female stimuli were analysed in two separate mixed models. For both male and female stimuli, controlling for subject and stimulus pair as random effects, no significant fixed effects were found to predict matching accuracy. All three masculinity difference variables were included in our analyses.

6.3.4.7 Independent and Partial Correlations in Study 3 (Male Incongruent Videos)

Masculinity of male voices and videos was not correlated, although a negative directionality was observed ($r(10) = -.47$, $p = .17$). Masculinity similarity between male voices and target videos was not correlated with accuracy ($r(10) = -.56$ $p = .09$), although a negative directionality was present, indicating that voice-video pairs that were similar in masculinity (low difference scores) were more easily matched than voices and faces which were dissimilarly masculine. Controlling for the difference in masculinity between the voice and the distractor video and for the difference in masculinity between the two videos presented increased the correlation coefficient and was statistically significant (partial $r(6) = -.77$, $p = .03$).

The difference in masculinity between the voice and the distractor face was not significantly correlated with accuracy (*r*(10) = -.39, *p* = .26), although again a negative directionality was observed, indicating that pairs with greater similarity in masculinity between the voice and distractor video were more easily matched. Controlling for the other two difference variables decreased the correlation coefficient and statistical significance (partial *r*(6) = -.03, *p* = .95).

The difference in masculinity between the target and distractor faces was positively but nonsignificantly correlated with matching accuracy (*r*(10) = .44, *p* = .20), indicating that video pairs which were similarly masculine were less easily matched. Controlling for the other two difference variables increased the correlation coefficient and was statistically significant (partial *r*(6) = .74, *p* = .04).

6.3.4.8 Independent and Partial Correlations in Study 3 (Female Incongruent Videos)

Masculinity ratings of female voice and video stimuli were not correlated (*r*(10) = -.10, *p* = .77). Masculinity similarity between female voices and target videos was not independently correlated with accuracy (*r*(10) = .23, *p* = .52), but a positive directionality was observed, indicating that voices and videos with greater differences in accuracy were more easily matched. Controlling for the difference between the voice and the distractor face and for the difference in masculinity between the two faces presented decreased the correlation coefficient and the significance level (partial *r*(6) = .29, *p* = .49).

The difference in masculinity between the voice and the distractor video not correlated with accuracy (*r*(10) = .55, *p* = .10), but a positive directionality was observed, indicating that voices and distractor videos with greater dissimilarity were more easily matched. Controlling for the other two difference variables did not substantially decrease the correlation coefficient (partial *r*(6) = .52, *p* = .19).

The difference in masculinity between the target and distractor videos was not related to matching accuracy for female pairs ($r(10) = .01$, $p = .99$). Controlling for the other two difference variables marginally increased the correlation coefficient but still was not significant (partial $r(6) = .15$, $p = .72$).

6.3.4.9 Mixed Effects Model for Study 3

Because male stimuli were matched with greater accuracy than female stimuli, and because there were differing directionalities of some correlations concerning the masculinity difference variables with respect to stimulus sex, male and female stimuli were analysed in two separate mixed models. For male stimuli, the model which best fit the data consisted of two fixed factors (controlling for subject and stimulus pair as random effects): the difference in masculinity between the target and distractor videos ($F = 3.94$, $p = .05$) and the difference in masculinity between the voice and the target video ($F = 2.04$, $p = .15$), but the model was not significant ($\chi^2(2) = 4.51$, $p = .11$). For female stimuli, no fixed effects were found which significantly described participants' accuracy.

6.4 General Discussion

6.4.1 Summary of Results

There was no relationship between vocal and facial masculinity in either men or women. Despite this lack of correlation, participants could match voices to static face images at rates significantly above chance, but only for male stimuli. Matching rates above chance for both genders were found in both of the video conditions. Similarity in masculinity between voice stimuli and their corresponding faces did not influence participants' matching abilities.

6.4.2 Matching Voices and Static Faces

In Study 1, participants matched male stimuli, but not female stimuli, at rates above chance. While Mavica and Barenholtz (2013) found no differences based on stimulus sex, average accuracy rates in the present study with male and female stimuli taken together are approximately the same (both 57% accuracy). These accuracy rates also roughly match those found by Kamachi *et al.* (2003), who found 56% accuracy, but did not report any sex differences. All of these scores differ from Lachs and Pisoni (2004), who report accuracy for matching voices to static face images at 50%, however, their study used singular spoken words as auditory stimuli, while all others, including the present study, used spoken sentences. This increase in auditory information available to the listener may have made matching easier, because connected speech contains dynamic speech patterns and variations in pitch and intonation that may not be available from singular spoken words.

This difference in stimulus content may also have made sex differences more pronounced, which could serve to explain why we found stimulus based sex differences in matching, while Mavica and Barenholtz (2013) found none. Men and women differ in the amount which they vary their voice pitch during a given utterance (measured as the standard deviation of pitch, F_0-*SD*), with women having a much higher F_0-*SD* than men (Henton, 1995). Why the presence of a more dynamic pitch range would negatively impact accuracy for matching images to female speakers, however, is unclear. It may be that the presence of F_0-*SD* served to make the voices sound much more feminine, and pairing these with faces may have been more difficult because the femininity of the voice did not match the femininity of the face. Indeed, results showed that in all of the conditions, masculinity of female voices and faces was not correlated. This may also serve to explain why, in the static face experiment (Study 1), the difference in masculinity between the voice and the *distractor* face was a significant factor in predicting participants' accuracy, and not the difference in masculinity between the voice and the correct face. While the masculinity of the distractor faces served to help participants choose the correct face, it is important to note that accuracy for female faces was still significantly below chance.

6.4.3 Matching Voices to Videos

In Studies 2 and 3, participants were able to match voices to muted videos of speakers reading the same sentence relative to the auditory stimulus. This supports previous work by Kamachi *et al.* (2003), Lachs and Pisoni (2004) and Lander *et al.* (2007), although average rates of accuracy in the present study are higher than those previously found (70%; all others 61-63%). These higher overall accuracy rates relative to both Kamachi *et al.* (2003) and Lander *et al.* (2007) might be because these previous studies have used Japanese stimuli and participants, and attendance to faces in social interactions is less salient to Japanese populations relative to Caucasian populations (Argyle & Cook, 1976; Mavica & Barenholtz, 2013). Additionally, while Lachs and Pisoni (2004) used Caucasian stimuli, only singular spoken words were used in their study, whereas the present study used complete sentences, which may have provided additional auditory information to the listeners which may have enhanced participants' ability to make correct matches.

The present results also confirm the results of both Kamachi *et al.* (2003) and Lander *et al.* (2007), who also report accuracy for incongruent video tasks at rates above chance, although accuracy rates herein again were higher (64%; others 58-59%). Furthermore, for both congruent and incongruent video tasks, no differences in accuracy were found based on stimulus sex, while the previous studies mentioned did not examine this factor.

It has previously been suggested, and is generally agreed upon, that success in matching voices to moving images is a result of "speech reading," that is, recognising the actions of the lips and lower face as being related to the auditory signal (Kamachi *et al.*, 2003; Scholl, 2004). The data presented here reflects that participants have greater success in matching speaker identities in both congruent and incongruent moving video images relative to static face images. Because matching accuracy for both male and female stimuli were above chance in both congruent and incongruent video trials, and no masculinity difference variables were associated with matching accuracy, this data support the conclusion that speech reading is utilised in order to make accurate matches. However, it is worth nothing that while congruent video trials elicited above-chance accuracy for both male and female pairs, and rates above chance were also observed for both sexes in incongruent videos, male pairs were matched at significantly higher rates than female pairs for incongruent videos. This pattern of results for incongruent videos is similar to the findings using static faces, as in both studies, male pairs were matched with greater accuracy than female pairs. It may be that participants in incongruent video trials use visual information about the movements of the face and mouth to make accurate matches, but because these moving images do not map on to the auditory stimuli precisely (as with the congruent videos), participants may have found it easier to match male pairs than female pairs. However, matching accuracy was still significantly above chance for stimuli of both genders.

6.4.4 Masculinity Did Not Aid Matching

In the stimuli examined in this study, men with masculine faces did not have masculine voices. Furthermore, no masculinity difference variables significantly predicted accuracy for male pairs in any of the three experiments. Despite the observed negative (but non-significant) relationship between male facial and vocal masculinity, participants were still able to make accurate matches at rates above chance. This suggests that male vocal and facial masculinity did not co-occur in this stimulus set, and moreover, observers did not *expect* them to co-occur. That participants were still able to match male voices and faces without using masculinity as their means of doing so suggests that participants utilised traits which were not measured in order to correctly assign voices to their owners.

Previous studies have also found that men's physical characteristics are unrelated to their voice characteristics, including other measures of masculinity such as physical prowess, muscularity, and 2D:4D (Collins, 2000; Puts *et al.*, 2004). It has been suggested that developmental timing may be responsible for the absence of concomitance of these sexually-dimorphic traits (Puts *et al.*, 2004). Testosterone fluctuates during development, and traits which are testosterone-dependent may develop at different pubertal stages. The marked drop in voice pitch experienced by males has been shown to occur relatively rapidly during puberty, marked by the clear and abrupt voice "breaking" (Harries *et al.*, 1997). Changes in facial appearance may develop more slowly, as they are constrained by the growth rate of bone structures, which may be less rapid than the growth rate of the cartilaginous vocal folds. Indeed, mandibular growth in adolescents largely maps on to the general growth "spurt," which lasts the duration of adolescence (Silveira, Fishman, Subtelny and Kassebaum, 1992). Facial morphology may thus be influenced by testosterone over a longer and/or different temporal growth stage than voice pitch.

Again, it is possible that the use of sentences as our auditory stimuli may have influenced perceptions of masculinity by including patterns of intonation and pitch variation not present in vowels or singular spoken words. This additional auditory information may have given listeners more cues to masculinity relative to shorter stimuli, such as the series of singular spoken words used by Saxton *et al.* (2006). While Saxton *et al.* (2006) found that men with attractive faces also had attractive voices, this was not replicated by Lander (2008) using spoken sentences as stimuli. The present results confirm the results of (Lander, 2008) using perceptions of masculinity rather than attractiveness. Because attractiveness and masculinity are linked in both faces and voices, results in similar directionalities are not surprising.

Furthermore, women's vocal and facial masculinity were not correlated in any of these three studies, despite previous findings which demonstrate that women who have attractive faces also have attractive voices (Collins & Missing, 2003; Lander, 2008), and that female vocal and visual attractiveness are related to femininity (Feinberg, Jones, DeBruine, *et al.*, 2005). These results were not supported using the construct of masculinity, however, different results may have been found if attributions of attractiveness or femininity had been measured, which may evoke a different characteristic than the term "masculinity". While it was expected that female voice and face masculinity would be related, and differences in masculinity between voices and correct faces or images were expected to adversely affect participants' matching accuracy, this was not the case – matching accuracy of female stimuli was not predicted by any voice or face masculinity factor except for the masculinity of the *distractor* face in the static face trials. Distractor faces with greater differences in masculinity relative to the voice presented made it easier for participants to make accurate matches, possibly because they made the correct option clearer. However, this relationship did not hold when video stimuli were used rather than static images, perhaps because additional information about mouth and face movements made matching to the target easier, and reliance upon the distractor image was no longer needed for viewers to make accurate matches.

This is not the first study to examine vocal attributions and their relationship with crossmodal matching abilities. Recently, Mavica and Barenholtz (2013) examined difference scores and matching accuracy of voices and static face images for differences in perceived traits across modalities. Three physical traits (height, weight and age) and five personality traits were examined (openness, conscientiousness, extraversion, agreeableness, and calmness). It was found that none of these difference variables correlated participants' matching abilities, however, the distractor images were not evaluated on these dimensions, nor were they included in the analyses. No mixed effects models were reported. This series of studies adds to this body of work by showing that participants do not match voices to static or moving images assortatively on the characteristic of masculinity.

6.5 Conclusions

While the rated masculinity of voices, faces and videos were not correlated for either male or female stimuli, listeners were able to match voices to static face images of men, but not women, at rates above chance. Additionally, participants were able to match voices to muted videos of the speakers at rates above chance while using muted videos of the same sentence being read relative to the auditory stimulus, and to muted videos of a different sentence being read, confirming the results of previous researchers. Furthermore, the difference in masculinity between voice and face stimuli was taken into account, and possible effects of the masculinity of the distractor images was also examined. The only significant effect of masculinity differences which were found was in the static face trails using female stimuli, where the difference in masculinity between women's voices and the distractor faces was a significant predictor of matching accuracy, such that greater differences between the voice and the distractor face made the correct face more clear to participants, however, the relative masculinity of the target face did not predict accuracy. Across all remaining trails, participants did not match voices to static or moving images assortatively on the characteristic of masculinity.

This thesis has covered a range of topics relating to the way people perceive voices, from how perceptions develop over time to what makes a voice sound pro-social and memorable. In this discussion, I will highlight the importance of each piece of work, and discuss the implications of these findings within the context of the broader field of voice perception research. I will also draw common themes across the five experimental chapters, and reflect upon how these themes relate to evolutionary psychology and to voice research in general.

7.1 Chapter Summary and Future Directions

Chapter 2 contained a critical examination of commonly-used methodologies in the field of voice perception. While a variety of types of stimuli are used, with sustained vowel stimuli being one of the most common, the validity of these stimuli as representations of real voices has been previously questioned, but never examined. I tested whether sustained vowel and scripted speech stimuli provide acoustic measurements which are representative of natural speech, and also examined whether ratings of attractiveness using these controlled speech types are comparable to attractiveness judgements made using more naturalistic recordings from the same speakers. I found that F_0, F_0-*SD* and HNR differed in at least one of the stimulus types for at least one gender, and vowel sounds provided the least representative attractiveness ratings relative to spontaneous speech. The most important finding of this study was that vowel-rated attractiveness was not correlated with attractiveness judgements made using spontaneous speech, while script-rated attractiveness was correlated with spontaneous speech attractiveness. Based on these results, I suggest that researchers may have cause to be wary about generalising the results of attractiveness studies which are based on sustained vowel stimuli, and recommend that researchers in the field of voice perception utilise scripted speech stimuli as a valid compromise between controlling for lexical content and presenting stimuli to listeners which are more ecologically valid. Perhaps an even more appropriate method would be the utilisation of spontaneous speech recordings which are made in a language unknown to the listeners, in the style of Knowles (2010), Fischer *et al.* (2011) and Leongomez *et al.* (*in press*). This method may be difficult for many researchers, for example those who work in predominantly English-speaking countries. However, as technology advances, this type of international work is becoming increasingly more feasible.

Chapter 3 also examined the ways in which the content of a stimulus can affect how it is perceived – in this case, the length of exposure time. A time course of trait perception has been previously examined for faces, with judgements being made at extremely short exposure times (≤100ms) which do not change given longer exposures, but the work in Chapter 3 is the first work to examine a time course of vocal trait attribution. These results indicate that vocal masculinity is assessed after as little as one second of exposure and is reflective of assessments made with unconstrained exposures. Furthermore, attractiveness is similarly assessed after five seconds of exposure, and trustworthiness attributions were the most changeable over time, with low similarity of results using shortened exposures of any length. In terms of evolutionary pressures, it makes sense that masculinity can be judged very quickly, as this trait is indicative of threat potential, and rapid appraisal of this trait may aid in the avoidance of conflict. Attractiveness is also judged quickly, but less so, and this added time in the time course may give listeners the opportunity to reflect on their appraisals. Trustworthiness seems to be the slowest assessment to develop, indicating that listeners take their time in making these judgements, and are unable to reliably do so reflexively. Again, this makes intuitive sense because trustworthiness is a complex personality characteristic, and may most reliably assessed after perceiving a large amount of information about the speaker.

The purpose of these chapters was to assess what kinds of stimuli should be used in voice perception research. The findings of Chapters 2 and 3 together indicate that a good stimulus is firstly ecologically valid, or at least ecologically valid enough to provide measurements and elicit perceptions which can be related to a fully naturalistic stimulus; secondly, a good stimulus is long enough to elicit attributions which are reflective of those made without time constraints. Moving forward in the field of voice perception research, these findings can be taken up by other researchers, and future studies may wish to use scripted or other-language stimuli of at least five seconds in duration. It would be interesting and pertinent to examine what types of changes in perception occur when listening to same- vs. other- language stimuli. In my MSc thesis, I found that other-language stimuli were rated as consistently less trustworthy, cooperative and attractive than same-language stimuli, but the relationships between vocal measurements and ratings were similar in both language groups (Knowles, 2010). If the dissimilarity between attractiveness ratings of vowels, scripted and spontaneous speech using other-language stimuli were similar to the same-language results presented here, it could rule out the presence of any content-related effects (initiated by the content of the script and the spontaneous nature of the responses), and could confirm whether the spoken content of the stimuli had more effect on ratings than purely perceived vocal attributes.

Opportunities for further research also exist considering the results presented in Chapter 3. Repeating the studies presented here with a larger number of stimuli may yield results which are more robust, and may improve clarity in cases which suffer from low statistical power. Existing studies of the time-course of trait perception using face stimuli have used a thin-slicing approach, with differences in exposure lengths of milliseconds rather than seconds. A more fine-grained approach could be attempted with auditory stimuli. Furthermore, another interesting avenue could be found in neuroimaging studies. It would be quite interesting to see how participants' ratings of vocal attributes map on to neurological activity, and whether exposure lengths which achieve judgements reflective of those made using unrestricted exposure times are associated with particular activations. For example, perceptions of masculinity in Chapter 3 seem to be made reflexively after the shortest exposure times. In the discussion, I suggest that it may be evolutionarily beneficial to make these judgements quickly due to the assessment of potential threat. If this is the case, there may be activation in threat-response cortical regions like the amygdala. Prior research has found that viewing untrustworthy faces was associated with automatic activation of the amygdala and the superior temporal sulcus (SCS), a cortical region associated with social interaction, while viewing trustworthy faces was associated with activation of SCS only (Winston, Strange, O'Doherty and Dolan, 2002). The time-course of these activations and any possible similarity using auditory stimuli would be interesting and relevant.

In Chapter 4, I shift focus to examine how vocal traits influence perceptions of prosocial traits of the speaker. Research in this area has previously focused on perceptions of trustworthiness, both in existing published research and in Chapter 3 herein; no research has yet examined perceptions of voice cooperativeness. Femininity in pitch traits (F_0 and F_0-*SD*) and masculinity in formant traits (D_f and P_f) were related to attributions of cooperativeness. It may be that feminine pitch traits make voices sound non-threatening, while masculine formant traits make voices sound dominant or prestigious; the combination of these two traits may make for a desirable cooperative partner.

The analyses presented in Study 1 of this chapter were gained by dividing stimuli into groups by high and low measures of the traits examined, and comparing cooperativeness ratings between the two groups. This was done in order to maximise statistical power, but a limitation of this method lies in the distribution of these traits among the voices – the stimuli which were in the high-F_0 group, for example, were not the same stimuli in the high-D_f group. While this was a useful tool for analysis, it may have been more beneficial to use a larger number of stimuli in order to perform a regression analysis; unfortunately, the number of stimuli used did not provide the necessary statistical power to provide meaningful results from such an analysis. Additionally, while I manipulated F_0 in Study 2 of this chapter, it would be useful to further examine the role that formants play in assessing cooperativeness, and conduct a study where these traits are manipulated in addition to F_0. This would allow for an assessment of the relative importance of these traits on perceptions of prosociality.

Many studies have examined ways to increase cooperation, using economic games as a method for measuring rates of cooperativeness amongst study participants. Looking forward, it would be interesting to see how manipulations of F_0 and formants relate to player contributions in public goods games, to determine whether the vocal traits which are perceived as cooperative elicit greater levels of cooperation from others. It may also be beneficial to examine whether vocal traits of speakers vary based on context. Knowles (2010) used stimuli in which voice donors "acted" the part of potential co-operator using mutual benefit cooperation and reciprocal altruism in addition to neutral contexts. This could be further developed by placing participants in more naturalistic interactions which require cooperation (and conversely, competition) to see whether any variations in vocal attributes are simply affected or whether they are "true" spontaneous responses to contexts. Some research has found that speakers vary their vocal parameters based on both mate-choice and competitive contexts (*e.g.* Leongómez et al., 2014; Puts et al., 2006); whether the same is true for cooperative contexts might be a fruitful opportunity for future work.

Chapter 5 examines what makes a voice memorable, and attempts to differentiate between memory for spoken words and memory for the content of spoken word stimuli. Words spoken by attractive male voices were better remembered than those spoken by unattractive voices, but attractiveness seemed to inhibit memory for the content which had been spoken. I linked enhanced memory for attractive spoken words to episodic memory, and the inhibited memory for content originally spoken by attractive voices may be related to the activation of cognitive processes which detract from memory encoding. Furthermore, raised voice F_0 seemed to enhance memory for content using both male and female voice stimuli in the learning phase. It may be that voices which sound masculine, or dominant, activate emotional processes which may inhibit the encoding of memory across sensory modalities from auditory (stimulus) to textual (test). So, while attractiveness plays an important role in the formation of episodic memories, voices which are masculine or dominant may hinder the formation of memories about the content of those episodes. These results may be of particular use to those who wish to communicate memorable messages – communicating public health messages, for example, may be one particular beneficial application of this work. Less noble applications may also exist for the marketing of products and services, and the communication of political messages. It may be useful to see whether these findings indeed work to enhance memory in an applied setting, for example whether raised voice pitch could enhance memory for an audio advertisement. Furthermore, integrating auditory and visual modalities may lead to interesting future work. Whether the findings presented here would still hold when paired with a visual image, such as in a television advertisement or recorded lecture, may be a valuable contribution to the field of audio-visual perception. Differences and similarities in the attractiveness of voices and faces, and their relative contributions to memory, could also be examined. For example, a raised voice pitch paired with an unattractive face may be better at encoding memory for content than a raised voice pitch paired with an attractive face.

Lastly, Chapter 6 tests whether vocal and facial masculinity co-occur, and moreover, whether observers expect them to co-occur. While it generally agreed upon that vocal and facial masculinity are common cues to mate value, whether they are concomitant has not been directly assessed. I found that masculinity in voices and faces did no co-occur for either men or women. However, participants were still able to match voices of men (but not women) to static facial photographs of their owners, suggesting that listeners did not have the expectation that vocal and facial masculinity would co-occur; or at the very least, they did not make their matches assortatively on the construct of masculinity. This finding is important because these two sexually dimorphic secondary sexual characteristics are generally thought to communicate the same thing: genetic quality. While this may indeed be the case, it may be necessary to further examine the relationships between these two traits before making generalised assumptions about their concomitance.

Further research could examine the relative developmental timing of these traits, and whether their co-occurrence (when they do co-occur) is related to enhanced attractiveness, measurable genetic quality, environmental stability, resource richness and/or reproductive success. Some research has shown that amongst males in the pubertal stage, mandibular growth rates occur during the general growth "spurt," which lasts the duration of adolescence, while vocal masculinity occurs more abruptly and is marked by the dramatic voice "breaking" during puberty. Whether the co-occurrence of masculinity is influenced by environmental factors, including the richness and stability of resources (*i.e.* nutrition) could provide valuable insights. Good health and nutrition, and the stability of these throughout men's development, tend to contribute to more attractive adult facial characteristics via low facial fluctuating asymmetry (FA) (Perrett et al., 1999), and low body FA has been shown to correlate with vocal attractiveness (Hughes, Harrison and Gallup, 2002). However, there is mixed evidence that vocal and facial *attractiveness* are correlated amongst men (Lander, 2008; Saxton et al., 2009; Saxton et al., 2006). It may well be that developmental stability could also result in more attractive phenotypes in the auditory as well as the visual domain, and developmental instability could result in voices which are masculine (because the change is sudden and may occur during a period of resource richness) and faces which are not masculine (because the change occurs over a longer time). Longitudinal data, including photographs, audio recordings and health data, would help to shed light on this topic.

While studies tend to focus mainly on either faces or voices, an integration of these two modalities is a logical and ecologically-valid step. Whether vocal attributes enhance or detract from facial masculinity (or indeed any number of assessments including attractiveness and trustworthiness) is an important line of research. While Saxton et al. (2009) found that facial, body and voice attractiveness were all positively related to one another, variation in the co-occurrence of attractiveness between individuals is likely to impact judgements of overall attractiveness when presented with all modalities concurrently. Indeed, O'Connor et al. (2012) found this to be the case, with women preferring audio-visual male stimuli that had been masculinised in both modalities. However, it has yet to be examined how this applies to naturalistic variation in these qualities. The current finding that male masculinity does not co-occur across modalities, despite women's success at matching them together, suggests that this might be a less-than-straightforward and interesting line of research.

7.2 Overarching Themes

7.2.1 Prosociality

I began this work with the aim to study the vocal characteristics which influence perceptions of prosocial traits, and these experiments appear in the centre of the thesis (Chapters 3-5). Trustworthiness is examined in Chapter 3, with cooperativeness being the focus in Chapter 4. The results of Chapter 5 are relatable to dominance attributions, which may be considered a trait negatively related to prosociality, however social dominance (or prestige) may play a positive role in ascribing cooperativeness, as discussed in Chapter 4.

Perceptions of social traits in voices appear to develop over a longer time course than masculinity, which is shown in Chapter 3 to be a fast, reflexive judgement. In this chapter, I showed that trustworthiness perceptions seem to require longer exposure times than physical traits, and this suggests that the perceptions of personality traits require some reflection on part of the listener and are less reflexive in nature. Chapter 4 shows that perceptions of a speaker's cooperativeness rely upon a combination of masculine and feminine vocal traits (pitch traits and formant traits, respectively). The perception of these divergent traits is less straightforward and may take longer than perceptions of pure sexual dimorphism, and this may explain why we observed a longer time course of perception of social traits in Chapter 3.

In Chapter 5, manipulated vocal masculinity appeared to inhibit memory for spoken content, with listeners better recalling words which were spoken by feminised voices of both sexes. While this manipulation did not affect attractiveness of the voices, masculinised voices of both genders may sound more dominant – this may have inspired listeners to attend to masculinised voices and devote cognitive resources to evaluating the potential threat of these voices, while feminised, non-dominant sounding voices better allowed listeners to encode the content of the spoken stimulus rather than focusing purely on evaluation. As such, voices which sound more friendly and prosocial may be more useful for delivering messages that will be remembered than voices which sound masculine and dominant.

Prosocial traits are beginning to be studied more thoroughly, with voice trustworthiness gaining particular traction in the scientific community. Examining the kinds of traits which positively influence social interactions has applications in the political and social spheres, and opportunities for continued examinations of these traits remain abundant.

7.2.2 Masculinity and Sex Differences

Much of the work described in this thesis examines the construct of masculinity. In Chapter 3, I found that people make judgements about the masculinity of voices reflexively and accurately, and faster than any other trait I examined. This underscores its importance as an attribution, and this resounds throughout the thesis, playing key factors in the way we ascribe cooperativeness (Chapter 4), how we form memories (Chapter 5) and how we assign – or rather, do not assign identity (Chapter 6).

A large proportion of the current published literature on voice perception has focused on female perceptions of male voices. This is undoubtedly due to an inequitable sex ratio in university psychology programmes, which provide experimenters with an abundance of female test subjects. More recently, this gap in the literature has begun to be filled, with a number of studies now focusing on female voices. However, these are still relatively small in number, and most studies only examine one gender, with researchers missing critical opportunities to examine sex differences in their effects. The importance of masculinity as an attribution is clear. However, in many of the studies recounted in this thesis, I report some notable sex differences, often finding clearer relationships for male voices than for female voices.

In Chapters 2 and 3, which focus on how stimulus content can interact with perceptions of speaker traits, I found differences in how attractiveness is perceived for women relative to men. Firstly, while the attractiveness of vowel stimuli was not correlated with the attractiveness of spontaneous speech stimuli for male voices, female voice attractiveness was correlated across all stimulus types (Chapter 2, section 2.3.2). Male vowel F_0 was negatively correlated with attractiveness ratings, but there was no relationship observed for women's voices (Chapter 2, section 2.3.3). Women's F_0-*SD* was positively correlated with attractiveness ratings, but there was no relationship observed for men's voices (Chapter 2, section 2.3.3). Finally, women were perceived as more attractive the longer they spoke, while men were perceived as less attractive with increased exposure length (Chapter 3, sections 3.2.2.3 and 3.3.2.3).

In Chapter 4, I found that multiple vocal measurements were differently related to cooperativeness dependent upon the sex of the speaker and/or the sex of the listener (section 4.2.2.1). While all listeners gave male voices with high F_0-*SD* higher cooperativeness ratings, the cooperativeness of women's voices was not driven by F_0-*SD* when they were rated by male listeners; however female listeners gave women's voices with high F_0-*SD* higher cooperativeness ratings. Additionally, female listeners gave higher cooperativeness ratings to voices with low D_f, but male listeners did not seem influenced by this trait in their ratings. Finally, women paid attention to P_f when rating the cooperativeness of men, but not of women; however, men attended to this trait when listening to voices of both sexes. I also found that men's (but not women's) dominance scores were negatively correlated with mean ratings of cooperativeness given, and women's (but not men's) age was positively correlated with how cooperative they found male (but not female) voices (section 4.2.2.2).

In Chapter 5, women had better memory for raised-F_0 male voices than men did (Experiment 1, section 5.3.1), suggesting that men may particular attention to dominant (i.e. masculine) same-sex voices. Additionally, male voice attractiveness enhanced memory for voices, while inhibiting memory for content. The reverse pattern of results was found for female voices (section 5.3.6). It may be that listeners attend to variation in vocal masculinity/femininity in different ways for male speakers than for female speakers because variation in potential mate quality is more important for women than for men.

In Chapter 6, participants were able to match male (but not female) faces and voices at rates above chance, despite no observed correlation of masculinity across the two sensory modalities (section 6.3.1). Although observers were successful at matching voices and faces of men, they did not use perceived masculinity to aid their matching; they did use masculinity differences between the distractor face stimulus and the voice stimulus to make positive matches of female stimuli, although matching accuracy was still at rates significantly below chance (section 6.3.4.3).

Taken together, these results indicate that male and female voices are perceived differently. Some of these findings could be based on the presence of a "choosiness" effect wherein variation in sexual dimorphism is more important among men than among women (Trivers & Willard, 1973; Trivers, 1972). The increased attentiveness to masculine traits in male voices relative to female voices is testament to this – in Chapter 3, this is evident in the unchanging attributions of female masculinity, while judgements of male masculinity seemed to become more refined over time. Additionally, the decrease in the attractiveness of male voices over longer time-steps suggests that women become more “choosy” when given time to reflect upon their initial quick judgements. That men find women’s voices more attractive the longer they listen to them perhaps shows the inverse of this that is also in line with Trivers’s theory, with men becoming less choosy when given time to reflect upon their judgements. This is also apparent in Chapter 5, which shows that women preferentially recall attractive male voices, but men did not preferentially remember attractive female voices. This suggests that variation in attractiveness is more important for men’s voices than women’s voices, and men listening to women may be less attuned to this variance than women listening to men.

Another explanation may also be suggested which is methodological in nature: the construct of *femininity*. Looking back, I should wish to alter the terminology in some of these studies to examine femininity in female voices rather than masculinity, as this may be a much more salient construct for those listening to female voices. In my research going forward, I will certainly do so. However, many of these sex differences cannot be simply based on the use of terminology, lending credence to the concept that male and female voices are perceived differently, and sexual dimorphism (i.e. masculinity) seems to be an important factor in these differences.

7.3 General Conclusions

The perception of social traits in voices is complex, and the work contained in this thesis confirms existing research which links perceptions of trustworthiness to lower (masculine) voice pitch. However, the relationship between masculine vocal characteristics and prosociality became less straightforward when I examined cooperativeness, which appears to depend upon a mixture of both masculine and feminine vocal traits which are not always concomitant. This makes examining these traits difficult, but avenues for future research are present and abundant, and the ability to manipulate vocal characteristics using technology is a useful tool for such an examination.

While examining sex differences may be difficult for researchers with a disproportionate amount of male and female test subjects, the sex differences in effects among both speakers and listeners throughout this thesis speak to the importance of examining these dissimilarities. Variation in masculinity/femininity may be more important for men than for women, and this may be evident in the differential effects of masculinity and femininity in voice research. Work contained in this thesis suggests that, indeed, listeners may be more attentive to variance in masculinity among male voices than in female voices, evidencing an effect of female “choosiness” in the perception of masculinity, attractiveness, and even memory. The importance of studying female voices is clear, and though it may be more challenging to recruit male test subjects, seeking out these differences is important if we are to effectively refine our understanding of voice perception.

Altogether, the experimental works in this thesis speak to the complexity of voice perception, and that vocal cues are not always what we think they are. While the perception of prosociality would seem to rely upon either perceptions of femininity or masculinity, in reality a combination of these traits may be perceived as prosocial. Furthermore, masculinity in vocal features might not be the same thing, at least developmentally, as facial masculinity. Further critical examination of these traits and their relative importance will be crucial as we continue to fill in the gaps in our understanding about voices and how they relate to person perception.

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice*, *13*(3), 424–446.

Abu-Al-makarem, A., & Petrosino, L. (2007). Reading and spontaneous speaking fundamental frequency of young Arabic men for Arabic and English languages: A comparative study. *Perceptual and Motor Skills*, *105*, 572–580.

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, *32*(3), 537–51.

Akcam, T., Bolu, E., Merati, A. L., Durmus, C., Gerek, M., & Ozkaptan, Y. (2004). Voice changes after androgen therapy for hypogonadotrophic hypogonadism. *The Laryngoscope*, *114*(9), 1587–91.

Allan, K., Jones, B. C., DeBruine, L. M., & Smith, D. S. (2012). Evidence of adaptation for mate choice within women's memory. *Evolution and Human Behavior*, *33*(3), 193–199. Elsevier B.V.

Anderson, U. S., Perea, E. F., Becker, D. V, Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., & Kenrick, D. T. (2010). I only have eyes for you: Ovulation redirects attention (but not memory) to attractive men. *Journal of Experimental Social Psychology*, *46*(5), 804–808. Elsevier Inc.

Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B., Hoffman, M., & Little, A. C. (2008). Testosterone and financial risk preferences. *Evolution and Human Behavior, 29*(6), 384–390.

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter-gatherers. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1659), 1077–82.

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters, 3*(6), 682–4.

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.

Archer, J. (1991). The influence of testosterone on human aggression. *British Journal of Psychology*, *82*, 1–28.

Archer, J., Birring, S. S., & Wu, F. C. W. (1998). The association between testosterone and aggression among young men: empirical findings and a meta-analysis. *Aggressive Behavior, 24*, 411–420.

Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Askenfelt, A., & Hammarberg, B. (1986). Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research*, *29*, 50–64.

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *The Journal of the Acoustical Society of America*, *99*(6), 3738–48.

Baken, R. J. (1987). *Clinical assessment of speech and voice*. Boston, MA: Little, Brown.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948–53.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–36.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–78.

Batliner, A., Kompe, R., Kießling, A., & Niemann, H. (1995). Can You Tell Apart Spontaneous and Read Speech if You just Look at Prosody ? In A. J. Rubio Ayuso & J. M. López Soler (Eds.), *Speech Recognition and Coding: New Advances and Trends* (pp. 321–324). Berlin: Springer.

Becker, D. V, Kenrick, D. T., Guerin, S., & Maner, J. K. (2005). Concentrating on beauty: sexual selection and sociospatial memory. *Personality & Social Psychology Bulletin*, *31*(12), 1643–52.

Beckford, N. S., Rood, S. R., Schaid, D., & Schanbacher, B. (1985). Androgen stimulation and laryngeal development. *The Annals of Otology, Rhinology, and Laryngology*, *94*(6), 634–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4073745

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–25.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–35.

Bennett, S., & Montero-Diaz, L. (1982). Children's perception of speaker sex. *Journal of Phonetics*, *10*(1), 113–121.

Berry, D. S. (1990). Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior*, *14*(3), 141–153.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11818–23.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*(4), 382–387.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. *Version 5.3.22*. Retrieved from http://www.praat.org/

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.3.56) [Software]. Retrieved from http://www.praat.org/

Booth, A., & Osgood, D. W. (1993). The influence of testosterone on deviance in adulthood: Assessing and explaining the relationship. *Criminology*, *31*(1), 93–117.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, *82*(1), 55–59. Elsevier Ltd.

Van Borsel, J., De Cuypere, G., Rubens, R., & Destaerke, B. (2000). Voice problems in female-to-male transsexuals. *International Journal of Language and Communication Disorders*, *35*(3), 427–442.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: pleasure and arousal in memory. *Journal of experimental psychology. Learning, memory, and cognition*, *18*(2), 379–90.

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of neurophysiology*, *97*(4), 3036–45. Retrieved June 4, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/17303809

Bruckert, L., Bestelmeyer, P. E. G., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., et al. (2010). Vocal attractiveness increases by averaging. *Current Biology*, *20*(2), 116–20. Elsevier Ltd.

Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciencesf America*, *273*, 83–89.

Buchanan, T. (2000). Internet research: self-monitoring and judgments of attractiveness. *Behavior Research Methods*, *32*(4), 521–527.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49. Retrieved from http://www.journals.cambridge.org/abstract_S0140525X00023992

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: an evolutionary perspective on human mating. *Psychological Review*, *100*(2), 204–32.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–43. Retrieved January 22, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/17997124

Cavelaars, A. E., Kunst, A. E., Geurts, J. J., Crialesi, R., Grötvedt, L., Helmert, U., Lahelma, E., et al. (2000). Persistent variations in average height between countries and between socio-economic groups: an overview of 10 European countries. *Annals of Human Biology*, *27*(4), 407–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10942348

Charous, S. J., Kempster, G., Manders, E., & Ristanovic, R. (2001). The effect of vagal nerve stimulation on voice. *The Laryngoscope*, *111*(11 Pt 1), 2028–31.

Chen, C. C. G., & Parker, C. R. (2004). Adrenal androgens and the immune system. *Seminars in Reproductive Medicine*, *22*(4), 369–377.

Childers, D. G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *The Journal of the Acoustical Society of America*, *90*, 1841–1856.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia Universtiy Press.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521–33.

Clarke, R., & Morton, J. (1983). Cross modality facilitation in tachistoscopic word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *35*(1), 79–96.

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *Journal of Cognitive Neuroscience*, *20*(6), 941–51. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3848031&tool=pmcentrez&rendertype=abstract

Coates, J. M., & Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 6167–72. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2329689&tool=pmcentrez&rendertype=abstract

Collins, S. A. (2000). Men’s voices and women's choices. *Animal Behaviour*, *60*(6), 773–780.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, *65*(5), 997–1004.

Cowan, M. L., & Little, A. C. (2013). The effects of relationship context and modality on ratings of funniness. *Personality and Individual Differences*, *54*(4), 496–500. Elsevier Ltd.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker’s voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*(2), 274–284.

D’Argembeau, A., Comblain, C., & Etienne, A. (2003). The effects of happy and angry expressions on identity and expression memory for unfamiliar faces. *Cognition & Emotion*, *17*(4), 609–623.

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences*, *27*, 801–804.

Dabbs, J. M., & Morris, R. (1990). Testosterone, social class, and antisocial behavior in a sample of 4,462 men. *Psychological Science*, *1*(3), 209–211. Retrieved March 9, 2012, from http://pss.sagepub.com/lookup/doi/10.1111/j.1467-9280.1990.tb00200.x

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9,* 133–50.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but... : A meta-anatytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128.

Ehrenkranz, J., Bliss, E., & Sheard, M. H. (1974). Plasma testosterone: correlation with aggressive behavior and social dominance in man. *Psychosomatic Medicine*, *36*(6), 469–75.

Ellis, A. W. (1982). Modality-specific repetition priming of auditory word recognition. *Current Psychological Research*, *2*, 123–128.

Ellis, A. W. (1989). Neuro-cognitive processing of faces and voices. In A. W. Young & H. D. Ellis (Eds.), *Handbook of Research on Face Processing* (pp. 207–215). Elsevier.

Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research*, *33*, 298–306.

Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biological Psychology*, *72*(2), 160–3.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, *93*(4-5), 783–8.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology*, *17*(2), 112–118.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior*, *29*(4), 233–241.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, *37*(4), 615–623.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., Tiddeman, B. P., et al. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *26*(5), 398–408.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., O’Connor, J. J. M., Tigue, C. C., & Borak, D. J. (2011a). Integrating fundamental and formant frequencies in women’s preferences for men's voices. *Behavioral Ecology*, *22*(6), 1320–1325.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., O’Connor, J. J. M., Tigue, C. C., & Borak, D. J. (2011b). Integrating fundamental and formant frequencies in women’s preferences for men's voices. *Behavioral Ecology*, *22*(6), 1320–1325.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G., et al. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior*, *49*(2), 215–22.

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*, 561–568.

Ferrand, C. T. (2002). Harmonics-to-noise ratio: an index of vocal aging. *Journal of Voice*, *16*(4), 480–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12512635

Fischer, J., Semple, S., Fickenscher, G., Jürgens, R., Kruse, E., Heistermann, M., & Amir, O. (2011). Do women’s voices provide cues of the likelihood of ovulation? The importance of sampling regime. *PloS one*, *6*(9), e24490.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2 Pt 1), 1213–22.

Fitch, W. T., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology and functional constraints on “honest” advertisement. *American Journal of Primatology*, *37*, 191–219.

Foa, E. B., Golboa-Schechtman, E., Amir, N., & Freshman, M. (2000). Memory Bias in Generalized Social Phobia : Remembering Negative. *Journal of Anxiety Disorders*, *14*(5), 501–519.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Folstad, I., & Karter, A. J. (2009). Parasites, bright males, and the immunocompetance handicap. *The American Naturalist*, *139*(3), 603–622.

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, *81*(1), 177–206.

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology*, *8*(3), 447–461.

Fraccaro, P. J., Jones, B. C., Vukovic, J., Smith, F. G., Watkins, C. D., Feinberg, D. R., Little, A. C., et al. (2011). Experimental evidence that women speak in a higher voice pitch to men they find attractive. *Journal of Evolutionary Psychology*, *9*(1), 57–67.

Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers’ voices on the retention of words presented visually. *Memory & Cognition*, *5*(5), 499–504.

De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289–311.

Gibson, J. M., & Bahrey, R. (2005). Modality-specificity effects in priming of visual and auditory word-fragment completion. *The Journal of General Psychology*, *132*(2), 117–37.

Goedeking, P. (1988). Vocal play behaviour in cotton-top tamarins. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate Vocal Communication* (pp. 133–144). Berlin: Springer.

Goldberg, L. R. (1999). A broad-bandwith, public-domain, personality inventory measuring the lower-level facest of several Five-Factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe, Vol. 7* (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(5), 1166–83.

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

Gray, H. (1918). *Gray's Anatomy* (20th ed., p. 1079). Philadelphia: Lea & Febiger.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: an anger superiority effect. *Journal of personality and social psychology*, *54*(6), 917–24.

Harries, M. L., Hawkins, S., Hacking, J., & Hughes, I. (1998). Changes in the male voice at puberty: vocal fold length and its relationship to the fundamental frequency of the voice. *The Journal of Laryngology and Otology*, *112*(5), 451–4.

Harries, M. L., Walker, J. M., Williams, D. M., Hawkins, S., & Hughes, I. A. (1997). Changes in the male voice at puberty. *Archives of Disease in Childhood*, *77*(5), 445–7.

Havlicek, J., Roberts, S. C., & Flegr, J. (2005). Women's preference for dominant male odour: effects of menstrual cycle and relationship status. *Biology Letters*, *1*(3), 256–9. Retrieved January 9, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1617143&tool=pmcentrez&rendertype=abstract

Helfrich, H., & Weidenbecher, P. (2011). Impact of voice pitch on text memory. *Swiss Journal of Psychology*, *70*(2), 85–93.

Henton, C. (1995). Pitch dynamism in female and male speech. *Language and Communication*, *15*(1), 43–61.

Hill, K. R., & Hurtado, A. M. (1996). *Ache Life History: The Ecology and Demography of a Forest People*. Hawthorne, NY: Aldine de Gruyter.

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, *21*(4), 406–427.

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2011). Voice correlates of mating success in men: examining “contests” versus “mate choice” modes of sexual selection. *Archives of Sexual Behavior*, *40*(3), 551–7.

Hodges-Simeon, C. R., Gurven, M., Puts, D. A., & Gaulin, S. J. C. (2014). Vocal fundamental and formant frequencies are honest signals of threat potential in peripubertal males. *Behavioral Ecology, 25*(4), 984–988.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *The Journal of the Acoustical Society of America*, *96*(5), 2646–2654.

Van Honk, J., Tuiten, A., de Haan, E., van den Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition and Emotion*, *15*(3), 279–297.

Hughes, S. M., Farley, S. D., & Rhodes, B. C. (2010). Vocal and physiological changes in response to the physical attractiveness of conversational partners. *Journal of Nonverbal Behavior, 34*(3), 155–167.

Hughes, S. M., Harrison, M. A., & Gallup, G. G. (2002). The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behavior*, *23*, 173–180.

Ishai, A. (2007). Sex, beauty and the orbitofrontal cortex. *International journal of psychophysiology*, *63*, 181–5.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition*, *12*(6), 568–74.

Jenkins, J. S. (1980). The voice of the castrato. *The Lancet*, *351*(9119), 1877–1880.

Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M., & Campanella, S. (2004). When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. *Neuroscience letters*, *369*(2), 132–7. Retrieved January 26, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/15450682

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, *43*, 280–3.

Johnston, V. S., & Franklin, M. (1993). Is beauty in the face of the beholder? *Ethology and Sociobiology*, *4*, 183–199.

Johnstone, T., van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, *1*(3), 242–9. Retrieved April 1, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1905858&tool=pmcentrez&rendertype=abstract

Jones, B. C., Feinberg, D. R., Debruine, L. M., Little, A. C., & Vukovic, J. (2008). Integrating cues of social interest and voice pitch in men’s preferences for women's voices. *Biology Letters*, *4*(2), 192–4. Retrieved March 3, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2429939&tool=pmcentrez&rendertype=abstract

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57–62. Elsevier Ltd.

Jones, B. C., Little, A. C., Boothroyd, L., Debruine, L. M., Feinberg, D. R., Law Smith, M. J., Cornwell, R. E., et al. (2005). Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the menstrual cycle when progesterone level is high. *Hormones and Behavior*, *48*(3), 283–90.

Jones, B. C., Main, J. C., DeBruine, L. M., Little, A. C., & Welling, L. L. M. (2010). Reading the look of love. Sexually dimorphic cues in opposite-sex faces influence gaze categorization. *Psychological Science*, *21*(6), 796–8. Retrieved March 6, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/20483824

Jones, B. C., Perrett, D. I., Little, A. C., Boothroyd, L. G., Cornwell, R. E., Feinberg, D. R., Tiddeman, B. P., et al. (2005). Menstrual cycle, pregnancy and oral contraceptive use alter attraction to apparent health in faces. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1561), 347–54.

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: preliminary test of a theoretical model. *The Journal of Applied Psychology*, *89*(3), 428–41. Retrieved November 7, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/15161403

Jurafsky, D., Ranganath, R., & Mcfarland, D. (2007). Extracting social meaning: Identifying interactional style in spoken conversation. *NAACL HLT*. Boulder, CO.

Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). “Putting the Face to the Voice”: Matching Identity across Modality. *Current Biology*, *13*, 1709–1714.

Kempe, V., Puts, D. A., & Cárdenas, R. A. (2013). Masculine men articulate less clearly. *Human Nature*, *24*(4), 461–75.

King, A., Ashby, J., & Nelson, C. (2001). Effects of testosterone replacement on a male professional singer. *Journal of Voice*, *15*(4), 553–557.

Kitayama, S. (1996). Remembrance of Emotional Speech: Improvement and Impairment of Incidental Verbal Memory by Emotional Voice. *Journal of Experimental Social Psychology*, *32*, 289–308.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704.

Knowles, K. K. (2010). *Perceptions of cooperation and trust in human voice pitch*. MSc Thesis. University of Liverpool: UK.

Kramer, C. (1977). Perceptions of female and male speech. *Language and Speech*, *20*(2), 151–161.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: report of Board of Scientific Affairs’ Advisory Group on the conduct of research on the internet. *The American Psychologist*, *59*(2), 105–117. Retrieved January 20, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/14992637

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *NeuroImage*, *37*(4), 1445–56. Retrieved March 3, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/17659885

De Krom, G. (1994). Consistency and reliability of voice quality ratings for different types of speech fragments. *Journal of Speech and Hearing Research*, *37*(5), 985–1000.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-11. Retrieved from http://cran.r-project.org/package=lmerTest

Kwon, H. (2010). Gender difference in speech intelligibility using speech intelligibility tests and acoustic analyses. *The Journal of Advanced Prosthodontics*, *2*(3), 71–6. Retrieved July 3, 2014, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2994697&tool=pmcentrez&rendertype=abstract

Laan, G. P. M. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*(1), 43–65.

Lachs, L., & Pisoni, D. B. (2004). Crossmodal source identification in speech perception. *Ecological Psychology*, *16*(3), 159–187.

Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, *75*(3), 817–822.

Lander, K., Hill, H., Kamachi, M., & Vatikiotis-Bateson, E. (2007). It's not what you say but the way you say it: matching faces and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 905–14.

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., Boothroyd, L. G., et al. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 135–140.

Leongómez, J. D., Binter, J., Kubicová, L., Stolařová, P., Klapilová, K., Havlíček, J., & Roberts, S. C. (2014). Vocal modulation during courtship increases proceptivity even in naive listeners. *Evolution and Human Behavior*, *35*(6), 489–496.

Levin, H., Schaffer, C. A., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech*, *25*(1), 43–54.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge: Massachusetts Institute of Technology Press.

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985). Measures of the sentence intonation of read and spontaneous speech in American English. *The Journal of the Acoustical Society of America*, *77*(2), 649–657.

Linville, S. E., & Fisher, H. B. (1985). Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *The Journal of the Acoustical Society of America*, *78*(1), 40–48. Retrieved from 10.1121/1.392452

Little, A. C., Burt, D. M., Penton-Voak, I. S., & Perrett, D. I. (2001). Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1462), 39–44.

Little, A. C., Connely, J., Feinberg, D. R., Jones, B. C., & Roberts, S. C. (2011). Human preference for masculinity differs according to context in faces, bodies, voices, and smell. *Behavioral Ecology*, *22*(4), 862–868.

Little, A. C., Jones, B. C., & DeBruine, L. M. (2008). Preferences for variation in masculinity in real male faces change across the menstrual cycle: Women prefer more masculine faces when they are more fertile. *Personality and Individual Differences*, *45*(6), 478–482.

Little, A. C., Roberts, S. C., Jones, B. C., & Debruine, L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in wartime and peacetime scenarios. *Quarterly Journal of Experimental Psychology*, *65*(10), 2018–32.

Locher, P., Unger, R., Sociedade, P., & Wahl, J. (1993). At first glance: accessibility of the physical attractiveness stereotype. *Sex Roles*, *28*(11), 729–43.

Loh, E. S. (1993). The economic effects of physical appearance. *Social Science Quarterly*, *74*(2), 420–438.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994–7. Elsevier Ltd.

Maner, J. K., Kenrick, D. T., Becker, D. V, Delton, A. W., Hofer, B., Wilbur, C. J., & Neuberg, S. L. (2003). Sexually selective cognition: beauty captures the mind of the beholder. *Journal of Personality and Social Psychology*, *85*(6), 1107–20.

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice*, *24*(5), 540–55.

Marzi, T., & Viggiano, M. P. (2010). When memory meets beauty: Insights from event-related potentials. *Biological Psychology*, *84*(2), 192–205. Elsevier B.V.

Mathes, E. W., Brennan, S. M., Haugen, P. M., & Rice, H. B. (1985). Ratings of physical attractiveness as a function of age. *The Journal of Social Psychology*, *125*(2), 157–168.

Mavica, L. W., & Barenholtz, E. (2013). Matching voice and face identity from static images. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 307–12.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences*, *21*, 353–397.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say “hello”? Personality impressions from brief novel voices. *PloS one*, *9*(3), e90779.

Mikach, S. M., & Bailey, J. M. (1999). What Distinguishes Women with Unusually High Numbers of Sex Partners? *Evolution and Human Behavior*, *150*, 141–150.

Milroy, J., Milroy, L., Hartley, S., & Walshaw, D. (1994). Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change*, *6*, 327–357.

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology*, *30*, 93–161.

Montepare, J. M., & Zebrowitz-McArthur, L. (1987). Perceptions of adults with childlike voices in two cultures. *Journal of Experimental Social Psychology*, *23*(4), 331–349.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869.

Mulac, A., & Giles, H. (1996). “You’re only as old as you sound”: Perceived vocal age and social meanings. *Health Communication*, *8*(3), 199–215.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Linguistics Compass*, *50*(1), 125–142.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097–108. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8445120

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive Memory: Remembering With a Stone-Age Brain. *Current Directions in Psychological Science*, *17*(4), 239–243.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–73.

Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1529), 2167–72.

Negus, V. E. (1949). *The Comparative Anatomy and Physiology of the Larynx*. New York: Hafner.

O’Connor, D. B., Archer, J., & Wu, F. C. W. (2004). Effects of testosterone on mood, aggression, and sexual behavior in young men: a double-blind, placebo-controlled, cross-over study. *The Journal of clinical endocrinology and metabolism*, *89*(6), 2837–45. Retrieved April 4, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/15181066

O’Connor, J. J. M., Feinberg, D. R., Fraccaro, P. J., Borak, D. J., Tigue, C. C., Re, D. E., Jones, B. C., et al. (2012). Female preferences for male vocal and facial masculinity in videos. *Ethology*, *118*(4), 321–330.

O'Connor, J. J. M., Re, D. E., & Feinberg, D. R. (2011). Voice pitch influences perceptions of sexual infidelity. *Evolutionary Psychology*, *9*(1), 64–78.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current opinion in neurobiology*, *14*(6), 769–76. Retrieved May 25, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/15582382

O'Doherty, J. P., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*(2), 147–55.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1–18.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of Voice. *Phonetica*, *41*, 1–16.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of personality and social psychology*, *80*(3), 381–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11300573

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522.

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, *5*(4), 498–502. Retrieved March 29, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/16366753

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–92.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, *85*(2), 913–925.

Parsa, V., & Jamieson, D. J. (2001). Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *Journal of Speech, Language and Hearing Research*, *44*(2), 327–339.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, *25*(4), 229–241.

Penton-Voak, I. S., Little, A. C., Jones, B. C., Burt, D. M., Tiddeman, B. P., & Perrett, D. I. (2003). Female condition influences preferences for sexual dimorphism in faces of male humans (Homo sapiens). *Journal of Comparative Psychology*, *117*(3), 264–71.

Penton-Voak, I. S., & Perrett, D. I. (2000). Female preference for male faces changes cyclically : Further evidence. *Evolution*, *21*, 39–48.

Perrett, D. I., Burt, D. M., Penton-voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and Human Facial Attractiveness. *Evolution and Human Behavior*, *20*(5), 295–307.

Perrett, D. I., Lee, K. J., Penton-Voak, I. S., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394*, 884–887.

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America*, *116*(4), 1905–1908.

Pinkham, A. E., Griffin, M., Baron, R., Sasson, N. J., & Gur, R. C. (2010). The face in the crowd effect: anger superiority when using real faces and multiple identities. *Emotion*, *10*(1), 141–6.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*, *33*(5), 509–519. Elsevier Inc. Retrieved August 21, 2012, from http://linkinghub.elsevier.com/retrieve/pii/S1090513812000050

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201–12.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *NeuroReport*, *11*(6), 1329–1333.

Power, C., Manor, O., & Li, L. (2002). Are inequalities in height underestimated by adult social position? Effects of changing social structure and height selection in a cohort study. *British Medical Journal, 325*, 131–134.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, *26*(5), 388–397.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 601–9.

Puts, D. A., Barndt, J. L., Welling, L. L. M., Dawood, K., & Burriss, R. P. (2011). Intrasexual competition among women: Vocal femininity affects perceptions of attractiveness and flirtatiousness. *Personality and Individual Differences*, *50*(1), 111–115. Elsevier Ltd. Retrieved March 27, 2012, from http://linkinghub.elsevier.com/retrieve/pii/S019188691000454X

Puts, D. A., Gaulin, S. J. C., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D:4D indicate? *Evolution and Human Behavior*, *25*(3), 182–199.

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-Simeon, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340–344.

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, *49*(2-3), 227–43. Retrieved July 20, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/22380590

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/

Ramig, L. A., Scherer, R. C., & Titze, I. R. (1984). Acoustic correlates of aging. *The Journal of the Acoustical Society of America*, *76*(S59).

Re, D. E., O'Connor, J. J. M., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PloS one*, *7*(3), e32719.

Remez, R. E., Bressel, R. S., Rubin, P. E., & Ren, N. (1987). Perceptual differentiation of spontaneous and read utterances after resynthesis with monotone fundamental frequency. *The Journal of the Acoustical Society of America*, *81*(Suppl. 1), 2–3.

Remez, R. E., Rubin, P. E., & Nygaard, L. C. (1986). On spontaneous speech and fluently spoken text: Production differences and perceptual distinctions. *The Journal of the Acoustical Society of America*, *79*(Suppl. 1), 26.

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society B: Biological Sciences*, *270*, S93–S95.

Röder, S., Fink, B., Feinberg, D. R., & Neave, N. (2013). Facial Visualizations of Women's Voices Suggest a Cross-Modality Preference for Femininity. *Evolutionary Psychology*, *11*(1), 227–237.

Rowe, R., Maughan, B., Worthman, C. M., Costello, E. J., & Angold, A. (2004). Testosterone, antisocial behavior, and social dominance in boys: pubertal development and biosocial interaction. *Biological Psychiatry*, *55*(5), 546–52. Retrieved April 4, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/15023584

Sassoon, D., Segil, N., & Kelley, D. (1986). Androgen-induced myogenesis and chondrogenesis in the larynx of Xenopus laevus. *Developmental Biology*, *113*, 135–140.

Saxton, T. K., Burriss, R. P., Murray, A. K., Rowland, H. M., & Roberts, S. C. (2009). Face, body and speech cues independently predict judgments of attractiveness. *Journal of Evolutionary Psychology*, *7*(1), 23–35.

Saxton, T. K., Caryl, P. G., & Roberts, S. C. (2006). Vocal and Facial Attractiveness Judgments of Children, Adolescents and Adults: the Ontogeny of Mate Choice. *Ethology*, *112*(12), 1179–1185.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 915–30.

Scherer, K. R. (1974). Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal Communication* (pp. 105–111). New York: Oxford University Press.

Scherer, K. R. (1978). Personality inference from voice quality: the loud voice of extroversion. *European Journal of Social Psychology, 8,* 467–487.

Scherer, K. R. (1989). Vocal Correlates of Emotional Arousal and Affective Disturbance. In H. L. Wagner & A. S. R. Manstead (Eds.), *Handbook of Social Psychophysiology* (pp. 165–197). Oxford: Wiley.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248.

Scherer, K. R. (2000a). A cross-cultural investigation of emotion inferences from voice and speech: implication for speech technology. *ICSLP* (pp. 379–382).

Scherer, K. R. (2000b). A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. *Proceedings of the International Conference on Spoken Language Processing* (pp. 379–382).

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331–346.

Scholl, B. J. (2004). Something in the way she moves. *Trends in Cognitive Sciences*, *8*(2), 49–51.

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language and Hearing Research*, *40*(2), 453–63.

Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *Quarterly Journal of Experimental Psychology*, *60*(10), 1446–56.

Silveira, A. M., Fishman, L. S., Subtelny, J. D., & Kassebaum, D. K. (1992). Facial growth during adolescence in early, average and late maturers. *The Angle Orthodontist*, *62*(3), 185–190.

Silventoinen, K., Lahelma, E., & Rahkonen, O. (1999). Social background, adult body-height and health. *International Journal of Epidemiology*, *28*, 911–918.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640.

Smith, D. S., Jones, B. C., Feinberg, D. R., & Allan, K. (2012). A modulatory effect of male voice pitch on long-term memory in women: evidence of adaptation for mate choice? *Memory & Cognition*, *40*(1), 135–44.

Stanton, S. J., Liening, S. H., & Schultheiss, O. C. (2011). Testosterone is positively associated with risk taking in the Iowa Gambling Task. *Hormones and Behavior*, *59*(2), 252–6. Elsevier Inc.

Streeter, L. A., MacDonald, N. H., Apple, W., M, K. R., & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *The Journal of the Acoustical Society of America*, *73*(4), 1354–1360.

Studer, L. H., Aylwin, A. S., & Reddon, J. R. (2005). Testosterone, sexual offense recidivism, and treatment effect among adult male sex offenders. *Sexual Abuse*, *17*(2), 171–81.

Swaddle, J. P., & Reierson, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proceedings of the Royal Society B: Biological Sciences*, *269*, 2285–9.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*(12), 452–460.

Thornhill, R., & Grammer, K. (1999). The Body and Face of Woman: One Ornament that Signals Quality? *Evolution and Human Behavior*, *20*, 105–120.

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210–216. Elsevier B.V.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social Cognitive and Affective Neuroscience*, *3*(2), 119–27. Retrieved December 11, 2013, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2555464&tool=pmcentrez&rendertype=abstract

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences*, *12*(12), 455–60. Retrieved December 11, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/18951830

Trainor, L. J., Austin, C. M., & Desjardins, N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97–100.

Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of F0-excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, *97*(3), 1905–1915.

Trivers, R. L. (1972). Parental Investement and Sexual Selection. In B. Campbell (Ed.), *Sexual Selection and the Descent of Man, 1871-1971* (pp. 136–179). Chicago, IL: Aldine.

Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, *179*(4068), 90–92.

Tsukiura, T., & Cabeza, R. (2011). Remembering beauty: Roles of orbitofrontal and hippocampal regions in successful memory encoding of attractive faces. *Neuroimage*, *54*(1), 653–660.

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech*, *3*, 223–34.

Vaissière, J. (2005). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 236–263). Oxford: Blackwell.

Vukovic, J., Feinberg, D. R., Jones, B. C., DeBruine, L. M., Welling, L. L. M., Little, A. C., & Smith, F. G. (2008). Self-rated attractiveness predicts individual differences in women's preferences for masculine men's voices. *Personality and Individual Differences*, *45*(6), 451–456.

Vukovic, J., Jones, B. C., DeBruine, L. M., Feinberg, D. R., Smith, F. G., Little, A. C., Welling, L. L. M., et al. (2010). Women's own voice pitch predicts their preferences for masculinity in men's voices. *Behavioral Ecology*, *21*(4), 767–772.

Vukovic, J., Jones, B. C., Feinberg, D. R., Debruine, L. M., Smith, F. G., Welling, L. L. M., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, *102*(1), 37–48.

Wadnerkar, M. B., Cowell, P. E., & Whiteside, S. P. (2006). Speech across the menstrual cycle: a replication and extension study. *Neuroscience Letters*, *408*(1), 21–4.

Watkins, C. D., Fraccaro, P. J., Smith, F. G., Vukovic, J., Feinberg, D. R., DeBruine, L. M., & Jones, B. C. (2010). Taller men are less sensitive to cues of dominance in other men. *Behavioral Ecology*, *21*(5), 943–947.

Watkins, C. D., Jones, B. C., & DeBruine, L. M. (2010). Individual differences in dominance perception: Dominant men are less sensitive to facial cues of male dominance. *Personality and Individual Differences*, *49*(8), 967–971. Elsevier Ltd. Retrieved November 9, 2012, from http://linkinghub.elsevier.com/retrieve/pii/S019188691000396X

Wheatley, J. R., Apicella, C. A., Burriss, R. P., Cárdenas, R. A., Bailey, D. H., Welling, L. L. M., & Puts, D. A. (2014). Women's faces and voices are cues to reproductive potential in industrial and forager societies. *Evolution and Human Behavior*, *35*(4), 264–271. Elsevier Inc.

Whiteside, S. P., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time : a developmental study of phonetic context effects in British English. *The Journal of the Acoustical Society of America*, *116*(2), 1179–1183.

Whiteside, S. P., & Marshall, J. (2001). Developmental trends in voice onset time: some evidence for sex differences. *Phonetica*, *58*(3), 196–210.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, *52*(4B), 1238–1250.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–8. Retrieved March 31, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/16866745

Wilson, M., & Daly, M. (2004). Do pretty women inspire men to discount the future? *Proceedings of the Royal Society B: Biological Sciences*, *271 Suppl*, S177–9. Retrieved March 3, 2012, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1810021&tool=pmcentrez&rendertype=abstract

Winston, J. S., O'Doherty, J. P., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, *45*(1), 195–206.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–83.

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, *64*(10), 1673–1683.

Zebrowitz-McArthur, L., & Montepare, J. M. (1989). Contributions of a babyface and a childlike voice to impressions of moving and talking faces. *Journal of Nonverbal Behavior*, *13*(3), 189–203.

Zraick, R. I., Wendel, K., & Smith-Olinde, L. (2005). The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *Journal of Voice*, *19*(4), 574–81. Retrieved July 24, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/16301103

Zuckerman, M., Miyake, K., & Elkin, C. S. (1995). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality*, *29*, 253–272.