SOCIAL PERCEPTION OF THE HUMAN VOICE: PERCEIVER ATTUNEMENT TO THE VOCAL SPECIFICATION OF SPEAKER PHYSICAL CHARACTERISTICS

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

at the

University of Canterbury

by

Brad H. Miles

University of Canterbury

2012

Acknowledgements

I would first like to thank my primary supervisor, Lucy Johnston, whose support, advice, and knowledge has been immensely appreciated. I also gratefully acknowledge the contributions of my secondary supervisors, Brian Haig and the late Denis Dutton. It has been a pleasure and a privilege to work alongside these three.

Special mention must also go to Alex Mackenzie and Rachel Bryant, whose assistance, particularly with the data collection phase of the research, has been invaluable.

Thank you to all the friends and family who contributed in various ways over the past few years, especially Sheryll Miles and John Miles, whose support has been beyond anything that could reasonably be expected.

I would also like to acknowledge all the members and visitors to the Social Perception Lab who have contributed to making it a stimulating and enjoyable place to work, and all the participants who volunteered their time and voices.

Finally, thank you to Lucy Catherwood – the best person I know.

Table of Contents

List of Tables

List of Figures

Abstract

The human voice is a common and important part of the social environment. In addition to being the primary carrier of language, there is growing evidence that the sound of a person's voice contains a great deal of socially relevant information. Drawing on a functional approach to perception, the current research investigated the attunement of social perceivers to the vocal specification of speaker physical properties. An initial study developed a set of vocal samples for use in subsequent perceptual studies, and conducted exploratory analyses investigating relationships between speaker vocal and physical characteristics. Significant differences in the acoustic properties of male and female voices were identified, but the relationships between acoustic properties and speaker age, body size, and body configuration were less robust. Study 2 investigated the ability of listeners to accurately perceive the physical characteristics of speakers from vocal information. Perceivers made assessments of speaker physical characteristics that were highly consensual and that accurately reflected speaker sex, age, and body size. Studies 3 and 4 investigated perceiver judgments of vocal attractiveness. In Study 3, both male and female perceivers rated the voices of male speakers with lower indices of body asymmetry (a marker of genotypic and phenotypic condition) as more attractive. However, for female perceivers it was shown that this relationship is influenced by changes in fertility levels associated with the menstrual cycle. At times of high fertility female perceivers displayed a stronger attraction to the voices of male speakers with low asymmetry than they did at times of low fertility. This finding was interpreted as a functional shift increasing attraction to males possessing phenotypic markers of high fitness when the likelihood of conception is highest. Study 4 considered the effects of menstrual cycle variation on the voices of female speakers. Both male and female perceivers rated female voices recorded during a phase of high fertility to be more attractive than the same voices recorded during a phase of low fertility. This finding extends previous research demonstrating cyclic shifts in visual and olfactory attractiveness to the auditory domain, and is discussed in terms of the vocal specificity of female fertility status. Study 5 extended the previous studies by considering how vocal cues specifying the sex and age of a social target interact with visual cues to influence social perception. Relative to concordant voice and face information, discordant information was found to facilitate social memory. This finding is discussed with regards to the integration of multiple sources of information in social perception. The results of all studies are discussed in terms of the adaptive significance of perceivers accurately detecting the physical characteristics of others that are specified vocally.

Chapter 1

Introduction

General Overview

The human voice is a highly sophisticated acoustic instrument. Capable of producing a wide variety of pitches, volumes, and timbres, voices can broadcast information to wide audiences over appreciable distances, or whisper information to close neighbours in relative privacy. While voices are typically associated with language as a primary mode of communication, much of the acoustic variation in human speech does not contribute to linguistic understanding. Such non-linguistic variation may nevertheless reflect important attributes of a speaker that perceivers can utilise as a basis for social judgments and decision-making.

Consider for example, a radio broadcast in which the announcer is speaking in an unfamiliar language. Despite a lack of both linguistic comprehension and visual information, listeners may still be able to perceive socially relevant characteristics of the speaker – such as whether they are male or female, or whether they are an adult or a child – simply from the sound of the speaker's voice. It is the perception of such fundamental features of a speaker that is the subject of this thesis. The central hypothesis is that independent of linguistic and affective content, human voices reliably reveal important information about speakers, and that listeners are able to use this information to guide social behaviour. Accordingly, this research investigates the nature of the vocal information that co-varies with other features of speakers, the assessment and interaction opportunities such information affords social perceivers, and the extent to which perceivers are attuned to such information.

The thesis is broadly structured into three main sections. The first section provides a general introduction and background to viewing the voice as a source of social information. The physiological and neural underpinnings of vocal production and perception are briefly considered, and a framework for understanding how vocal signals can encode information regarding socially relevant features of a speaker is presented. Previous research investigating the judgements perceivers make on the basis on vocal information is considered, and it is argued that voices can be viewed as salient indicators of socially relevant traits such as sex, age, physical condition, and attractiveness. Integrating these various strands, a number of specific hypotheses concerning the social perception of human voices are then developed.

The second section comprises the empirical components of the thesis. An initial study describes the development of a set of vocal samples and presents the results of exploratory analyses examining the relationships between speaker vocal characteristics and phenotypic traits such as sex, age, and body configuration. Study 2 is the first of a series of perceptual studies, and reports on the ability of perceivers to utilise vocal information to make reliable and valid judgments about the physical characteristics of speakers. Studies 3 and 4 investigate the relationships between speakers' vocal characteristics and evaluations of attractiveness, with particular emphasis on the extent to which evaluations of attractiveness are influenced by hormonal variation in both speakers and perceivers. Studies 5a and 5b investigate the interaction between vocal and visual information in social perception, and examine how that interaction influences social memory.

The final section provides a general discussion of the main empirical findings. The results are discussed in terms of perceivers' abilities to use vocal information as a basis for making judgments about speakers and guiding social interaction.

Voices as a Source of Social Information

Voices are a common and important part of the human social environment. It has been suggested for example, that people spend more time attending to voices than to any other type of sound (Belin, Zatorre, & Ahad, 2002). Not only are voices a carrier of language, and therefore of the direct semantic information that language contains, they also convey a host of information over and above the literal meaning of the speech content (Caffi & Janney, 1994). This is perhaps best illustrated by the phrase “it’s not what she said, it’s how she said it,” in which the tonal qualities of a speaker’s voice contribute meaning additional to the semantic content of the words.

Traditionally, the judgments perceivers make about a speaker from vocal cues have not been a major focus of linguistic or speech perception research. While all human speech displays the same basic organization, characteristic vocal patterning is determined by inter- and intra-individual variation in acoustic parameters. Despite this variation, speech is typically perceived in a stable manner. Accordingly, a primary goal of speech perception research has been to explain how listeners achieve perceptual constancy from varying input. Indeed, most theories of speech perception take the acoustic variation between speakers as noise that is to be filtered out through a process of vocal normalization (Goldinger, 1996; Johnson, 2005). In other words, in order to comprehend speech, listeners must determine which aspects of the acoustic signal are linguistically meaningful and which aspects are not. Some of the non-linguistic variation in a signal may be noise, but some of it is likely to reflect attributes of a vocaliser that perceivers can utilise to help guide socially adaptive behaviour. For example, various accents and dialects are differentiated by systematic variation in the enunciation of particular phonemes. Dialects are associated with speech communities, often reflecting a speaker’s regional origins and/or socioeconomic status, and can affect the way listeners perceive a speaker’s behaviour (Giles & Powsland, 1975). Moreover, voices are influenced by a speaker’s affective state. Specific patterns of autonomic activity and muscular action corresponding to various emotional states produce changes in the acoustic properties of a speaker’s voice. While vocal information regarding affective state is often found in the form of prosodic speech, non-speech sounds such as screams, groans, cries, and laughs, also contain a great deal of affective information. Given that different emotional states are likely to afford different behavioural opportunities, awareness of the emotional state of a speaker can help perceivers guide their behavioural interactions with that speaker in a functional manner. Indeed, vocal cues of affect may be the most common method of inferring other people's emotions (Planalp, 1998).

In addition to relatively dynamic and time-varying information such as speech and affect, voices may also carry important cues relative to more invariant features of a speaker. Individuals are able to be recognized and differentiated on the basis of unique vocal features that are independent of the phoneme being produced. That is, individuals can be said to possess a unique "vocal signature." Even after long periods apart perceivers can recognize familiar individuals from vocal information with considerable accuracy (Papcun, Kreiman, & Davis, 1989), and prior to the ability to understand speech, infants show a capacity to discriminate between speakers and to recognise the voices of their primary care-givers, possibly even before birth (DeCasper & Fifer, 1980; Kisilevsky et al., 2003; Ockleford, Vince, Layton, & Reader, 1988).

In a similar fashion, voices reveal information about fundamental and relatively stable aspects of a speaker's makeup. Males sound characteristically different to females, and children sound characteristically different to adults, so voices can reveal information about a speaker's sex and age (Bachorowski & Owren, 1995; Ladefoged & Broadbent, 1957). Vocal parameters also co-vary with judgments of a range of physical characteristics such as body size and configuration (Evans, Neave, & Wakelin, 2006; Hughes, Harrison, & Gallup, 2009), physical strength and fighting ability (Sell et al., 2010), and physical attractiveness (Collins & Missing, 2003), as well as various features of psychology and behaviour such as personality (Scherer, 1972), intelligence (Reynolds & Gifford, 2001), sexual orientation (Linville, 1998), and possibly even dispositional states relevant to attempts at deception (Ekman, Friesen, & Scherer, 1976; Streeter, Krauss, Geller, Olson, & Apple, 1977).

Voices are also likely to provide information about a speaker's current physiological state. Increases in muscle tension are associated with changes in vocal parameters (Titze, 1994), and the vocal folds are regulated by the vagus nerve, which carries information from the viscera and is involved in the regulation of heart-rate, blood pressure, and other aspects of what is commonly termed the fight or flight system (Berthoud & Neuhuber, 2000). As such, changes in the maintenance of fine motor control, the state of the muscular system, and the state of the parasympathetic/sympathetic system, may all be revealed vocally. Vocal information may also be useful in identifying various disorders, including Parkinson's disease, essential tremor, major depressive disorder, and autism, as well as identifying the presence and effects of drug and alcohol use (Cannizaro, Harel, Reilly, Chappell, & Snyder, 2004; Gamboa, Jimenez-Jimenez, Mate, & Cobeta, 2001; Oller et al., 2010; Sapir, Ramig, Spielman, & Fox; Scherer & Zei, 1988). Because stress, drugs, illness, and the like can all alter vocal characteristics in a manner that can be hard to control volitionally, voices may convey valid information pertaining to a speaker's current mental and physical condition.

The above considerations show that voices are a potentially rich source of social information. In this sense voices can be viewed in a comparable fashion to the human face. Although they involve different physical mediums, the information content of the visual signals conveyed by the face and the acoustic signals conveyed by the voice both contain similar types of socially relevant information relating to language, affect, identity, and both mental and physical condition (Belin, Bestelmeyer, Latinus & Wilson, 2011). Nevertheless, while a considerable amount of research has examined face perception in a social context, relatively little research has examined voice perception (Belin et al., 2011; Bliss-Moreau, Owren, Barrett, 2010; Ko, Judd & Blair, 2006). Indeed face perception has blossomed into a major research topic encompassing a number of disciplines and addressing a range of research issues (Zebrowitz, 2006). For example, research on face perception includes investigation of the attributes that are perceived from faces (e.g., Miles & Johnston, 2007), the nature of the information that specifies the attributes (e.g., Carré, McCormack, & Mondloch, 2009), individual differences in abilities to perceive attributes (e.g., Davis et al., 2011), the developmental trajectories of such abilities (e.g., Mondloch, 2012), the underlying neural mechanisms (e.g., Kanwisher & Yovel, 2006), and the behavioural consequences of face perception (e.g., Stirrat & Perrett, 2010). Although a fully comprehensive account of face perception remains to be achieved, current understanding of voice perception is, by comparison, in its infancy.

The present research aims to contribute to the small, but growing, body of work examining the social perception of human voices. It is argued that independent of linguistic and affective information, human voices convey reliable and valid information regarding socially relevant features of a speaker, and that listeners are perceptive to such information. In this sense, voice perception is considered in a functional capacity – that of guiding adaptive social behaviour. This functional perspective suggests that social perceivers should be attuned to the properties of others for which the correct (or incorrect) detection of would impact on behavioural opportunities relevant to reproduction and survival (e.g., Gangestad, Simpson, DiGeronimo, & Biek, 1992). In essence, adaptive social behaviour requires the accurate detection of information specifying the properties of others. Some of the most adaptively relevant properties of others are related to fundamental physical characteristics. Different physical characteristics afford different behavioural opportunities, and correct detection of these characteristics is likely to proffer considerable advantage to social perceivers. That is, knowledge of the physical characteristics of conspecifics allows perceivers to direct their behaviour toward adaptive interactions with those conspecifics and away from mal-adaptive or harmful interactions. For example, knowing that an individual is very young may direct behaviour toward nurturing, knowing that an individual is physically large may direct behaviour toward submissiveness, and knowing that an individual is of a particular sex may direct behaviour toward courtship. In light of these considerations, the specific focus of the current research is the ability of perceivers to use vocal information to accurately detect fundamental, and socially relevant, biological and physical features of speakers such as sex, age, physical stature, and attractiveness. A simple but useful analogy can be made to listening to a piece of music and asking what instrument is being played. Regardless of the particular piece, a violin sounds characteristically different to a double bass, and listeners are typically perceptive to the difference. That is, rather than considering what is being said, or how it is being said, the actual physical qualities of voices may reveal important information about speakers. As such, the current research investigates voice perception across a chain of analysis, from the nature of the acoustic information that specifies the physical characteristics of a speaker, to examination of listener abilities to perceive speaker characteristics from vocal information, and to the behavioural consequences of voice perception.

The remainder of this chapter considers how information about behaviourally relevant physical features of a speaker can be reliably encoded in an acoustic signal, and examines previous research addressing the types of judgments perceivers make about speakers from vocal information. Subsequently, drawing on theories of sexual selection and biological signalling, an orienting account of the functional nature of vocal signals is presented, and several specific hypotheses concerning the social perception of voices are developed.

The Utility of Voices as Distinct from Language

While voices are typically associated with language in the form of speech, it is important in the present context to distinguish between the two. Language, as a system for representing and communicating semantic information, can be viewed independently of the mode of communication. For example, language could theoretically have been first encoded gesturally rather than vocally (Corballis, 1992). A lack of empirical data has meant that understanding the evolution of language has proven both difficult and controversial (Tallerman & Gibson, 2011), with one of the few certainties being that language must have originated sometime between the divergence of modern humans and chimpanzees approximately six-million years ago.

Although spoken language may have appeared relatively recently in human evolution, non-linguistic vocal sounds are likely to have figured prominently in the environments of our vertebrate ancestors, and accurately perceiving the information contained in vocal sounds, whether from conspecifics, predators, or prey, would have been of considerable selective advantage (Hauser, 1996). One of the key features of vocal signals is that both their production and perception are independent of light (Gallup & Cameron, 1992). Consequently, voices may provide a particularly important source of social information in situations where visual information is impoverished or absent. Sell et al. (2010) have argued that evolutionary recurrent conditions in which visual assessment of individuals would have been difficult (e.g., distance, darkness, intervening obstructions) would likely have driven the selection of systems specialized for social perception in the auditory domain. Formal models have suggested that vocal signals are likely to provide honest information about a vocaliser (Johnstone, 1995), and studies have demonstrated that numerous species use vocal information to assess the characteristics of a vocaliser (e.g., Fitch; 1997; Reby & McComb, 2003; Ryan & Keddy-Hector, 1992).

As illustrated in Figure 1, Taylor and Reby (2010) have proposed an evolutionary feedback loop linking the production of vocal signals to their acoustic structure, and the perception and subsequent behaviour of perceivers.

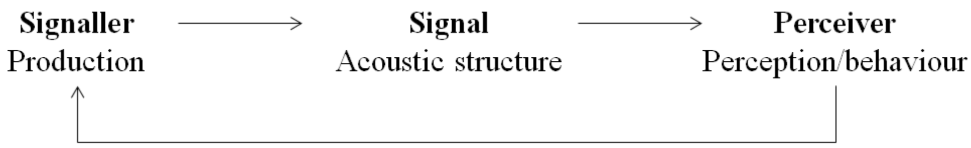

Figure 1. Evolutionary feedback loop linking the production, structure, and functional perception of vocal signals. Adapted from Taylor and Reby (2010).

According to this model, the operation of vocal production mechanisms determines the structure of an acoustic signal. Information about the signaller is encoded in the signal and it is this information that affords perceivers behavioural opportunities relative to the signaller. The behaviour of perceivers then drives selection of the signal at the level of its production. It is this level – the production level – at which the information content of vocalizations, and hence the information that is available to perceivers is determined. Accordingly, understanding the production of vocal signals, and the information that they may contain, is a critical first step in elucidating the processes of social perception in the auditory domain.

The Source-Filter Theory of Vocal Production

The dominant framework for understanding voiced acoustic signals is the source-filter theory (Fant, 1960; Titze, 1994). This approach suggests that vocal signals originate from two, essentially independent, aspects of the vocal apparatus: the “source,” and the “filter.” The source is comprised of the larynx (including several sub-laryngeal structures), while the filter consists of the supra-laryngeal vocal tract – the passage connecting the larynx to the openings of the mouth and nose. Application of the source-filter theory has allowed the development of testable hypotheses linking variation in the acoustic structure of vocal signals to their means of production and their functional perception. What follows is a brief introduction to source-filter theory with particular emphasis on the production of specific acoustic parameters and consideration of the information they may encode.

The production of voiced sound originates in the larynx, with the opening and closing of the vocal folds. The vocal folds themselves consist of bilateral membranes stretching across the larynx, and together with the spacing between them they form the glottis (Titze, 1994). Air expelled from the lungs and through the glottis forces the vocal folds apart, before biomechanical actions result in the vocal folds closing again (Chan & Titze, 2006). This pattern of opening and closing produces a cyclic variation in air pressure that constitutes the glottal waveform or source signal. The rate of oscillation of the vocal folds determines the fundamental frequency (F_0) of the signal. F_0, the acoustic correlate of what is perceived as pitch, is influenced by the size and tension of the vocal folds; longer, thicker vocal folds vibrate at a slower rate than smaller vocal folds, and consequently produce signals of lower F_0 (Titze, 1994). Additional characteristics of the source signal such as duration and amplitude result from changes in air flow and sub-glottal pressure that are typically under voluntary muscular control (Titze, 1994).

The second component of voiced sound arises from the filtering process of the supra-laryngeal vocal tract. The vocal tract is comprised of the tube, structures, and air cavities that link the larynx to the openings of the mouth and nose, from which sound is radiated into the environment. Resonant properties of the vocal tract selectively augment or attenuate specific frequencies of the source signal, producing spectral peaks known as resonant or formant frequencies (Fant, 1960). Manipulations of the vocal tract, such as those produced by movements of the tongue and lower jaw, affect the structure of formant frequencies, with modulation of the lower formants giving rise to the phonemes that constitute the vowels of human speech. In contrast to other primates, where the larynx is positioned high enough in the throat to enable simultaneous swallowing and breathing, the human larynx is situated much lower (Negus, 1949). While infants are born with a high positioned larynx, and can suckle and breathe simultaneously, beginning at around three months of age the larynx undergoes a gradual descent, reaching its lower adult position at around three to four years of age, with a further descent occurring in males at puberty (Fitch & Giedd, 1999; Sasaki, Levine, Laitman, & Crelin, 1997). This descended larynx allows for greater flexibility in modulating the shape of the vocal tract. For example, the human tongue can move both horizontally and vertically in the vocal tract, enabling the production of vowel sounds such as /i/ and /u/ that other mammals typically cannot produce. Such flexibility in configuring the vocal tract enables the production of a wide range of highly discriminable formant patterns, and is a critical component of speech. The importance of formant frequencies is perhaps best illustrated by the case of whispered speech. Whispering involves the generation of a source signal without vibration of the vocal folds, but with vocal tract articulation that remains the same as in normal speech (Tartter & Braun, 1994). As such, the formants are present, and whispered speech is readily intelligible.

Fitch and Hauser (2002) have employed the source-filter theory to outline how vocal signals can provide reliable and valid information regarding various characteristics of a vocalizing individual. The general premise of this "honest signalling" hypothesis is that the acoustic structure of vocal signals is constrained by the physical mechanisms that produce them. For example, as outlined above, F_0 is critically dependent on the physical characteristics of the vocal folds. Development of the vocal folds is under hormonal influence with estrogen and progesterone influencing female vocal development (Abitbol, Abitbol, & Abitbol, 1999) and testosterone influencing male vocal development (Harries, Hawkins, Hacking, & Hughes, 1998). Increasing testosterone levels at puberty produces a lengthening and thickening of male vocal folds such that male F_0 decreases around 50% in comparison to female F_0 (Fitch & Giedd, 1999; Harries et al., 1998). In males, F_0 has been found to co-vary with circulating levels of testosterone (Dabbs & Mallinger, 1999; Evans, Neave, Wakelin, & Hamilton, 2008) and to decrease with exogenous testosterone treatment (Need, Durbridge, & Nordin, 1993), while in females, hormonal changes associated with puberty and menopause influence F_0 production (Abitbol, Abitbol, & Abitbol, 1999; Caruso et al., 2000), and natural hormonal variation across the menstrual cycle has been associated with changes in F_0 (Bryant & Haselton, 2009). Thus, F_0 appears to co-vary with sex, age, and hormonal profile.

Unlike the vocal folds, which are able to develop largely independent of the rest of the body, development of the vocal tract is closely tied to the development of skeletal structures (Fitch, 2000). This relationship means that formant frequencies may provide information about the body size of a vocalizer. Rather than the actual formants themselves, it is the overall relationships between formants that may provide the most legitimate cue to body size. Typically, this relationship is considered as "formant dispersion" (F_d), and is defined as the average distance between successive formants (Fitch, 1997); however, Puts, Apicella, and Cárdenas (2011) have proposed an alternative formulation termed "formant position" (F_p), which they defined as the average standardized formant frequency. Perceivers are sensitive to formant shifts as small as 7% (Rendall, Vokey, & Nemeth, 2007), and a number of studies have reported relationships between formant structure and body size (Collins & Missing, 2003; Evans et al., 2006; Gonzalez, 2004; Rendall et al., 2005; Sell et al., 2010).

The above considerations show that application of the source-filter theory allows for the acoustic structure of vocal signals to be understood in terms of their means of production. Importantly, this approach suggests that anatomical and physiological attributes of speakers – including sex, age, body size, and hormonal profile – reliably and predictably produce acoustic variation in vocal signals in a manner that is functionally available to perceivers.

Specializations for the Processing of Vocal Acoustic Information

The information content of voice signals is such that there are likely to be neuropsychological specializations for the processing of vocal information (Ghazanfar & Hauser, 1999; Hauser, 1996). Although relatively little is known about the neuropsychology of extra-linguistic voice processing, two lines of evidence suggest that vocal sound is selectively processed. Neuroimaging studies provide evidence for specialized neural mechanisms selective to the sound of voices. Discrete regions of the auditory cortex, specifically regions located along the mid and anterior parts of the superior temporal sulcus, show increased activation to vocal sounds relative to non-vocal sounds (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). While this response is particularly strong for speech sounds, studies have reported evidence for the processing of voice information as being distinct from the processing of speech information. Middle regions of the superior temporal sulcus, particularly in the right hemisphere, appear more responsive simply to the sound of voices rather than to the meaning of speech, showing for example, increased activation in response to backwards played speech but not to understandable and semantically meaningful modulated noise (Belin, Fecteau, & Bedard, 2004). That such voice selective responses have been observed in infants as young as 7-months of age, after the development of vocal discrimination abilities but before the full development of speech processing abilities, suggests that such regions are not exclusively concerned with speech processing (Grossman, Oberecker, Koch, & Friederici, 2010). Moreover, claims of homologous voice selective neural regions in non-human primates such as macaques suggest that areas specialized for the processing of vocal sounds are phylogenetically ancient and likely to have been present at least as far back as the common ancestor of humans and macaques around 30 million years ago, and well before the evolution of speech and language (Petkov et al., 2008).

Studies asking participants to listen to speech and make decisions either about the identity of the speakers or about the content of the speech have demonstrated increased activation in the right anterior STS and the anterior temporal lobes during identity related tasks relative to content related tasks (Imaimuzi et al., 1997; Nakamura et al., 2001; Von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). Commensurate with this, Belin and Zatorre (2003) found decreased activity in regions of the right anterior superior temporal sulcus during trials in which participants listened to different syllables being spoken by a single speaker relative to trials in which the same syllable was spoken by different speakers. This reduced activity was interpreted as an adaptation by neural regions sensitive to the idiosyncratic acoustic characteristics of a speaker's voice.

A second line of evidence comes from clinical populations. Cases of impaired recognition of familiar speakers and/or impaired discrimination between unfamiliar speakers have been documented in studies of patients with brain lesions (Van Lancker & Kreiman, 1987; Peretz et a1., 1994; Neuner & Schweinberger, 2000). Importantly, these impairments appear dissociable both from each other and from speech perception deficits. Cases of impaired speech perception with preserved voice recognition, and cases of impaired voice recognition with preserved speech perception demonstrate a double dissociation whereby vocal indexical features and speech recognition appear to be processed in partially dissociable neural regions (Belin, et al., 2004). Taken together, the above lines of research provide converging evidence for the existence of neural mechanisms dedicated to the processing of non-linguistic vocal information.

Social Perception of the Voice

The preceding sections demonstrate that independent of speech and affect, human voices contain socially relevant information about speakers. Specifically, anatomical and physiological constraints on vocal production may produce acoustic signals that contain reliable and valid information about a vocaliser's sex, age, body size, and hormonal profile. Moreover, perceivers appear to have neuro-specialisations for the processing of non-linguistic vocal information. This section briefly reviews the growing body of research suggesting that perceivers readily utilise non-linguistic vocal cues to make judgments about a range of speaker physical attributes.[1]

Early research into the social perception of human voices took as a starting point the judgments that listeners make about speakers from radio broadcasts. Pear (1931, cited in Allport & Cantril, 1934), examined descriptions provided by over 4000 listeners of several speakers whose voices were played over the radio. For the most part, judgments regarding speaker sex, age, and physical appearance proved reasonably accurate. Drawing on an earlier distinction made by Sapir (1927) between voice and speech, Allport and Cantril (1934) were among the first researchers to systematically investigate the judgments that perceivers make from voices independent of language. Across a series of studies in which the linguistic content of utterances was held constant by having speakers recite a standard passage, attempts were made to assess the extent to which perceivers could accurately match a range of personal attributes (e.g., age, height, personality, vocation) with particular voices. Two general findings emerged. Firstly, while there was substantial variation, the majority of matchings were above chance (with the notable exception of assessments concerning speaker height, which were largely inaccurate, although it was suggested that height assessments may have been skewed by one particularly idiosyncratic speaker), indicating a moderate degree of validity in perceiver judgments. Secondly, perceiver judgments displayed a high degree of uniformity – even when those judgments were wrong. That is, regardless of whether they

[1] This is intended only as a broad overview of relevant literature. Fuller consideration of previous research examining the judgments that perceivers make from voice information is presented in the context of the relevant empirical sections of the thesis, primarily chapters 3, 4, and 5.

were accurate or not, perceiver judgments were found to be highly similar. This raises an important consideration for the current research, that of the distinction between consensus and accuracy. While different perceivers can agree in their judgments (i.e., the judgments show consensus), those judgments may not necessarily be accurate. For example, two parents can agree that their child will win an Olympic medal, but the child may never actually do so. Historically, much of the research in social perception that has focused on consensus has done so possibly as a proxy for accuracy in cases where there is a lack of criteria by which to evaluate accuracy (Kenny, 1991). However, in the present case, the objective measurement of speaker characteristics provides a clear criterion by which to assess the accuracy of perceptions of those characteristics. As such, consensus and accuracy represent two different ways by which to evaluate perceiver judgments.

Nearly half a century after Allport and Cantril, a series of studies by Lass and colleagues (e.g., Lass, Beverly, Nicosia, & Simpson, 1977; Lass, Brong, Ciccolella, Walters, & Maxwell, 1980; Lass & Colt, 1980; Lass & Davis, 1976; Lass et al., 1979; Lass, Hughes, Bowyer, Waters, & Bourne, 1976) produced similar results concerning both the accuracy and consensus of perceiver evaluations. Investigating perceiver judgments of speaker characteristics such as sex, age, height, and weight from vocal cues, it was found that across perceivers, judgments were both highly consistent, and reflected the true characteristics of the speakers. However, many of these studies were subject to methodological criticisms (e.g., Cohen, Crystal, House, & Neuberg, 1980) and subsequent re-evaluation of much of the data found that while perceiver judgments were highly consistent, judgments of speaker height and weight were largely unrelated to the actual values (e.g., van Dommelen, 1993; Gonzalez, 2003).

Nevertheless, some research has found that perceivers are indeed able to accurately evaluate a number of physical traits from vocal information. Krauss, Freyberg, and Morsella (2002) found that after hearing a speaker's voice, perceivers were able to select the speaker from a pair of photographs with over 75% accuracy. Krauss et al. argued that perceivers achieved this by first estimating a speaker's physical characteristics. That is, perceivers were believed to use vocal characteristics in order to determine a speaker's age, height, and weight, and then use these estimates to select the photograph that most closely aligned with them. In a second study, Krauss et al. (2002) found support for this notion by showing that perceivers could accurately determine a speaker's sex, and make age, height, and weight estimates that both closely matched the true values of speakers, and that were comparable in accuracy to estimates made from photographs. More recent research has shown that in addition to judgments of overall body size, perceivers may also be able to use vocal information to make judgments about body configuration. Hughes, Harrison, and Gallup (2009) found that perceiver judgments of speaker waist-to-hip and shoulder-to-hip ratios accurately reflected the actual waist-to-hip ratios of female speakers and the actual shoulder-to-hip ratios of male speakers, while Sell et al. (2010) found that perceivers could utilise vocal cues to accurately assess the upper-body strength of speakers.

Voices and Sexual Selection

The recurrent theme in much of the research cited above is that, independent of language, the human voice is a medium that conveys a host of important biological and social information, and that perceivers are sensitive to this information in a manner that is likely to promote adaptive social behaviour. In this sense, the social perception of voices may profitably be considered in the light of sexual selection theorizing. Sexual selection emerges as result of variation in reproductive success between individuals of the same sex and species (Darwin, 1871). Contemporary models of sexual selection hold that many physical and behavioural traits function as "fitness indicators" – reliable markers of genotypic and phenotypic condition as related to viability and fertility, and that in order in to enhance reproductive success, perceivers are likely to be highly attuned to these indicators (Anderson, 1994; Miller, 2000). In essence, fitness indicators serve as reliable signals of reproductively important traits such as age, health, fertility, and status. For example, one putative fitness indicator is waist-to-hip ratio. In females, waist-to-hip ratio is under hormonal influence, and has been found to be a reliable indicator of age, health, and fertility that affects perceiver judgments of attractiveness and desirability (Singh, 1993; 1995).

The information content of vocal signals, and the types of judgments that perceivers make from vocal information is such that voices may themselves function as salient indicators of fitness related information. Indeed, much recent research into the perception of vocal cues has been conducted within the context of sexual selection theorizing (e.g., Collins, 2000; Feinberg, DeBruine, Jones, & Little, 2008; Feinberg, Jones, Little, Burt, & Perrett, 2005; Feinberg et al., 2006; Fraccaro et al., 2011; Hodges-Simeon, Gaulin, & Puts, 2010; Hughes, Dispenza, & Gallup, 2004; Hughes, Harrison, & Gallup, 2002; 2009; Pipitone & Gallup, 2008; Puts, 2005; Puts, Barndt, Welling, Dawood, & Burriss, 2011; Sell et al., 2010; Wolff & Puts, 2010).

If voices do function as a signal of important fitness related information then they should be an "honest" signal. That is, in order to be reliable, signals should have high relative marginal costs so that low fitness individuals cannot easily fake a high quality version of the signal (Grafen, 1990; Zahavi & Zahavi, 1997). One source of honesty comes from the physiology and structure of the vocal anatomy. As noted previously, individuals can produce acoustic parameters across a range of values and volitionally manipulate their voices in different ways depending on social context. However, the degree of flexibility in the acoustic parameters that an individual can produce is not without limits, and acoustic production is constrained in such a way that it provides legitimate cues to speaker sex, age, and body size. Moreover, evidence strongly indicates a significant hormonal influence on the development of the vocal apparatus. Testosterone appears to be causally linked to the masculinizing of the vocal folds and vocal tract during puberty (Harries et al., 1998; Jenkins, 1998), and with the maintenance of masculine vocal properties during adulthood (Bruckert, Lienard, Lacroix, Kreutzer, & Leboucher, 2006; Dabbs & Malinger, 1999; Need, Durbridge, & Nordin, 1993). Because testosterone is an immunosuppressant (Grossman, 1985), and can be produced in high levels only by individuals that are otherwise healthy and well-buffered (Folstad & Karter, 1992), masculine traits appear to be a reliable cue of health and fitness (e.g., Rhodes, Chan, Zebrowitz, & Simmons, 2003). As such, masculine vocal characteristics that are testosterone dependent, such as low F_0 and low F_d (an acoustic correlate of vocal tract length), may be honest signals of disease resistance and long-term physical health.

In contrast to the masculinizing effects of testosterone, estrogen and progesterone appear to play a role in shaping feminine sounding voices (Abitbol, Abitbol, & Abitbol, 1999). The cytological profile of the larynx is similar to that of the genitals (Caruso et al., 2000), and across the menstrual cycle, histological changes that occur in the endometrium are mirrored by those that occur in the larynx (Abitbol et al., 1999). In line with this, vocal parameters have been found to change in accordance with hormonal variation associated with the menstrual cycle (Higgins & Saxman, 1989), with some forms of hormonal contraception (Amir & Kishon-Rabin, 2004), and with the onset of both puberty and menopause (Caruso et al., 2000). Moreover, F_0 in women is negatively associated with long-term health risk factors (Vukovic, Feinberg, DeBruine, Smith, & Jones, 2010), and positively associated with levels of estrogen (Abitbol et al., 1999; Feinberg et al., 2006), which in turn, are positively associated with fertility (Baird et al., 1999).

The hormones that influence vocal development also influence the development of sexually dimorphic body characteristics such as shoulder-to-hip ratio (Kasperk et al., 1997), waist-to-hip ratio (Singh, 1993), facial configuration (Fink et al., 2005), upper-body musculature and physical prowess (Bhasin et al., 1996; Clark & Henderson, 2003), as well as behavioural traits such as physical aggressiveness (Archer, 1991; 2006; 2009). Thus, common hormonal influences may mediate a link between vocal, somatic, and behavioural development. Together, these lines of evidence suggest that vocal acoustic parameters may be honest cues of hormonal profile, and therefore of immunological competence, fertility, body configuration, and the tendency to engage in aggressive and dominating behaviour.

Not only do vocal traits appear to be legitimate signals of socially and evolutionarily relevant information, but there is growing evidence that perceivers attend to them as such. Recent evidence suggests that perceivers use vocal cues to assess the quality of both potential opposite-sex mates, and of potential same-sex competitors. Voices can be subjectively evaluated, with perceivers readily making assessments of the aesthetic qualities of a speaker's voice (Zuckerman & Driver, 1989). Studies of naturally occurring variation in vocal signals have found negative relationships between judgments of vocal attractiveness and F_0 for male speakers (Bruckert et al., 2006; Collins, 2000), and positive relationships for female speakers (Collins & Missing, 2003). Consistent with these findings, studies that have experimentally manipulated F_0 have found increased attractiveness ratings for male voices with lowered F_0 (Feinberg et al., 2005), and for female voices with raised F_0 (Feinberg, DeBruine, Jones, & Perrett, 2008). In addition to F_0, ratings of voice attractiveness also covary with other putative indices of health and fertility including facial attractiveness (Collins & Missing, 2003), body symmetry (Hughes, Harrison, & Gallup, 2002), waist-to-hip ratio in women (Hughes, Dispenza, & Gallup, 2004), and shoulder-to-hip ratio in men (Hughes, Dispenza, & Gallup, 2004). Voice attractiveness judgments also appear to be functionally context sensitive. When considering short-term mating contexts relative to long-term mating contexts, men exhibit stronger preferences for vocal femininity (voices higher in F_0 and/or F_d) in female speakers

(Puts et al., 2011), while women exhibit stronger preferences for vocal masculinity (voices lower in F_0 and/or F_d) in male speakers. In a similar fashion, women have been found to prefer vocal masculinity in male voices, but not in female voices, more so in high fertility phases of the menstrual cycle than in low fertility phases (Feinberg et al., 2006). To the extent that evaluations of attractiveness function in the service of adaptive mate choice (e.g., Rhodes, 2006), these findings suggest that vocal attractiveness judgments are likely to be adaptive in that they promote the likelihood of mating with individuals who possess hormonally mediated markers of heritable health and fertility.

In addition to attractiveness evaluations, perceivers use vocal signals to assess a range of socially and sexually relevant speaker attributes, including assessing the quality and behavioural intentions of potential same-sex rivals. Masculinized voices are associated with increased ratings of speaker age, body size, and dominance in male speakers (Collins, 2000; Feinberg et al., 2005; Puts, Hodges, Cardenas, & Gaulin, 2007), and are perceived as more dominant than feminized voices in both male and female speakers (Feinberg et al., 2006; Wolf & Puts, 2010), while women perceive feminized female voices as being more flirtatious and as more attractive to men (Puts et al., 2011). Aside from laboratory based judgments, there is also evidence that vocal parameters are related to differences in sexual behaviour and reproductive success. For both men and women, ratings of vocal attractiveness have been found to predict self-reported age of first sexual intercourse, number of sexual partners, and number of extra-pair copulations (Hughes, Dispenza, & Galllup, 2004), and in a natural fertility population living a traditional hunter-gatherer subsistence lifestyle, male F_0 was found to be negatively associated with reproductive success (Apicella, Feinberg, & Marlowe, 2007).

Finally, studies of the ontogeny of vocal attractiveness evaluations have found that judgments made by female adolescents mirror adult judgments in preferring male voices of lower F_0.

However, female children do not show the same preference for low F_0, suggesting that vocal attractiveness judgments are functionally associated with life-stage in such a way that masculinity preferences develop in conjunction with reproductive capability (Saxton, Caryl, & Roberts, 2006).

The above findings are strongly suggestive of the role of sexual selection in shaping both vocal production and perception. Vocal acoustic signals appear to be honest cues of speaker sex, age, and hormonal profile, and voice perception appears to be calibrated in such a way that it tracks these cues in a manner functionally relevant to both inter-sexual mate choice and intra-sexual competition forms of sexual selection. Although sexual selection is not without criticism (e.g., Roughgarden, Oishi, & Akçay, 2006), viewing both voice production and perception in the light of sexual selection theory offers two strategic advantages. Firstly, it provides an orienting theoretical account that integrates a number of diverse phenomena in terms of their ultimate function – that of increasing reproductive success. Secondly, it offers a basis from which to generate novel and testable hypotheses concerning the social perception of human voices.

The Present Research

The considerations of the previous sections are suggestive of two things. The first is that voices appear to contain information about a host of speaker characteristics. The second is that perceivers appear sensitive to much of this information. The present research aims to further understanding of both these aspects of voice related social behaviour (i.e., both the social production and the social perception of vocal signals) as they relate to the specification and detection of fundamental physical characteristics. As highlighted previously, differences in the basic physical composition of individuals, including differences in sex, age, body configuration, and hormonal profile, afford social perceivers different behavioural opportunities. Considerations of the source-filter theory of vocal production suggest that information specifying these physical characteristics is contained in the acoustic properties of vocal signals. Accordingly, in order to guide social behaviour in an adaptive fashion, perceivers are expected to be functionally attuned to the vocal specification of speaker physical characteristics.

In line with this reasoning, the approach taken in the present research is to examine voice related behaviour across a chain of analysis. Consideration is given to the production of vocal signals and the acoustic information that covaries with speaker physical characteristics, listener abilities to perceive speaker characteristics from vocal information, the nature of the information that informs perception, and the behavioural consequences of perception.

Study 1 of the present research examines the first step in this chain of analysis by assessing the relationships between speaker vocal and physical characteristics. On the basis of the arguments outlined previously for the adaptive utility of detecting speaker sex, age, and size, the relationships between these characteristics and the acoustic properties of speakers' voices are investigated. Importantly, this investigation is informed by the examination of acoustic parameters that are theoretically thought to be constrained by anatomical and physiological variation related to sex, age, and size. Moreover, consideration is also given to physical characteristics such as body shape, body symmetry, and physical strength, which are thought to be mediated by similar developmental processes to the vocal apparatus, and for which relationships with acoustic properties have seldom been the subject of systematic empirical investigation. Explicit consideration is also given to two hypotheses relevant to the functional nature of voice perception that previous research has typically only implicitly addressed. The first of these is the notion that vocal signals should show phenotypic variance. That is, there should be significant differences in the acoustic structure of vocal signals between individuals. Without this variation, vocal signals would lack utility as cues to underlying variation in adaptively relevant traits and would be redundant in the context of adaptive social perception. The second is that vocal signals produced by individuals should be relatively stable across different times and contexts. This is not to say that vocal signals cannot, will not, or should not vary, but rather that variation should not be capricious. Signals that varied capriciously across time and context would lack utility as cues for social perception.

Study 2 investigates the attunement of perceivers to the vocal specification of speaker physical characteristics. To the extent that vocal signals convey reliable and valid information regarding adaptively relevant features of speakers, it is expected that perceivers should attend to them as such. Specifically, given the importance for adaptive social behaviour of the accurate detection of the sex, age, and physical stature of individuals, it is expected that perceiver assessments of these characteristics made from vocal cues should accurately reflect the actual characteristics of speakers. Moreover, because judgments that varied capriciously between perceivers would be unlikely to effect adaptive social behaviour, assessments of sex, age, and physical stature should show strong inter-perceiver agreement. That is, perceptual assessments should display both accuracy and consensus. While previous research has investigated the perception of physical characteristics from vocal information, the methods used have often pooled measures across participants. Such an approach informs on the ability of a group of individuals to make accurate assessments of speaker characteristics but makes it difficult to reach conclusions regarding adaptive social perception at the level of the individual perceiver. Accordingly, the current research employs methods more appropriate for evaluating the accuracy of perception at the level of the individual. Moreover, previous research has often failed to consider the nature of the vocal information that that underpins perceptual evaluations of speaker physical characteristics. As such, acoustic analyses are employed to identify the physical properties of vocal signals that perceivers are attending to in their judgments.

Studies 3 and 4 investigate perceiver judgments of vocal attractiveness. To the extent that evaluations of attractiveness function in the service of adaptive mate choice by promoting attraction to cues of health and fertility, it is expected that judgments of vocal attractiveness should co-vary with traits that are indicative of or directly related to health and fertility. Previous research has shown that vocal attractiveness is correlated with speaker body symmetry – a reliable marker of phenotypic and genotypic quality (Hughes et al., 2002). Study 3 of the current research aims to replicate and extend this finding by examining whether the association between vocal attractiveness and body symmetry varies as a function of female perceivers' cyclically varying fertility. Given that women are hypothesized to favour male cues of high fitness during the most fertile phase of the menstrual cycle (Gangestad & Thornhill, 1998), female perceivers may attend to vocal cues of male body symmetry differentially across the menstrual cycle. Specifically, it is expected that during phases of high fertility, female perceivers should exhibit stronger preferences for the voices of symmetrical men than during phases of low fertility.

The possibility of differential attractiveness evaluations across the menstrual cycle raises an additional point of consideration. To the extent that vocal cues specify mate value, they are expected to be attended to both by opposite-sex suitors motivated to assess a potential mate and by same-sex competitors motivated to assess a potential rival. That is, it is expected that judgments made by male and female perceivers should display reasonable agreement, regardless of the sex of the vocaliser. Indeed, Hughes et al. (2002) found both male and female perceivers to favour the voices of more symmetrical speakers. However, given the prediction that female judgments of vocal attractiveness may change across the menstrual cycle, the current research also explores that idea that judgments of vocal attractiveness will vary between male and female perceivers. In terms of promoting reproductive success, female perceivers may benefit from differentially attending to cues of heritable male fitness at different times of the menstrual cycle, but male perceivers are likely to benefit from attending to such cues in a consistent manner (i.e., to be consistently attuned to the quality of potential competitors). Thus, while it is anticipated that female perceiver judgments of vocal attractiveness as a function of male speaker body symmetry will vary across evaluation sessions that occur at different times of the menstrual cycle, it is expected that male judgments will remain stable across evaluation sessions.

Study 4 further examines the influence of cyclically varying fertility on vocal attractiveness, but does so in terms of fertility variation in female speakers rather than perceivers. Given the influence of hormonal factors on vocal production mechanisms, hormonal variation across the menstrual cycle may affect women's vocal signals in a perceptible manner. The current research examines whether perceivers differentially evaluate women's voices at different stages of the menstrual cycle. Specifically, it is expected that perceivers should preferentially evaluate women's voices during phases of high fertility relative to phases of low fertility. Because both male suitors and female competitors may benefit from attending to cues of fertility in female speakers, the judgments made by male and female perceivers are expected to be highly consensual. In both Studies 3 and 4, acoustic analyses are used to examine the physical properties of voices that are informing perceiver evaluations.

Study 5 aims to extend the previous studies examining perceiver attunement to the vocal specification of physical characteristics by considering how vocal cues interact with cues from other modalities to influence social perception. While vocal information specifying speaker physical characteristics is sometimes encountered by social perceivers separate from other sources of information (e.g., during a telephone conversation), it is also frequently encountered in conjunction with other sources of information – most notably the visual information afforded by a person's face (e.g., during a face-to-face conversation). As such, examining the integration of different sources of information in social perception is of considerable importance, both in terms of advancing theoretical understanding of voice perception and in terms of understanding how perceivers navigate the social world on a daily basis. Nevertheless, the majority of previous research concerning the social perception of voices has considered vocal signals in isolation. Accordingly, Study 5 provides a novel hypothesis and examination of the influence that the relationship between vocal and visual information has on social perception. Several researchers have argued that the acoustic cues provided by voices and the visual cues provided by faces both contain highly similar information regarding social targets – including information specifying fundamental physical characteristics (e.g., Belin et al., 2011; Feinberg, 2008), and there is some suggestion that perceptual evaluations made from voices and faces tend to correlate (e.g., Collins & Missing, 2003; Lander, 2008). The present research utilises this notion of confluence between voice and face to examine how information from the two sources is integrated to influence social memory. An argument is presented suggesting that social perceivers have expectations that the information afforded by face and voice cues should covary in a reliable manner. This notion is examined by manipulating the congruence between the vocal and facial cues specifying physical characteristics of sex and age. It is suggested that cases in which a social target's vocal and facial information are discordant with regards to the target's sex or age will be attended to more thoroughly than cases in which a social target's vocal and facial information are concordant with regards to the target's sex or age. This extra attention may then facilitate memory for the incongruent cases relative to the congruent cases.

The remainder of this thesis presents empirical examinations of the hypotheses outlined above, and provides a general discussion of the main findings relative to adaptive social behaviour and to advancing current understanding of the perception of voices in a social context.

Chapter 2

Study 1. Acoustic Correlates of Speaker Physical Characteristics

This chapter outlines an empirical investigation of speaker vocal characteristics. The purpose of this study was two-fold. The first aim was to develop a set of reliable and valid voice recordings, and associated physical measures, to be used in subsequent perceptual studies. The second aim was to conduct exploratory analyses of these recordings, investigating potential relationships between speaker physical and vocal characteristics.

The association between non-linguistic acoustic information and speaker physical characteristics has received considerable research attention. The basic rationale of such investigations is that individual differences in anatomical and physiological characteristics may delimit speaker acoustic parameters in a reliable and predictable manner. While much of this research has produced null or inconsistent results, several patterns concerning the acoustic correlates of speaker sex, age, and physical stature have emerged.

In humans, as in many primate species, a number of acoustic properties exhibit sexual dimorphism (Childers & Wu, 1991; Green, 1981; Mitani & Gros-Louis, 1995; Rendall, Kollias, Ney, & Lloyd, 2005). The vocal folds contain androgen receptors (Newman, Butler, Hammond, & Gray, 2000; Saez & Martin, 1976), and at puberty, greater testosterone levels in men result in the vocal folds growing longer and thicker than for women (Harries, et al., 1998). Consequently, with vocal fold vibrations at around half the rate of that for women, men tend to produce vocalizations of lower F_0. Men may also exhibit lower variation in F_0 (i.e., greater monotonicity) than women, although the proximal mechanisms for this difference are not well understood (Daly & Warren, 2001; Puts, Apicella, & Cardenas, 2011; Saez & Martin, 1976). A number of studies have also documented sex differences in formant frequencies, with females typically having higher formant patterns than males (e.g., Childers

& Wu, 1991; Deterding, 1997; Huber, Stathopoulos, Curione, Ash, & Johnson, 1999; Peterson & Barney, 1952). Formant frequencies reflect the configuration of the vocal tract, with the length of the vocal tract being inversely related to formant frequencies (Huber, et al., 1999). The increase in vocal tract length that occurs with the descent of the larynx in males during puberty (Fitch & Giedd, 1999) results in lower and more closely spaced formants for males relative to females (Fant, 1960). Sex differences in vocal tract shape may also contribute to differences in formant frequencies. For example, male-female differences in the shape of the glottis have been noted (Titze, 1989), while differing proportionate facial bone growth, with women displaying greater upper-jaw growth and lesser lower-jaw growth relative to men, tends to produce greater facial convexity in women (Walker, 1994). Thus, both differences in the size and shape of the vocal tract may result in characteristically lower and more closely spaced formant frequencies for male speakers.

Acoustic parameters have also been found to be age dependent. The physical changes associated with the development from infancy to old age are accompanied by corresponding changes in vocal characteristics. These changes are likely to be, in large part, the product of physiological changes in the vocal tract that occur across the life-span. Such physiological changes include the lengthening of the vocal tract (Xue & Hao, 2003), ossification of connective tissue of the larynx (Kahane, 1987), and stiffening of the vocal folds (Kahane, 1987). In turn, age-related physiological changes have a number of acoustic correlates. For both males and females, F_0 becomes steadily lower throughout childhood until puberty, at which point it decreases rapidly for males but decreases relatively slower for females (Huber et al., 1999). For men, F_0 continues to decline with maturity until middle-age, where it begins to gradually rise, while for women, F_0 declines throughout adulthood until old age, where it begins to rise (Baken & Orlikoff, 2000). In addition, variation in F_0 has been found to change with age, with both males and females showing greater variation as they grow older (Hollien

& Shipp, 1972; Scukanec, Petrosino, & Squibb, 1991), and there is some suggestion that F_0 perturbation (cycle-to-cycle variation in F_0), increases with age (Wilcox & Hori, 1980), although some studies have not found this pattern (Brown, Morris, & Michel, 1989; Ramig & Ringel, 1983). Formant measures have been found to decrease in frequency from young adulthood to old age (Endres, Bambach, & Flosser, 1971; Linville & Fisher, 1985; Xue & Hao, 2003). Typically, this pattern has been considered to be the result of increases in vocal tract size associated with ageing, although some researchers have argued that age related changes in articulatory patterns are responsible (e.g., Rastatter & Jaques, 1990). Xue and Hao (2003) used a method of acoustic reflection analysis to assess age-related changes in vocal tract morphology and associated acoustic correlates. As well as finding the standard pattern of lowered formants associated with increasing age, they also documented age-related increases in the volume of the vocal tract and oral cavity, and concluded that changes in the laryngeal structure, respiratory functioning, and articulatory patterns may all contribute to age-related changes in voice characteristics.

As with speaker sex and age, acoustic correlates of speaker physical stature have been the subject of regular empirical investigation. Historically, F_0 has often been taken as a cue to speaker body size (Darwin, 1872; Gradol & Swan, 1983; Morton, 1977); however, subsequent investigations have produced conflicting results. Across species, larger sized animals tend to produce lower frequency vocalizations than smaller animals, though in many cases, the relationship may not necessarily hold within species (Masataka, 1994; McComb, 1991). Given that longer vocal folds naturally vibrate at lower frequencies than shorter vocal folds, it seems intuitive that larger individuals should have larger vocal folds and consequently, lower frequency vocalizations. In humans, there is some support for this notion at a gross level; in accordance with general size differences children have higher F_0 than adults, and women have higher F_0 than men (Hollien, Green, & Massey, 1994; Titze, 1989).

However, when sex and age are controlled for, most research has failed to find a consistent relationship between F_0 and body size (Bruckert, et al., 2006; Collins, 2000; Gonzalez, 2007; Lass & Brown, 1978; Rendall et al., 2005; van Dommelen & Moxness, 1995), although one study reported a relationship between F_0 and weight in male speakers (Evans, et al., 2006), and another between F_0 and height (Puts, et al., 2011).

One possible explanation for the lack of observed association between F_0 and body size is that growth patterns of the larynx are, at least partially, dissociable from growth patterns of general skeletal size. Unlike most mammals, the human larynx is positioned low in the vocal tract. Consequently, the larynx is able to develop largely independent of the constraints imposed by the boundaries of the skull. In contrast, the length of the vocal tract is constrained considerably by the skeletal structure of the neck and skull. As such, Fitch (1997) has suggested that formant frequencies might provide more robust cues to body size than F_0. There is some support for this notion from animal studies, with a number of species displaying relationships between vocal tract length, formant patterns, and body size (Fitch, 1997; Fitch & Reby, 2001; Harris, Fitch, Goldstein, & Fishing, 2006; Reby & McComb, 2003; Riede & Fitch, 1999), although these studies may have over-estimated the relationship between formants and body size by collapsing across age and sex groupings (Rendall et al., 2005). Nevertheless, human studies have not produced the same consistency. While several studies have found a modest correlation between formant patterns and height in adult men (Evans et al., 2006; Gonzalez, 2004; Rendall et al., 2005; Sell et al., 2010) and two studies have found a correlation for women (Collins & Missing, 2003; Gonzalez, 2004), other studies have reported no significant relationships (Bruckert et al., 2006; Collins, 2000). Similarly, some studies have reported a relationship between formant patterns and speaker weight (Evans et al., 2006; Gonzalez, 2004), while others have not (Bruckert et al., 2006; Collins, 2000; Rendall et al., 2005; Sell et al., 2010). Typically, these studies have indexed formant structure with a measure of formant dispersion (F_d, the average distance between successive formant frequencies). However, Puts et al. (2011) have recently criticised this formulation of formant structure, arguing that it fails to take into consideration the middle formant values, is overly influenced by the value of the highest formant measured, and provides information about formant spacing but not actual formant position. As such, they have proposed an alternative formulation termed formant position (F_p, the average standardized formant value), which is claimed to be a stronger acoustic correlate of vocal tract length and body size than F_d. Accordingly, the current research investigates relationships between speaker body size and both F_d and F_p.

Acoustic parameters other than fundamental and formant frequencies have been considered as potential correlates of speaker body size, but the results have been equivocal. Gonzalez (2007) for example, examined correlations between 27 voice parameters and four measures of body size. While a significant relationship was found between F_0 perturbation and body size for women, with heavier women exhibiting greater frequency perturbation, the correlations between voice parameters and body size were for the most part null or very weak.

It has been suggested that these inconsistent findings may reflect small sample sizes, varied sample characteristics, or varied research methods[2] (Puts et al., 2011). An alternative possibility is that vocal acoustic parameters do not reflect general body size per se, but instead reflect other physiological and phenotypic dimensions that are only indirectly related to overall body size. Many of the same hormones that influence the development of vocal mechanisms also influence the development of specific aspects of body configuration. Thus, voice parameters may be more likely to be associated with body configuration, rather than actual body size. For example, F_0 is a reliable marker of testosterone – increases in

[2] Although it might be expected that any robust effects should hold up across different methods.

testosterone around puberty result in a lowering of F_0 in adolescent males (Hollien, Green, & Massey, 1994), and circulating testosterone levels are negatively related to F_0 in adult males (Dabbs & Mallinger, 1999). As such, it is possible that F_0 may be associated with hormone mediated physical characteristics such as shoulder-to-hip ratio (Kasperk et al., 1997), the ratio of the second to fourth digit (2D:4D) (Zheng & Cohn, 2011), and upper-body musculature (Bhasin et al., 1996). In support of this notion Evans et al. (2006) reported a negative relationship between F_0 and shoulder-to-hip ratio in adult men, whereby men with lower frequency voices tended to have broad shoulders and narrow hips. Similarly, Puts et al. (2011) reported a relationship between F_p and men's upper body strength, whereby men with lower F_p tended to have greater upper body strength. Nevertheless, little research has systematically addressed the relationships between vocal acoustic parameters and anthropometric characteristics that may be mediated by developmental processes similar to those affecting the larynx and vocal tract. As such, in addition to speaker characteristics of sex, age, height, and weight, the present research investigated relationships between vocal parameters and speaker characteristics of physical strength, body symmetry, shoulder-to-hip ratio, waist-to-hip ratio, and 2D:4D.

The present study had two primary goals. The first was to develop a set of reliable and valid voice recordings, and associated physical measures, to be used in subsequent perceptual studies. The second goal was to conduct exploratory analyses investigating possible relationships between speaker vocal and physical characteristics. In light of the considerations discussed above, it was expected that acoustic parameters would show differences between male and female speakers, be moderately correlated with speaker age, and be weakly to moderately correlated with speaker body stature, configuration, and strength. Moreover, because vocal signals are hypothesized to be constrained in such a way that they reliably and validly reflect fundamental features of speakers, it is expected that the measured acoustic parameters should show variation between speakers, but be relatively consistent within speakers, showing strong relationships across different vocal samples and across different measurement sessions.

Method

Participants. Voice samples and anthropometric measures were obtained from 30 female and 30 male volunteers, who were recruited from the University of Canterbury. Women ranged in age from 19 to 36 years (*M* = 23.9, *SD* = 5.3) and men from 18 to 34 years (*M* = 21.8, *SD* = 3.8). Participants were remunerated with a $5 voucher for each session that they took part in. This study was reviewed and approved by the University of Canterbury Human Ethics Committee. All participants were provided with a written information sheet (see Appendices A and B), and signed a written consent form prior to participation (see Appendix C).

Apparatus. Anthropometric measures were taken using digital vernier callipers accurate to 0.01mm, digital scales accurate to 0.1 kg, a 1500 mm anatomical measuring tape with a graduation of 1 mm, and a stadiometer with a graduation of 1 mm. Grip strength was measured with a hand-held Stoelting dynamometer. Voice samples were recorded with a unidirectional microphone (Sony ECM-MS907) to a digital audio recorder (Sony Hi-MD MZ-RH10). Acoustic analyses were carried out using Praat voice analysis software (www.praat.org).

Procedure.

Screening. Participants were initially asked several screening questions to ascertain their suitability for the study. Participants were asked what their first language was and if they had previously experienced or were currently experiencing any conditions that may affect their voice (e.g., illness, chronic smoking, throat surgery, hearing problems). All participants reported New Zealand English as their first language and no participants reported any voice-affecting conditions. Additionally, because the vocal samples analysed in this study were to form the basis of the vocal stimuli for subsequent perceptual studies in which several hypotheses (outlined in Chapter 1) concerned systematic changes as a function of fertility across the menstrual cycle, female participants were asked several questions relevant to their fertility status. Nine women, out of 41 who initially reported interest, were excluded from the study because they reported as having used hormonal contraception in the previous three months, as having an irregular menstrual cycle, as being currently pregnant, or as currently breast-feeding. During this preliminary screening, female participants were also asked to identify the day on which their last menstrual period began, the day on which they expected their next menstrual period to begin (a calendar was provided for reference), and the typical length of their menstrual cycle. Following Thornhill and Gangestad (1999), the day of ovulation was estimated to be 15 days prior to the next menstrual onset and using the reverse counting method (Lenton, Landgren, & Sexton, 1984), which takes into account individual cycle length and controls for the greater variation in the first half of the cycle, female participants were scheduled for two recording sessions: one during a period of relative high fertility, and one during a period of relative low fertility. Actuarial fertility data shows that there is a steady increase in risk of conception (if copulation occurs) in the days preceding ovulation and a rapid decrease following ovulation (Jochle, 1973; Wilcox, Dunson, Weinberg, Trussell, & Baird, 2001). The high fertility session was scheduled on or just prior to the expected day of ovulation and the low fertility session was scheduled during the mid-luteal phase. Follow-up consultation found that for two women, menstruation did not occur at the expected time and they were excluded from analysis. The order of the two sessions was determined by each participant’s position in the menstrual cycle, with the first session scheduled to coincide with the next appropriate cycle phase (high fertility or low fertility).

Nineteen women completed their first session during a phase of high fertility and 11 women completed their first session during a phase of low fertility. Male participants completed two recording sessions one week apart.

Voice recording. Recording sessions took place individually. During each session participants were seated at a table in a noise-isolated recording room (2.5 m X 3 m) with a head-mounted microphone fixed approximately 5 cm from their mouth. Using their normal speaking voice, participants were asked to count from one to ten in English at the rate of approximately one numeral per second, to read the passage "when sunlight strikes raindrops in the air they act as a prism and form a rainbow, the rainbow is a division of white light into many beautiful colours,"[3] to voice the vowels /a/, /e/, /i/, /o/, and /u/, and the sustained (approximately 5 s) vowel /a/. The voice samples used in previous research have often differed across studies depending on the precise goal of the research, and in general there exists a trade-off between the experimental control and robustness of measurement offered by vowel sounds, and the more naturalistic recordings of spontaneously generated speech. The present samples were chosen to in order to obtain voice samples in which the content was both neutral and comparable across speakers, that were consistent with voice samples used in similar previous research, and that allow for the appropriate assessment of conventional acoustic measures, as well as for the assessment of more naturalistic, connected speech patterns.

Anthropometry. In addition to the voice recordings a number of anthropometric measures were obtained from participants. To assess and increase reliability, each trait was measured twice by one of two experimenters – a female research assistant who measured the female participants, and a male research assistant who measured the male participants.

[3] This is the first two sentences of "the Rainbow Passage" (Fairbanks, 1960), a standard reading passage commonly used to elicit samples of connected speech.

Participant height was measured to the nearest 0.5 cm using a stadiometer and weight was measured to the nearest 0.5 kg using digital scales. The circumference of the shoulders, waist and hips were taken to the nearest 0.5 cm. Shoulder circumference was taken by measuring around the greatest width of the shoulders as the participant stood with their arms to the side. Waist circumference was taken at the level of the umbilicus, and hip circumference was taken at the greatest distance around the hips and buttocks. A waist-to-hip (WHR) ratio was calculated by dividing the waist measurement by the hip measurement, and a shoulder-to-hip ratio (SHR) was calculated by dividing the shoulder measurement by the hip measurement. A body mass index (BMI) was calculated by dividing each participant's weight in kilograms by the square of their height in meters (Tovee, Reinhart, Emery, & Cornelissen, 1998). Hand-grip strength, a good indicator of overall upper-body strength (Sell et al., 2010), was assessed to nearest 0.5 kg by having participants squeeze the dynamometer as tightly as possible with each hand. Body symmetry was assessed through a measure of fluctuating asymmetry (FA). The left and right side measurements of seven anatomical traits were made using digital vernier callipers. The measured traits were ear length, elbow width, wrist width, and the length of all fingers, excluding the thumb, taken along the ventral surface from the proximal crease to the tip. These traits have been shown to be easily and reliably measured, and to exhibit minimal directional asymmetry (Gangestad & Thornhill, 1998; Livshits & Kobyliansky, 1991). Participants were also asked if they had experienced any conditions or injuries that could potentially affect the values of the measured traits (e.g., broken bone or recent sprain). In order to guard against the effects of exaggerated asymmetry measures due to injuries, the values of potentially affected traits were excluded from analysis and replaced by the mean trait asymmetry for that participant (12 of the 420 trait measures were replaced with mean values). An overall fluctuating asymmetry measure was then calculated for each participant. The absolute asymmetry of each morphological trait was initially divided by the mean trait size for that participant. This provides an index of asymmetry as a proportion of the size of the trait, ensuring that larger traits do not have a greater influence on overall asymmetry measures than smaller traits. The resulting values were then averaged to yield an overall index of fluctuating asymmetry for each participant. Averaged fluctuating asymmetry indices provide a better estimate than single trait values (Livshits & Kobyliansky, 1991). In addition, the ratio of the length of the second digit to the fourth digit (2D:4D) was calculated for both left and right hands. Finally, each participant's age and sex was recorded.

Acoustic analyses. All acoustic analyses were conducted using Praat analysis software with input parameters defined according to standard recommendations (www.praat.org). Measures were made across the duration of the spoken sentence, the spoken numeral count, and each vowel, and where appropriate, mean acoustic measures were calculated by averaging across values for each participant. Fundamental frequency was measured using Praat's autocorrelation algorithm. The input range was from 75-300 Hz for male speakers and from 100-500 Hz for female speakers. Fundamental frequency mean (F_0-mean), standard deviation (F_0-SD), minimum (F_0-min) and maximum (F_0-max) were measured. Frequency perturbation (cycle to cycle variation in fundamental frequency) was assessed through three measures of jitter: relative jitter – the mean absolute difference between consecutive periods, divided by the average period (jit-local); relative average perturbation – the mean absolute difference between a period and the mean of it and the two periods either side of it, divided by the mean period (jit-rap); and the five-point period perturbation quotient – the mean absolute difference between a period and the mean of it and the four closest periods to it, divided by the mean period (jit-ppq5). Amplitude perturbation (cycle to cycle variation in waveform amplitude) was assessed through four measures of shimmer: relative shimmer – the mean absolute difference between the amplitude values of consecutive periods, divided by the mean amplitude (shim-local); the three-point amplitude perturbation quotient – the mean absolute difference between the amplitude of a period and the mean of the amplitudes of that period and the two periods either side of it, divided by the mean amplitude (shim-apq3); the five-point amplitude perturbation quotient – the mean absolute difference between the amplitude of a period and the mean of the amplitude of that period and the four closest periods to it, divided by the mean amplitude (shim-apq5); and the 11-point amplitude perturbation quotient – the mean absolute difference between the amplitude of a period and the average of the amplitude of that period and the ten closest periods to it, divided by the mean amplitude (shim-apq11). Measures of jitter and shimmer are useful in characterising both healthy and pathological voices (Horii, 1980) and are key components of what is perceived as voice roughness and hoarseness (Moore & Thompson, 1965). While measures of F_0-SD, jitter, and shimmer tend to correlate, each measure offers independent information and none are typically considered redundant (Horii, 1980). The first four formant frequencies (F_1 – F_4) were measured using Praat's linear predictive coding algorithm and visual inspection. Formant dispersion (F_d) was calculated as the mean difference between adjacent formant frequencies, and formant position (F_p) was calculated as the average standardized formant frequency, with formants standardized using between-sex means and standard deviations. Lastly, the harmonics-to-noise ratio (HNR) was calculated as a measure of harmonicity.

Results

Reliability assessment of anthropometric measures. To assess and increase reliability, each anthropometric trait was measured twice during each session. Table 1 shows the relationships between the two measures for both male and female participants. As can be seen the two measures show a high degree of reliability for all traits; all $r(60) \geq .94$, $p <$ 0.001. Subsequently, the two measures were averaged to give a mean trait value for each participant per session.

Table 1. Correlation coefficients between repeated measures of anthropometric traits for male and female participants.

Trait	Males (n = 60)[a]	Females (n = 60)[a]
Height	1.00*	1.00*
Weight	1.00*	1.00*
Shoulders	.99*	.98*
Waist	.97*	.98*
Hips	.98*	.98*
Right ear	.96*	.97*
Left ear	.96*	.96*
Right elbow	.94*	.97*
Left elbow	.96*	.96*
Right wrist	.94*	.95*
Left wrist	.97*	.96*
Right 2D	.98*	.98*
Left 2D	.98*	.97*
Right 3D	.98*	.98*
Left 3D	.96*	.97*
Right 4D	.97*	.97*
Left 4D	.94*	.97*
Right 5D	.97*	.95*
Left 5D	.96*	.98*

[a] For this analysis, the two measuring sessions were treated independently with the two anthropometric measures from the first session pooled with the two measures from the second session.
$*p < .001$

Comparisons of anthropometric measures across recording sessions. The mean values of each measured anthropometric trait from both measurement sessions (sessions one and two for males, low and high fertility sessions for females) are shown in Tables 2 and 3 for males and females respectively. Paired t-tests revealed that for both male and female participants no trait differed significantly between sessions; all $t(29) \leq 1.75$, $p > 0.05$. Thus, the measures from each session were averaged to provide a single mean value for each participant for use in subsequent analyses.

Table 2. Comparison of male speakers' anthropometric measures for each measurement session.

	Session 1		Session 2				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
Height (cm)	181	6.8	181	6.8	0	1	0
Weight (kg)	80	9.7	80	9.7	0.84	.41	0.15
Shoulders (cm)	120	7.9	120	7.7	0.74	.47	0.14
Waist (cm)	86	6.3	86	6.5	0.18	.86	0.03
Hips (cm)	89	6.3	89	5.7	0.67	.51	0.12
Right ear (mm)	63.30	3.10	62.95	3.08	1.85	.08	0.34
Left ear (mm)	62.72	3.35	62.52	3.31	1.13	.27	0.21
Right elbow (mm)	70.47	5.10	70.12	4.81	0.93	.36	0.17
Left elbow (mm)	70.08	5.51	69.56	5.33	1.59	.12	0.29
Right wrist (mm)	60.21	3.90	60.05	3.74	0.69	.49	0.13
Left wrist (mm)	59.99	3.64	60.18	3.96	0.94	.36	0.17
Right 2D (mm)	74.63	4.22	74.75	4.17	0.76	.46	0.14
Left 2D (mm)	74.82	3.91	75.09	3.93	1.75	.09	0.32
Right 3D (mm)	82.70	3.89	82.47	3.91	1.41	.17	0.26
Left 3D (mm)	83.05	3.99	83.03	4.24	0.11	.92	0.02
Right 4D (mm)	77.23	3.94	77.33	4.14	0.59	.56	0.11
Left 4D (mm)	77.08	3.98	76.84	3.61	0.82	.42	0.15
Right 5D (mm)	63.66	4.52	63.51	4.54	0.69	.50	0.13
Left 5D (mm)	63.35	4.75	63.18	4.49	0.65	.52	0.12

Table 3. Comparison of female speakers' anthropometric measures for each measurement session.

	Low fertility		High fertility				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
Height (cm)	167	6.6	167	6.6	1	.32	0.18
Weight (kg)	64	12.7	64	12.4	0.86	.40	0.16
Shoulders (cm)	103	6.5	104	8.2	1.30	.20	0.24
Waist (cm)	75	8.2	74.5	7.7	1.04	.31	0.19
Hips (cm)	91	9.5	90	9.1	0.41	.69	0.07
Right ear (mm)	57.43	4.81	57.64	3.66	0.40	.69	0.07
Left ear (mm)	58.01	4.83	57.71	3.89	0.54	.59	0.10
Right elbow (mm)	56.98	8.48	56.43	8.57	1.02	.32	0.19
Left elbow (mm)	56.80	9.15	56.50	8.94	0.75	.46	0.14
Right wrist (mm)	52.60	3.63	52.16	4.23	0.95	.35	0.17
Left wrist (mm)	52.50	3.60	51.75	3.47	1.68	.11	0.31
Right 2D (mm)	73.21	6.40	73.23	6.32	0.05	.96	0.01
Left 2D (mm)	73.07	5.82	72.87	6.27	0.46	.65	0.08
Right 3D (mm)	80.21	6.71	79.30	7.20	1.58	.13	0.29
Left 3D (mm)	79.52	6.51	78.86	7.18	1.14	.27	0.21
Right 4D (mm)	73.18	6.89	72.89	7.00	0.52	.61	0.09
Left 4D (mm)	73.28	6.09	72.81	6.40	0.99	.33	0.18
Right 5D (mm)	59.14	4.66	58.82	4.95	0.98	.30	0.18
Left 5D (mm)	59.20	4.52	58.93	4.74	0.69	.50	0.13

Assessment of asymmetry measures. The mean left side minus right side trait measures for ear length, elbow width, wrist width, and all digit lengths were assessed for directional asymmetry (Palmer, 1994). One-sample t-tests found that no trait asymmetries were significantly different to zero: for ear length $t(59) = 0.41$, $p = .69$, $d = 0.05$; for elbow width $t(59) = 1.25$, $p = .22$, $d = 0.16$; for wrist width $t(59) = 0.83$, $p = .41$, $d = 0.17$; for 2D $t(59) = 0.05$, $p = .96$, $d = 0.01$; for 3D $t(59) = .025$, $p = .80$, $d = 0.03$; for 4D $t(59) = 0.68$, $p = .50$, $d = 0.09$; and for 5D $t(59) = 0.69$, $p = .49$, $d = 0.09$. Asymmetry values for all traits were approximately normally distributed with a mean g_2 value of 0.57 indicating slight leptokurtosis, which is generally consistent with previous measures of fluctuating asymmetry (Gangestad & Thornhill, 1998).

Comparisons of acoustic measures across vocal samples. Voice recordings were made of four different vocal tasks (spoken numerals, a spoken sentence, vowel sounds, and a sustained vowel sound). An assumption of the current research is that the measured acoustic parameters reflect underlying individual speaker specific properties. As such, it is expected that respective acoustic values should correlate across vocal samples. Following van Dommelen and Moxness (1995), correlation coefficients were calculated between the four vocal samples for each of the primary acoustic parameters (F_0, F_1, F_2, F_3, & F_4), and are shown in Table 4.

Table 4. Correlation coefficients between voice sample types for each of the primary acoustic parameters (F_0, F_1, F_2, F_3, & F4). N = 60.

	Numerals	Vowels	Sentence
F_0			
Vowels	.98*		
Sentence	.99*	.98*	
Sustained vowel	.94*	.94*	.95*
F_1			
Vowels	.58*		
Sentence	.03	.13	
Sustained vowel	.21	.25	.16
F_2			
Vowels	.66*		
Sentence	.73*	.56*	
Sustained vowel	.72*	.37*	.69*
F_3			
Vowels	.70*		
Sentence	.32*	.39*	
Sustained vowel	.45*	.15	.21
F_4			
Vowels	.79*		
Sentence	.61*	.71*	
Sustained vowel	.56*	.56*	.52*

*$p < 0.01$

With the exception of the first formant frequency, acoustic measures across the four types of vocal samples were largely positively correlated, suggesting that the measures represent speaker specific properties in a reliable manner. Given the association of acoustic measures across the different sample types, the remainder of the current analysis focuses on measures from only the spoken numeral vocal samples.[4]

[4] Analogous assessments for the spoken sentence and vowel sound samples are presented in Appendices D – F. Analyses of these samples reveal highly similar patterns to the results of the spoken numeral analysis presented in the current chapter.

Comparisons of acoustic measures across recording sessions. Acoustic measures for the two recording sessions (sessions 1 and 2 for males, low and high fertility sessions for women) were compared using paired-samples *t*-tests. Table 5 shows the comparisons for male speakers and Table 6 the comparisons for female speakers. For both male and female speakers, there were no significant differences between sessions for any of the assessed acoustic parameters; all $t(29) \leq 1.8$, $p \geq 0.10$. For further analysis, measures from the two recording sessions were averaged to give mean values for each participant.

Table 5. Comparison of male speakers' acoustic measures for each measurement session.

	Session 1		Session 2				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	111	8.3	111	7.9	0.17	.87	0.03
F_0-SD (Hz)	11.6	6.2	10.7	5.3	0.74	.46	0.14
F_0-min (Hz)	94.7	7.3	95.0	7.8	0.27	.79	0.05
F_0-max (Hz)	158.3	39.4	145.3	22.5	1.70	.10	0.31
F_1 (Hz)	686	79.2	674	72.1	0.84	.41	0.15
F_2 (Hz)	1667	92.1	1650	76.9	1.31	.20	0.24
F_3 (Hz)	2734	121.8	2716	105.8	1.12	.27	0.20
F_4 (Hz)	3609	153.9	3627	225.5	0.45	.66	0.08
Jit-local (%)	2.13	0.61	1.94	0.67	1.91	.07	0.35
Jit-rap (%)	0.92	0.32	0.82	0.32	1.67	.11	0.30
Jit-ppq5 (%)	0.87	0.26	0.83	0.26	1.02	.32	0.19
Shim-local (%)	9.73	1.72	9.24	21.8	0.99	.33	0.18
Shim-apq3 (%)	3.50	0.90	3.26	0.72	1.51	.14	0.27
Shim-apq5 (%)	5.04	1.35	4.81	1.08	0.83	.42	0.15
Shim-apq11 (%)	9.17	2.35	9.05	2.08	0.22	.83	0.04
HNR	12.55	2.29	12.98	2.02	1.47	.15	0.27

Table 6. Comparison of female speakers' acoustic measures for each measurement session.

	Low fertility		High fertility				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	209	16.9	208	16.4	0.30	.77	0.05
F_0-SD (Hz)	31	12.2	30	10.5	0.37	.71	0.07
F_0-min (Hz)	132	28.6	138	32.4	0.97	.34	0.18
F_0-max (Hz)	337	72.0	350	81.7	0.66	.52	0.12
F_1 (Hz)	697	59.7	687	59.7	0.79	.44	0.14
F_2 (Hz)	1861	3.95	1870	62.0	0.82	.42	0.15
F_3 (Hz)	2877	90.74	2863	102.8	1.01	.32	0.18
F_4 (Hz)	3969	107.3	3968	121.6	0.08	.94	0.01
Jit-local (%)	1.63	0.35	1.77	0.62	1.70	.10	0.31
Jit-rap (%)	0.79	0.20	0.88	0.37	1.75	.09	0.32
Jit-ppq5 (%)	0.76	0.23	0.82	0.31	1.42	.17	0.26
Shim-local (%)	6.07	1.51	6.59	1.67	1.89	.07	0.34
Shim-apq3 (%)	2.48	0.64	2.60	0.92	0.90	.38	0.16
Shim-apq5 (%)	3.07	0.83	3.22	1.18	1.13	.27	0.21
Shim-apq11 (%)	5.07	1.18	5.45	1.87	1.85	.08	0.34
HNR	16.66	2.20	16.18	2.59	1.60	.12	0.29

Comparison of male and female measures. Table 7 shows the mean anthropometric and acoustic measures for male and female participants. In addition to the direct anthropometric measures, the derived measures of BMI, SHR, WHR, 2D:4D and FA are included. Similarly, in addition to the direct acoustic measures, four values derived from these measures are included. The respective measures of frequency perturbation (jit-local, jit-rap, & jit-ppq5) and amplitude perturbation (shim-local, shim-apq3, shim-apq5, & shim-apq11) were averaged to give two composite perturbation measures (jit-mean and shim-mean), and the formant frequencies F_1-F_4 were used to calculate measures of formant dispersion (F_d) and formant position (F_p). For both male and female speakers the physical and acoustic measures were comparable with those reported in previous studies (e.g., Baken & Orlikoff, 2000; Evans et al., 2006; Gallup et al., 2007; Gonzalez, 2004; Hughes et al., 2004; Manning, 2002; Puts et al., 2011; Rendall et al., 2005). A number of traits displayed sexual dimorphism. In terms of anthropometric measures, male participants exhibited greater body size and strength measures (height, weight, BMI, SHR, WHR, & grip strength), and lower 2D:4D ratios compared to female participants. No sex difference was found in the degree of fluctuating asymmetry. In terms of acoustic parameters, male participants exhibited lower fundamental frequency (F_0-mean, F_0-SD, F_0-min, and F_0-max), formant frequency (F_2, F_3, F_4, F_d, & F_p), and HNR values, and greater amplitude perturbation (shim-local, shim-apq3, shim-apq5, shim-apq11, & shim-mean) values than female participants.

Table 7. Comparison of anthropometric and acoustic measures between male and female speakers.

	Males		Females				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
Height (mm)	181	6.8	167	6.6	7.99	< .001	2.07
Weight (kg)	80	9.7	64	12.6	5.65	< .001	1.47
BMI	24.33	2.18	22.61	3.57	2.25	.028	0.60
Grip (kg)	54	6.9	32	5.0	14.12	< .001	3.80
SHR	1.34	0.09	1.15	0.06	16.76	< .001	1.87
WHR	0.97	0.03	0.83	0.04	9.96	< .001	5.71
2D:4D	0.97	0.02	1.00	0.02	5.08	< .001	1.50
FA	0.02	0.01	0.02	0.01	0.68	.51	0.15
F_0-mean (Hz)	111	7.82	208.9	15.8	30.41	< .001	8.29
F_0-SD (Hz)	11.15	4.66	30.58	8.89	10.61	< .001	2.87
F_0-min (Hz)	94.85	7.24	134	24.54	8.53	< .001	2.51
F_0-max (Hz)	151.83	24.31	343.75	53.39	17.92	< .001	4.94
Jit-local (%)	2.03	0.58	1.70	0.45	2.50	.015	0.66
Jit-rap (%)	0.86	0.28	0.83	0.26	0.51	.61	0.13
Jit-ppq5 (%)	0.85	0.24	0.79	0.24	0.97	.34	0.25
Jit-mean (%)	1.25	0.35	1.11	0.30	1.71	.09	0.44
Shim-local (%)	9.49	1.42	6.33	1.40	8.68	< .001	2.24
Shim-apq3 (%)	3.26	0.72	2.54	0.70	3.92	< .001	1.19
Shim-apq5 (%)	4.93	0.96	3.14	0.95	7.24	< .001	1.87
Shim-apq11 (%)	9.11	1.72	5.26	1.45	9.38	< .001	2.43
Shim-mean (%)	6.70	1.04	4.32	1.05	8.84	< .001	2.31
F_1 (Hz)	680	65.36	692	49.83	0.77	< .45	0.20
F_2 (Hz)	1659	76.80	1866	43.15	12.88	< .001	3.45
F_3 (Hz)	2726	106.04	2870	88.00	5.74	< .001	1.49
F_4 (Hz)	3618	351.31	3969	109.19	5.22	< .001	1.52
F_d (Hz)	979	117.00	1092	36.81	5.05	< .001	1.47
F_p	-0.46	0.69	0.50	0.34	6.79	< .001	1.86
HNR	12.76	2.01	16.42	2.25	6.64	< .001	1.72

Relationships between anthropometric and acoustic measures. Given that both the anthropometric and acoustic measures were sexually dimorphic, relationships between the two groups of measures were assessed separately for male and female participants. Table 8 displays the relationships for male participants and Table 9 the relationships for female participants. Despite the large number of comparisons in these tables, following Wolff and Puts (2010), statistical corrections for multiple tests were not used for these analyses. Although this has the possibility of increasing the Type I error rate, given the exploratory nature of the analyses and the lack robust findings from previous research, it was deemed prudent to maximize the chances of detecting potential relationships and interpret them cautiously, rather than increase the chances of obtaining false negative results (Rothman, 1990).

Table 8. Correlations between physical and acoustic measures for male speakers. N = 30.

	Age	Height	Weight	BMI	Grip	SHR	2D:4D	FA
F_0-mean	.01	-.05	-.03	.00	.10	.13	.02	-.09
F_0-SD	.27	-.08	-.09	-.07	.00	-.10	-.02	-.02
F_0-min	-.22	-.09	-.04	.02	-.07	.22	-.24	-.33
F_0-max	.06	-.01	-.13	-.18	-.21	-.24	-.04	-.06
Jit-loc	.29	-.01	.13	1.8	-.06	-.18	.01	.06
Jit-rap	.10	.04	.01	-.01	-.14	-.17	-.07	.10
Jit-ppq5	.18	-.07	-.05	-.01	-.05	-.03	.05	.28
Jit-mean	.23	-.01	.06	.09	-.08	-.15	.00	.12
Shim-loc	.15	.04	.06	.05	-.09	-.41*	.03	-.09
Shim-apq3	-.05	.20	.14	.00	-.02	-.10	.00	-.10
Shim-apq5	-.01	.04	.08	.08	-.16	-.37*	.04	-.14
Shim-apq11	-.20	.17	.09	-.02	.02	-.48**	.10	-.10
Shim-mean	-.04	.13	.10	.03	-.06	-.44*	.06	-.12
F_1	-.03	.06	.20	.24	.30	.32	-.09	.14
F_2	.15	.02	.18	.25	.28	.28	.12	.17
F_3	.13	-.13	.05	.19	.26	.24	.02	.21
F_4	-.03	-.06	-.11	-.09	-.14	.01	.43*	.01
F_d	-.03	-.07	-.15	-.13	-.21	-.01	.44*	-.01
F_p	.08	-.04	.07	.15	.22	.25	.21	.16
HNR	-.16	.28	.05	-.17	.01	.05	.11	.02

*p < 0.05, **p < 0.01

Table 9. Correlations between physical and acoustic measures for female speakers. N = 30.

	Age	Height	Weight	BMI	Grip	WHR	2D:4D	FA
F_0-mean	-.10	-.12	-.29	-.30	-.19	-.04	-.15	.16
F_0-SD	-.17	-.07	-.11	-.10	-.04	.10	-.01	-.06
F_0-min	.21	.39*	.35	.27	.13	.08	.11	.06
F_0-max	.10	.02	.00	-.01	-.26	-.12	.26	-.16
Jit-loc	-.13	-.22	-.35	-.34	-.17	-.04	-.11	.23
Jit-rap	.06	-.09	-.19	-.21	-.21	-.03	-.05	.22
Jit-ppq5	.06	-.04	-.08	-.11	-.24	-.05	-.11	.28
Jit-mean	-.03	-.15	-.25	-.26	-.21	-.04	-.10	.26
Shim-loc	-.2	-.23	-.41*	-.40*	-.07	.04	-.04	.29
Shim-apq3	.06	.02	-.12	-.18	-.13	.05	.09	.39*
Shim-apq5	-.20	-.18	-.37*	-.38	-.17	.18	-.03	.53**
Shim-apq11	-.18	-.21	-.33	-.31	-.15	.12	-.03	.42*
Shim-mean	-.17	-.19	-.35	-.36	-.13	.10	-.02	.43*
F_1	.06	-.09	-.17	-.17	-.12	-.26	.06	.00
F_2	.01	-.16	-.10	-.06	-.15	.03	.07	-.25
F_3	-.09	.00	-.05	-.08	.07	-.10	-.42*	-.27
F_4	-.43*	-.06	-.27	-.32	.10	-.14	-.02	-.04
F_d	-.45*	-.02	-.18	-.24	.15	-.02	-.04	-.04
F_p	-.10	-.11	-.21	-.22	-.06	-.22	-.15	-.21
HNR	.14	.25	.19	.13	.05	-.09	-.14	-.24

*$p < 0.05$, **$p < 0.01$

Relationships between physical anthropometric measures and acoustic measures were, for the most part, non-significant. For male speakers, only two physical measures were associated with acoustic parameters; SHR was negatively associated with measures of amplitude perturbation (shim-loc, shim-apq5, shim-apq-11, and shim-mean), and 2D:4D was positively associated with the fourth formant frequency (F_4) and formant dispersion (F_d). For female speakers, age was negatively correlated with F_4 and F_d, suggesting that older speakers had more closely spaced formant frequencies. Two measures of body size (weight & BMI) were negatively correlated with measures of amplitude perturbation (shim-loc & shim-apq5), while height was positively correlated with F_0-min. FA was positively correlated with amplitude perturbation (shim-apq3, shim-apq5, shim-apq11, & shim-mean).

Discussion

This study aimed to develop a set of reliable stimuli to be used in subsequent perceptual studies, and to explore relationships between speaker physical and acoustic properties. Physical anthropometric measures showed a high degree of consistency, both for repeat measures made within each recording session and for measures across sessions. For the most part, acoustic measures showed significant positive correlations across the four different types of voice sample and no significant differences between recording sessions (also see Appendices D – F), suggesting that the measured parameters represent speaker-specific vocal properties in a reliable manner. Moreover, both the acoustic measures and the physical measures were comparable to those reported in previous studies (e.g., Baken & Orlikoff, 2000; Evans et al., 2006; Gallup et al., 2007; Gonzalez, 2004; Hughes et al., 2004; Manning, 2002; Puts et al., 2011; Rendall et al., 2005).

In terms of the relationships between speaker physical and acoustic characteristics, the present results are, in general, in line with previous investigations. Sex differences were found for a number of acoustic parameters, with male speakers displaying lower fundamental (F_0-mean, F_0-min, F_0-max, & F_0-SD) and formant frequency (F_2, F_3, F_4, F_d, & F_p) values than female speakers. With the exception of F_0-SD, these differences are likely to reflect differences in the anatomy of the vocal apparatus, caused by the exaggerated growth and descent of the larynx in males relative to females (Fitch & Giedd, 1999; Harries et al., 1998). By contrast, sex differences in F_0-SD are not well understood, but there is some suggestion that differences in neuropsychological control of vocal mechanisms may play a role (Puts et al., 2011). Additionally, male speakers displayed greater amplitude perturbation (shim-local, shim-apq3, shim-apq5, shim-apq11, and shim-mean) than female speakers, however the reasons for this difference are not clear and previous research has produced inconsistent results, with some studies finding a similar pattern to the present result and others finding the reverse (Brockman, Drinnan, Storck, & Carding, 2011).

Previous research has typically found decreases in fundamental and formant frequencies, and increases in perturbation measures associated with increasing age (Baken & Orlikoff, 2000; Xue & Hao, 2003). In contrast, the present results found only very limited associations between acoustic properties and age. Two formant measures were significantly correlated with age for female speakers, with older female speakers tending to have lower F_4 and F_d values, while no acoustic properties were significantly correlated with age for male speakers. However, given the comparatively restricted age range of the current sample (18 to 36 years), this result is perhaps not surprising, and it seems likely that greater age variation would produce more robust relationships.

While some previous studies have found relationships between body size and either fundamental frequency (Evans et al., 2006; Puts et al., 2011) or formant frequencies (Collins & Missing, 2003; Gonzalez, 2004; Rendall et al., 2005; Sell et al., 2010), others have not (Bruckert et al., 2006; Collins, 2000; Gonzalez, 2007; Lass & Brown, 1978; Rendall et al., 2005; Sell et al., 2010; Van Dommelen & Moxness, 1995). In the present study, no fundamental or formant frequency measures showed significant relationships with height, weight, or BMI for male speakers, and only F_0-min was significantly related to height for female speakers. Two measures of amplitude perturbation (Shim-loc & Shim-apq5) were negatively correlated with weight and one (Shim-loc) with BMI for female speakers, although these relationships were in the reverse direction to those previously reported in a comparable analysis (Gonzalez, 2007).

One explanation for these inconsistent findings is that rather than being associated with overall body size, voice acoustic features are instead associated with more specific indices of body configurations, particularly configurations that may be affected by similar hormonal influences as the vocal apparatus. However, the present results provide limited support for this notion. Hand-grip strength, a good predictor of general physical health, testosterone levels, and upper body strength (Gallup et al., 2007; Sell et al., 2010), was not related to any voice parameter for either male or female speakers. This result is in line with previous research. Sell et al. (2010), found no relationships between either fundamental or formant frequency measures, and a physical strength measure, although Puts et al. (2011) found small, but inconsistent relationships of F_0-mean, F_0-SD, and F_p with a measure of arm strength.

While the same hormones that underlie vocal changes during puberty are also implicated the development of sex-typical SHR configurations in men and WHR configurations in women (Abitbol et al., 1999; Hughes et al., 2004; Kasperk et al., 1997; Singh, 1993), the present results found no acoustic measures to be significantly correlated with WHR in female speakers, and only measures of amplitude perturbation (Shim-loc, Shim-apq5, Shim-apq11, & Shim-mean) to be significantly correlated with SHR in male speakers. This result contrasts with that of Evans et al. (2006) who found a negative relationship between F_0 and SHR in male speakers, but is consistent with that of Vukovic, Feinberg, DeBruine, Smith, and Jones (2010), who found no relationship between F_0 and WHR in female speakers.

Similarly, consistent with Ferdenzi et al (2011), who found no relationship between fundamental frequency and 2D:4D ratio, acoustic measures were for the most part, unrelated to measures of 2D:4D ratios, with only F_4 and F_d being positively correlated with 2D:4D in male speakers. This result is perhaps not surprising given current understanding of hormonal influences on the development of both digit ratios and vocal mechanisms. 2D:4D is thought to be determined by the influence of foetal hormones during the first trimester, and to remain fixed through later development (Garn, Burdi, Babler, & Stinson, 1975; Manning, 2002).

Vocal mechanisms develop prenatally at around the same, however, these mechanisms are not fixed to the same extent and change considerably under later hormonal influence, especially during puberty (Fitch & Giedd, 1999; Harries et al., 1998; Moore, 1977). While some researchers have suggested positive correlations between prenatal and adult hormone levels (e.g., Jamison, Meier, & Campbell, 1993; Manning, 2002), the present results suggest that vocal features are somewhat decoupled from the organizational effects of prenatal hormones by the activational effects of circulating hormones later in the life-span. Interestingly, vocal parameters for female speakers were found to be relatively stable across the menstrual cycle, showing no differences between the spoken numeral recordings made at high fertility and those made at low fertility.[5] This finding is considered further in Chapter 5.

Measures of fluctuating asymmetry showed no association with any acoustic measure for male speakers, but significant positive correlations with measures of amplitude variation for female speakers, whereby speakers with greater asymmetry tended to have higher levels of shimmer (Shim-apq3, Shim-apq5, Shim-apq11, and Shim-mean). Hughes et al. (2008) also found a measure of amplitude perturbation (Shim-apq11) to be correlated with fluctuating asymmetry, but only for male speakers. Together, these results suggest that body asymmetries may be reflected in acoustic perturbation measures. The vocal folds are bilaterally paired structures and unequal masses, lengths, or tensions between the two folds can lead to irregular phonation. The extreme case of this occurs in patients with unilateral vocal fold paralysis where one fold exhibits normal tension while the other is relaxed (Tigges, Mergell, Herzel, Wittenberg, & Eysholdt, 1997). While, the larynx often exhibits asymmetry (Hirano, Kurita, Yukizane, & Hibi, 1989) which can, to some extent, be

[5] Analyses of the other vocal samples found no differences in vocal parameters between menstrual cycle phases for the spoken sentence or vowel sound samples, but significantly higher F_0-mean and significantly lower F_0-SD for recordings of the sustained vowel /a/ made at high fertility relative to low fertility – see Appendices D – F.

compensated for (Ludlow, Sedory, Holzer, & Fujita, 1997), individual differences in the bilateral symmetry of the vocal apparatus may generate differences in the acoustic signal.

The current results should be interpreted in the light of a number of considerations. Firstly, it is possible that a larger sample size may have revealed further relationships between speaker physical and acoustic measures. However, the current sample size is comparable to that of previous research in this area, and the results are, for the most part, consistent with previous findings – including findings from those studies with a larger sample size. Secondly, it has been suggested that the inconsistent results in previous research, particularly in terms of detecting relationships between body size/stature and vocal properties, may be partly attributable to methodological variations between studies (Puts et al., 2011). One source of such variation is the vocal samples produced by speakers. Vocal samples employed in previous research have ranged from the production of vowel sounds (e.g., Collins, 2000), to numeral counts (e.g., Hughes et al., 2002), to the reading of standard passages (e.g., Vukovic et al., 2010), to the production of spontaneously generated connected speech samples (e.g., Puts, 2005). The choice of sample typically depends on the specific goals of the research, but in general there exists a trade-off between the experimental control and reliability of acoustic assessment offered by vowel sounds and the more ecologically valid production of spontaneous speech. Nevertheless, given that they come from the same source, acoustic parameters are expected to correlate across sample types (Baken & Orlikoff, 2000; Van Dommelen & Moxness, 1995). In the present study, speakers were asked to produce four vocal samples (spoken numerals, vowel sounds, a standard reading passage, and a sustained vowel sound) with acoustic measures correlating positively across the different samples. It should also be emphasized that separate analyses for each type of vocal sample revealed highly similar patterns of results (see Appendices D – F).

The present results suggest that while vocal characteristics vary between speakers, they do not do so entirely independently of anatomical and physiological characteristics of individual speakers. Consistent with previous research, differences in acoustic properties as a function of speaker sex were found, while relationships between acoustic properties and individual features such as age and body configuration were less robust. Moreover, the acoustic measures were found to be correlated across vocal samples, and consistent between measurement sessions, suggesting that the obtained voice recordings reflect individual speaker-specific features in a reliable manner, and are appropriate for use in subsequent perceptual studies.

Chapter 3

Study 2. The Perception of Speaker Physical Characteristics from Vocal Information

Study 1 explored the relationships between the physical characteristics of 60 speakers and the acoustic properties of their voices. While acoustic analyses found only a few robust relationships between speaker vocal and physical measures, the ability of perceivers to utilise vocal information to make judgments about the speakers was not examined. Previous research has demonstrated that perceivers use non-linguistic information contained in speech signals to make a host of inferences about speakers (Ladefoged & Broadbent, 1957; Giles & Powsland, 1975). For example, perceivers readily make inferences about a speaker's race (Lass, Tecca, Mancuso, & Black, 1979), socioeconomic status (Ellis, 1967), and personality (Allport & Cantril, 1934; Hughes, Pastizzo, & Gallup, 2008; Zuckerman & Driver, 1989). Given the adaptive importance of detecting the fundamental physical characteristics of other individuals, the present study assesses the ability of listeners to accurately perceive sex, age, and physical stature from vocal signals.

As reported in Study 1, several acoustic parameters demonstrate significant differences between male and female speakers, and a number of studies have shown that perceivers readily make judgments about the sex of a speaker from voice information (e.g., Bennett & Montero-Diaz, 1982; Krauss et al., 2002; Ingemann, 1968; Lass et al, 1976; Schwartz & Rhine, 1968; Traunmuller, 1997; Whiteside, 1998). In general, these studies show both a high degree of reliability and validity in perceiver judgments. For example, across 1,800 judgments made from sustained vowel sounds, Lass et al. (1976) reported an accuracy of 96% in judgments of speaker sex.

Several studies have examined the ability of perceivers to use voice information to make assessments of speaker age. Perceivers can readily differentiate the voices of younger adults from older adults (Ptacek & Sander, 1966), and can assign general age categories to speakers based on the sound of their voice (Cerrato, Falcone, & Paoloni, 2000). Mean estimates of speaker age generally closely approximate mean actual age values (Krauss et al., 2002; Lass et al., 1982), and age estimates made from speech appear to be just as accurate when perceivers are familiar with the language of a speaker as when they are unfamiliar (Braun & Cerrato, 1999), suggesting that age estimation from vocal cues is not dependent on linguistic comprehension. Moreover, several studies have found significant positive correlations between estimated speaker age and actual speaker age (e.g., Bruckert et al, 2006; Krauss et al., 2002; Neiman & Applegate, 1990; Ryan & Burk, 1974), with one study demonstrating estimates of age made from voice recordings to be comparable in accuracy to estimates made from photographs (Krauss, et al., 2002).

Research examining perceiver estimates of speaker body size has proven more controversial. An early series of studies reported that naïve perceivers were capable of making accurate height and weight judgments from voice information for both male and female speakers (e.g., Lass & Davis, 1976; Lass, et al., 1978; Lass, et al., 1980; Lass et al., 1979; Lass et al., 1981). For example, Lass et al. (1978) asked perceivers to make direct estimates of speaker height and weight from recorded voice samples, and found only a small discrepancy between the average actual and estimated values. However, these studies typically pooled values across both speakers and perceivers, and assessed accuracy through a measure of "average difference" – the mean actual value obtained from speakers compared to the mean estimated value made by perceivers. In contrast to more appropriate measures (e.g., the average *absolute* difference), such an approach is likely to cancel out individual deviations and to produce misleading estimates of perceiver accuracy. This problem was demonstrated by a study in which participants were asked to think of any male speaker, and to estimate the weight of that speaker. The mean estimated value generated through this procedure closely approximated the mean weight obtained from a group of randomly selected males, even though the participants heard no actual vocal information from those males (Cohen, et al., 1980).

Re-evaluation of the data presented in several of the studies by Lass et al. showed that while perceivers were highly consistent in their estimates of height and weight, such estimates did not reflect the actual height and weight values of the target speakers (van Dommelen, 1993; Gonzalez, 2003). Similar findings of consistent but erroneous body size judgements have also been reported in other studies (e.g., van Dommelen & Moxness, 1995; Collins, 2000). Nevertheless, some research has demonstrated perceivers to be capable of making accurate speaker judgments from voice information. Krauss et al. (2002), presented perceivers with a recorded speech sample followed by two full-length, frontal photographs – one of the speaker, and one of another individual of the same sex as the speaker. They then asked perceivers to select the photograph that depicted the speaker. Performance was considerably above chance, with perceivers identifying the speaker's photograph on nearly 77% of trials. Krauss et al. suggested that perceivers performed this task by first estimating speaker characteristics (including height and weight), and then selecting the photograph that was most closely matched to those characteristics.

Using a slightly different approach, Rendall et al. (2007) played randomly paired voice samples to perceivers and asked them to select the sample that came from the larger speaker. For both male and female speakers, when the actual height differential between speakers was negligible, estimates were only at chance levels, but the proportion of correct responses increased in accordance with the absolute difference in height between speakers. That is, the greater the disparity in the speaker height, the better the participants performed in deciding which speaker was larger.

Studies that have asked perceivers to directly estimate speaker size have also demonstrated an ability to make correct size inferences from voice information. Krauss et al. (2002), asked perceivers to estimate the height and weight values of speakers and found the average absolute difference between estimated and actual values to be comparable to estimates made from visual information. Both Collins (2000) and Bruckert et al. (2006) found female perceivers to be able to accurately estimate the weight, but not height, of male speakers, while van Dommelen and Moxness (1995) found accurate height and weight estimates for male voices but not female voices, with accurate ratings being mostly confined to ratings made by male perceivers. In an assessment of voice samples from four distinct language groups, Sell et al. (2010) found that perceiver estimates showed positive relationships with the actual height and weight measures of both male and female speakers and that accuracy was not dependent on language familiarity. Thus, in spite of a number of early studies showing ambiguous findings, more recent studies, using a variety of methods, have demonstrated that perceivers can make correct judgments about speaker body size from voice information.

How perceivers are able to make judgments regarding speaker size remains unclear. Few studies have examined both perceiver judgments and the acoustic properties of speaker voices that perceivers are attending to in order to make such judgments. Studies that have done so, have often failed to identify acoustic properties that account for significant variation in perceiver judgments, or found that perceivers are utilising acoustic information that is not related to the physical property being assessed. For example, van Dommelen and Moxness (1995) found perceivers used both fundamental and formant frequencies to estimate body size, but that these acoustic properties were not related to the actual body size of speakers in their sample. Similarly, both Collins (2000), and Bruckert et al. (2006) found female perceivers to correctly estimate the weight of male speakers, however neither study was able to identify acoustic cues that were actually related to speaker weight. Moreover, perceivers in the Bruckert et al. (2006) study appeared to use speech intonation information to estimate speaker height, but their estimates were inaccurate. Sell et al. (2010) found ratings of height and weight to be negatively associated with F_0 (lower frequency voices were rated as being produced by larger speakers), but no relationship between F_0 and actual body size of the speakers was observed. Thus, perceivers were able to make height and weight judgments that were related to the actual height and weight of speakers, but appeared to do so, at least in part, on the basis of erroneous F_0 cues.

An alternative approach to identifying the acoustic information that perceivers are attending to in their judgments of speakers is to experimentally manipulate specific voice parameters and observe the effect on perceiver evaluations. Fitch (1994) asked perceivers to make size attributions from synthesized speech that included a range of fundamental and formant frequency values, with manipulations of both parameters affecting size judgements independently of each other. Rendall et al. (2007) manipulated the same two parameters, and found that listeners were sensitive to formant cues in making relative size judgments but that their usage was somewhat overridden by the inclusion of F_0 cues. Similarly, experimentally lowering speaker F_0 and increasing apparent vocal tact length (through formant manipulations) both produced increased size evaluations in male speakers (Feinberg et al., 2005). Thus, it appears that perceivers may be attending to both F_0 and formant structure in their evaluations of speaker size. Nevertheless, experiments using manipulated stimuli do not demonstrate that perceivers use those acoustic cues to make accurate inferences about the body size of real speakers. As Gonzalez (2006) has pointed out, there is a need for studies in which perceivers make judgments from natural voices and in which true body parameters are used as a criterion by which to evaluate performance.

The present study utilises the set of voice samples and associated speaker physical information obtained in Study 1 to examine the judgments perceivers make about the physical characteristics of speakers. It is predicted that perceivers will make both reliable and valid judgments about the physical characteristics of speakers from vocal information alone. Specifically, it is expected that judgments of speaker sex, age, height, and weight will show strong agreement between perceivers and will be positively related to the actual measures obtained from the speakers. In addition, the acoustic analyses conducted in Study 1 will be used as a basis to examine specific vocal cues that perceivers may be attending to in their evaluations.

Method

Participants. Sixty students (30 male, 30 female) were recruited from a research participation pool at the University of Canterbury to act as perceivers. Female perceivers ranged in age from 19 to 42 years (M = 22.9, SD = 5.5) and male perceivers from 19 to 52 years (M = 23.1, SD = 4.2). This study was reviewed and approved by the University of Canterbury Human Ethics Committee. All perceivers were provided with a written information sheet and signed a written consent form prior to participation (see Appendices G & B).

Voice samples. Sixty voice samples generated through the procedure described in Study 1 were used as voice stimuli in this study. These samples consisted of 30 female speakers and 30 male speakers counting from 1 to 10 in English, at a rate of approximately one numeral per second. Spoken numeral counts provide voice samples in which the content is both neutral and comparable across individual speakers, and have been used successfully in prior research examining perceiver judgments from voice information (e.g., Ellis, 1967;

Hughes, et al., 2002; Pipitone & Gallup, 2008). Speaker physical and acoustic properties were assessed as detailed in Study 1.

Apparatus and materials. Voice samples were played to perceivers using the Acoustica audio software program (www.acoustica.com) through a pair of noise attenuating headphones (Sennheiser HD-201) and responses were made on a series of rating scales (see Appendix H).

Procedure. Perceivers were invited to take part in a study that was described as an investigation into the judgements people make about others based on hearing their voices. They were initially asked if they had any conditions that could potentially affect their hearing, with no participants reporting any issues. Perceivers were then provided with an information sheet giving a brief description of the research and outlining their rights as research participants (see Appendix G). Each perceiver was then asked to sign a consent form (see Appendix B) and seated at a table in a noise isolated testing room. Participants were told that they would be played a series of recorded voices through a set of headphones, and that after hearing each voice they should answer the questions on the response sheet (see Appendix H). Participants completed the listening sessions individually and during each session listened to one of three sets of 20 voice recordings. The 20 voice recordings in each set consisted of samples from 10 female speakers and 10 male speakers, with the order of presentation randomized across perceivers. The sets were constructed so that a given voice sample was only included in one set, ensuring that each speaker was assessed an equal number of times.

Immediately after hearing each voice sample, participants were asked to respond to four questions. They were first asked to indicate whether they thought the speaker was male or female. They were then asked to estimate how old, tall, and heavy they thought the speaker to be. Previous research has used a variety of measures to address such estimates. For example, Lass & Davis (1976) utilised a multiple-choice response option (e.g., under 5 ft 0 in, 5 ft 0 in - 5ft 5 in etc.) while other research has asked participants to give precise estimates (i.e., numerical values) (e.g., Krauss, et al., 2002), or used rating scales anchored with numerical values (e.g., van Dommelen & Moxness, 1995). Pilot testing for the current study found that several participants had considerable difficulty in assigning numerical height and weight values to their estimates. That is, they reported being able to make judgments about a target speaker's relative height or weight (e.g., "very tall," "taller than me, but not that tall"), but were unsure about reporting their judgments in appropriate units of measurement. As such, height and weight estimates were obtained by asking perceivers to place a mark on 10 cm analogue scales. Each scale clearly indicated where its mid-point was situated and was anchored at the lower and upper extremes with text labels. For height, these anchor labels were "very short" and "very tall," and for weight "very light" and "very heavy." After hearing and assessing each voice sample, participants were asked to indicate if they thought the speaker had an accent (i.e., if they thought that the speaker's first language was not New Zealand English) and if they recognised the speaker. No speakers were identified as having an accent other than New Zealand English, while one participant reported recognising a speaker and the ratings for that trial were excluded from analysis.

Results

Inter-perceiver agreement. The extent to which perceivers displayed agreement in their evaluations of target speakers' physical characteristics was assessed using intra-class correlation coefficients (ICC). ICCs were calculated for the estimates of speaker sex, age, height, and weight, separately for each of the three sets of speakers, and are shown in Table 10. In all three sets, perceivers were unanimous in their judgments of target speaker sex, and showed high levels of agreement in their judgments of age, height, and weight. In addition, comparison between male and female perceiver estimates of target age $t(118)$, = 0.67, p = .50, d = 0.12, height $t(118) = 0.78$, $p = .44$, $d = 0.14$, and weight $t(118) = 0.50$, $p = .62$, d = 0.09, showed no significant differences. As such, male and female perceiver estimates were combined for further analyses.

Table 10. Inter-rater reliability measures for perceiver estimates of target speakers' sex, age, height, and weight.

	ICC	95% CI	p-value
Sex			
Set 1	1	1 - 1	.
Set 2	1	1 - 1	.
Set 3	1	1 - 1	.
Age			
Set 1	.86	.75 - .94	< .001
Set 2	.69	.45 - .86	< .001
Set 3	.85	.73 - .93	< .001
Height			
Set 1	.87	.77 - .94	< .001
Set 2	.91	.84 - .96	< .001
Set 3	.89	.81 - .95	< .001
Weight			
Set 1	.90	.83 - .96	< .001
Set 2	.88	.79 - .95	< .001
Set 3	.87	.77 - .94	< .001

Relationships between actual and estimated physical characteristics. In judging whether target speakers were male or female, perceivers showed complete accuracy. That is, across 60 target speakers, each of which was judged 20 times, no male or female perceivers made any errors in determining the sex of a speaker.

In order to compare results with previous studies, the accuracy of perceivers' estimates of speaker age, height and weight was initially assessed by computing Pearson correlations between averaged perceiver ratings and the actual speaker values. The relationships between mean perceiver estimates and actual target characteristics are shown in Figure 2 separately for male and female speakers.

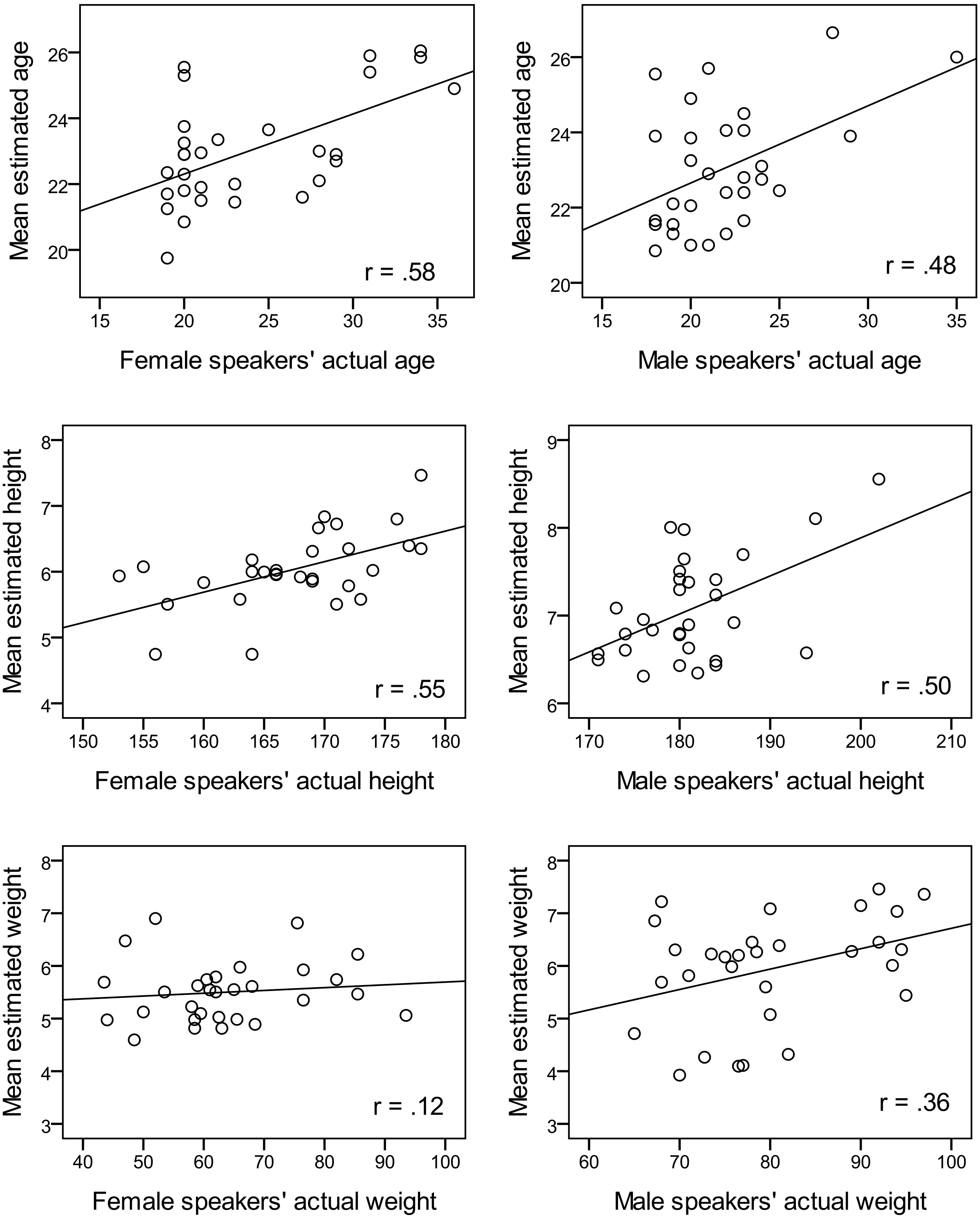

Figure 2. Scatterplots showing the relationships between mean estimated values and male and female speakers' actual age, height, and weight measures.

Mean estimated age was positively correlated with actual age for both male speakers, $r(30) = .48$, $p < .01$, and female speakers, $r(30) = .58$, $p < .001$. As can be seen, there was a general trend whereby estimations of speaker age tended to track actual speaker age. A similar pattern can be seen for estimations of target height; mean estimated height correlated positively with actual target height both for male speakers, $r(30) = .50$, $p < .005$, and for female speakers, $r(30) = .55$, $p < .001$. With regards to weight, mean estimated weight was significantly correlated with actual weight only for male speakers, $r(30) = .36$, $p < .05$. No significant correlation was found for estimates of female speakers' weight, $r(30) = .12$, $p = .54$. Pooling estimates across all speakers (i.e., not distinguishing between male and female speakers) shows a similar pattern with regards to estimates of age, $r(60) = .52$, $p < .001$, and weight, $r(60) = .34$, $p < .01$, but a much stronger correlation for estimates of height $r(60) = .76$, $p < 0.001$.

While aggregating estimates across perceivers has been the conventional technique in evaluating the accuracy of voice ratings, this method reflects the accuracy of perceivers' pooled estimates and provides limited information about the average accuracy of individual perceivers. Moreover, using mean estimates of speaker values is likely to inflate the magnitude of any relationship between actual and perceived values (van Dommelen, 1993). As such, a secondary analysis was conducted in which correlations between actual and estimated measures were calculated for each individual perceiver, and tested against the null hypothesis that the average individual correlation was zero. The results of this analysis are shown in Table 11. Although the relationships are of less magnitude than pooled mean estimates, they depict the same pattern of association. Estimates of speaker age, height, and weight were positively correlated with the actual measures obtained from speakers, with the exception of estimates for the weight of female speakers.

Table 11. One-sample t-tests examining whether the mean of individual perceivers' correlations (*r*) between estimated and actual measures of age, height, and weight are significantly different from zero.

	Mean *r*	*SD*	*t*	*p*	*d*
Age					
Male speakers	.24	.35	5.29	< .001	0.68
Female speakers	.28	.34	6.31	< .001	0.82
All speakers	.26	.27	7.47	< .001	0.96
Height					
Male speakers	.36	.27	10.27	< .001	1.33
Female speakers	.24	.34	5.36	< .001	0.69
All speakers	.49	.18	21.36	< .001	2.76
Weight					
Male speakers	.24	.30	6.10	< .001	0.79
Female speakers	.06	.36	1.23	.22	0.16
All speakers	.19	.26	5.76	< .001	0.74

Acoustic parameters and perceiver judgments. In order to assess what acoustic information perceivers may be attending to in their judgments, multiple linear regression analyses were performed with perceiver estimates of speaker physical characteristics as dependent variables and the acoustic measures of F_0-mean, Jit-mean, Shim-mean, and F_d as independent predictor variables, Table 12. These measures were chosen as predictors based on the relationships between speaker acoustic and physical properties predicted by the source-filter approach to vocal production (Fant, 1960; Titze, 1994), the relationships examined in Study 1, and consideration of previous research examining the acoustic features that underlie perceiver judgments of speaker characteristics outlined earlier.

Table 12. Regression analyses examining acoustic parameters as predictors of the physical characteristic judgments made by perceivers.

Trait	β	t-value	p-value
Sex			
R^2 = .95, $F(4, 59) = 295.44$, $p < .001$			
F_0-mean	-0.85	16.86	< .001
Jit-mean	-0.12	3.05	< .01
Shim-mean	0.141	2.30	< .05
F_d	-.09	2.61	< .05
Age			
R^2 = .06, $F(4, 59) = 0.15$, $p = .96$			
F_0-mean	-0.04	0.16	.87
Jit-mean	0.07	0.36	.72
Shim-mean	-0.11	0.38	.71
F_d	-0.10	0.66	.51
Height			
R^2 = .49, $F(4, 59) = 13.01$, $p < .001$			
F_0-mean	-0.60	3.48	< .001
Jit-mean	-0.24	1.73	.09
Shim-mean	0.19	0.89	.38
F_d	-0.01	0.05	.96
Weight			
R^2 = .08, $F(4, 59) = 1.19$, $p = .32$			
F0-mean	-0.24	1.04	.30
Jit-mean	-0.12	0.67	.50
Shim-mean	0.03	0.12	.91
F_d	-.06	0.37	.71

In judging a target speakers' sex, F_0-mean, Jit-mean, Shim-mean, and F_d were all significant predictors. No acoustic variables significantly predicted estimations of age or weight, while only F_0-mean significantly predicted estimations of speaker height. However, this relationship may primarily be due to differences between male and female speakers. That is, male speakers tended to be both taller and to have voices with lower F_0 values than female speakers (see Study 1). As such, perceivers may be using F_0 as a cue to sex rather than as a direct cue to height. To examine this possibility, further regression analyses examined estimates of height separately for male and female speakers. Within-sex analyses found F_0-

mean not to be a significant predictor of height for either male speakers, $\beta = -0.72$, $t = 0.32$, $p = .75$, or female speakers $\beta = -0.20$, $t = 0.97$, $p = .34$.

Discussion

Perceivers in this study were asked to estimate the physical characteristics of target speakers. Estimates displayed considerable agreement between perceivers, and with the exception of female speaker body weight, estimates of speaker sex, age, height, and weight reflected the actual characteristics of the speakers. These estimates were made after listening to recorded voice samples that were of identical semantic content, suggesting that perceivers used variation in speaker vocal characteristics to assess physical characteristics.

Several features of the results are of particular note. First, is the degree of consensus among perceivers. Perceivers were unanimous in their judgments of speaker sex, and were highly consistent in their judgments of speaker age, height and weight. It appears that social perceivers are sensitive to certain acoustic parameters of voices and are able to utilise those parameters to make judgments about physical characteristics of speakers in a highly consistent manner. Moreover, there were no sex differences in perceiver estimations; male and female perceivers were able to utilise vocal information to evaluate the physical characteristics of speakers to a comparable extent. This is in contrast to a previous finding in which male perceivers were better at estimating male body size than female perceivers, a finding which was taken to be the result of the importance of accurately assessing rivals during male intra-sexual competition (van Dommelen & Moxness, 1995). However, the present results are more in line with those of Sell et al. (2010), who found no sex differences in perceiver evaluations of speaker physical strength, and argued that both males and females should be functionally attuned to making such judgments.

Second, the inferences made were, for the most part, in line with the true features of the speakers. Perceivers in the present study judged the sex of target speakers with complete accuracy, suggesting that differences in vocal signals between male and female speakers are readily distinguished by perceivers, and are able to be used to make valid judgments regarding speaker sex. Pooled estimates of speaker age were strongly and positively correlated with the actual age of speakers, with analysis that averaged across individual perceiver correlations having the expected effect of reducing the strength of this association. While this average individual correlation was weaker than that reported by Krauss et al. (2002), this discrepancy may be due to the smaller age range of the speakers in the present study. Speakers in the Krauss et al. study ranged from 20 to 60 years of age, compared to 18 to 36 years in the current study. It seems likely that with a greater spread of ages, the magnitude of the relationship between estimated and actual age values would increase.

Previous research has shown that perceivers are capable of making gross distinctions about speaker body size from vocal information (Rendall et al., 2007), but research asking for direct estimates has produced equivocal results (Gonzalez, 2006). In the present study estimates of height and weight were obtained by asking participants to indicate where they thought a given speaker would lie on an analogue scale. These responses were then compared with the actual values obtained from speakers. Rather than assessing perceiver ability to assign absolute numerical values to a target – which is likely to be of limited practical utility – the current method essentially assessed perceiver ability to rank-order speakers. In doing so, perceiver estimates of body size were found to positively correlate with the actual speaker measures. Similar to estimates of age, pooling estimates across perceivers produced correlations between estimated and actual height values that were of a greater magnitude than averaging across individual correlations. Additionally, because the distributions of height for male and female speakers differed (see Study 1), combining male and female speakers into a single category, and thus increasing the range of speaker height, also increased the magnitude of the relationship between estimated and actual height. The results were slightly different for estimates of weight. While weight estimates were positively correlated with actual weight for male speakers, perceivers were largely unable to accurately judge female speaker weight. Compared to height, weight is more readily malleable (e.g., through diet and exercise), and as such, vocal cues may able to be somewhat dissociated from speaker weight, making weight estimations from voice more difficult. However, the reason for differences in the accuracy of weight perceptions for male and female speakers is unclear from the present results.

In the present study all target speakers spoke the same 1 to 10 numeral count. As such, perceiver judgments must have been derived from variation in the acoustic signal between speakers that was not semantically meaningful. Such variation may come from at least two sources. One source of variation is in the physical properties of the vocal production mechanisms. As detailed in Study 1, physiological and anatomical differences in sex, age, and body size are associated with variation in both the source and filter components of vocal signal characteristics. For example, sex differences in both the size of the vocal cords, and in the size and shape of the vocal tract contribute to characteristically different sounding male and female voices. A second source is social, with some of the differences in the vocal sounds people produce being under normative social influence. For example, there are perceptually identifiable differences between the voices of male and female children despite little obvious difference in their vocal anatomy (Fitch & Giedd, 1999). Similarly, in adults, the difference between basal F_0 and average speaking F_0 is greater for women than for men, with women typically placing their voices mid-range and men placing their voices toward the lower-end of their available range (Gradol & Swann, 1983). Thus, it is possible that speakers in the present study were judged as male or female, in part because they spoke in a stereotypically masculine or feminine manner. However, it seems less likely that perceivers based their judgments on vocal cues that were consistent with speech norms associated with age, height, and weight (see Krauss et al., 2002).

Regardless of the source of variation, identification of the acoustic features that perceivers attend to in their judgments of speaker physical properties is not straightforward. In the present study F_0-mean, Jit-mean, Shim-mean, and F_d all significantly accounted for variation in perceiver judgments of speaker sex. Study 1 revealed significant differences in F_0-mean, Shim-mean, and F_d between male and female speakers. Taken together, the results suggest that these acoustic parameters represent valid cues that perceivers use in judging speaker sex. However, the results are less clear for estimates of age and body size. No acoustic parameters were found to significantly account for perceiver judgments of age or weight. F_0-mean significantly predicted height judgments but did not do so when judgments were analysed separately for male and female speakers. Combined with the lack of acoustic features identified as being related to speaker physical characteristics in Study 1, the present results make it difficult to specify the acoustic features that allowed perceivers to estimate the age and body size of speakers as well as they did. Nevertheless, the results of this study show that listeners are perceptive to variations in the acoustic structure of vocal signals specifying speaker sex, age, and body size.

Recent research has questioned whether perceiver evaluations are directly related to body size, or whether body size evaluations are actually a by-product of the assessment of other physical features. Sell et al. (2010) have argued that perceivers may be more efficient at making strength and physical fighting assessments from vocal information rather than simple height and weight assessments. Indeed, they found perceivers to make height, weight, and strength assessments that were highly correlated, but that tracked cues of physical strength independent of height and weight. Nevertheless, although Sell et al. found F_0 to be utilised as an acoustic cue in assessments of strength, F_0 was not related to strength itself. While perceiver evaluations of speaker strength were not assessed not the present study, Study 1 found no relationship between acoustic measures and speaker hand-grip strength (a reliable indicator of overall upper-body strength). Thus, even if height and weight estimates are a by-product of general strength estimates, the acoustic parameters that perceivers use in such evaluations remain unknown.

The current results found that perceivers were able to use vocal acoustic information to make judgments about the sex, age, and body size of target speakers. While previous research has typically considered this issue in terms averaged or pooled perceiver estimates, the present study also considered judgments at the level of the individual perceiver. These judgments displayed high levels of consensus between perceivers and for the most part accurately reflected the true characteristics of the speakers. It appears that social perceivers are functionally attuned to the information content of vocal signals in such a way that they are able to accurately detect adaptively relevant physical characteristics of speakers. Establishing the precise nature of the information that perceivers attend to remains an open issue for future research.

Chapter 4

Study 3. Menstrual Cycle Variation in Women's Attraction to the Voices of Symmetrical Men

Study 2 found that perceivers were capable of using vocal cues to make judgments about the physical characteristics of speakers. Previous research has demonstrated that perceivers also readily make judgments regarding attractiveness based on the sound of a speaker's voice (Zuckerman & Driver, 1989; Zuckerman, Hodgins, & Miyake, 1990). Much of this voice attractiveness research has been developed within the framework of sexual selection theorising and suggests that attractiveness evaluations of both male and female voices covary with underlying markers of speaker phenotypic profile (e.g., Collins, 2000; Feinberg et al., 2005; Feinberg et al., 2006; Hughes, Dispenza, & Gallup, 2004; Hughes, Harrison, & Gallup, 2002; Puts, 2005; Puts, et al., 2011).

For example, Hughes et al. (2002) investigated how ratings of voice attractiveness co-vary with a speaker's fluctuating asymmetry (FA). FA represents non-directional deviation from perfect bilateral symmetry for morphological traits that are, on average, bilaterally symmetrical in a population, and is taken to be a proxy measure of developmental stability – an individual's ability to cope with genetic and environmental perturbations during development (Moller & Swaddle, 1997; Van Valen, 1962). Individual differences in FA have been found to exhibit moderate heritability (Johnson, Gangestad, Segal, & Bouchard, 2008; Livshits & Kobyliansky, 1991; Moller & Thornhill, 1997), and across a range of species, including humans, are associated with differences in physical health, longevity, fecundity, and cognitive ability (Furlow, Armijo-Prewitt, Gangestad, & Thornhill, 1997; Gangestad & Simpson, 2000; Johnstone, 1995; Moller & Thornhill, 1998; Rhodes et al., 2001; Zebrowitz, Hall, Murphy, & Rhodes, 2002). Given these associations, there is reason to expect mate selection preferences that are based on FA or on traits that covary with FA (Thornhill & Gangestad, 1999). Indeed, in their investigation of the relationship between FA and voice attractiveness, Hughes et al. (2002) found that for both male and female speakers, individuals with lower indices of FA (i.e., more symmetrical morphological features) were rated as having more attractive voices. As such, it was suggested that vocal quality may be a salient indicator of underlying phenotypic and genetic quality, particularly as it pertains to developmental stability.

If this is indeed the case, then female preferences for male voice characteristics that co-vary with FA may themselves vary as a function of female perceivers' cycling fertility. A number of studies have reported that women exhibit changes in attraction and sexual preferences across the menstrual cycle. For example, during periods of high fertility, women show increased preference for masculine male faces (Johnston, Hagel, Franklin, Fink, & Grammer, 2001; Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999), increased preference for male behavioural displays of social dominance and competitiveness (Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004), increased preference for masculine male voices (Feinberg et al., 2006; Puts, 2005), increased preference for the scent of symmetrical males (Gangestad and Thornhill, 1998; Rikowski & Grammer, 1999), and increased preference for male creative intelligence (Haselton & Miller, 2006).

Rather than being seen as incidental by-products of hormonal variation across the menstrual cycle or as socially mediated learned behaviours, such changes are typically viewed as the result of sexual selection operating on female mate choice strategies. The putative function of these cyclical fluctuations in female sexual psychology is to increase reproductive success by promoting the likelihood of sexual intercourse, particularly with males who possess phenotypic markers of high genetic fitness, during the most fertile phase of the menstrual cycle (for a review, see Thornhill & Gangestad, 2008).

Given the association between voice attractiveness and FA, it may be that women's attractiveness evaluations for the voices of men with lower FA measures would be greater at times of high fertility than at times of low fertility. Such a possibility would be directly analogous to findings in the olfactory domain that while women generally prefer the scent of men that exhibit low levels of FA, they do so to a greater extent during times of high fertility relative to low fertility (Gangestad & Thornhill, 1998; Rikowski & Grammer, 1999; Thornhill & Gangestad, 1999). That women's attraction to the voices of men as a function of FA may change across the menstrual cycle receives support from findings showing that women's preferences for male voices that are lower-pitched and masculinized increase during phases of high fertility (Feinberg et al., 2006; Puts, 2005). Extending this approach to incorporate male speakers' FA, the present study examines whether the relationship between voice attractiveness and FA varies as a function of female listeners' fertility. Following Hughes et al. (2002), it is predicted that both male and female perceivers will rate the voices of male speakers that exhibit lower indices of FA as more attractive. However, given that women are hypothesized to favour markers of high fitness during the most fertile phase of the menstrual cycle, it is predicted that during phases of high fertility women will exhibit stronger preferences for the voices of symmetrical men than during phases of low fertility.

Method

Participants. Thirty female participants, ranging in age from 18 to 29 years (*M* = 22, *SD* = 3.5) and thirty male participants, ranging in age from 18 to 34 years (*M* = 22, *SD* = 4.2) were recruited from the University of Canterbury to act as perceivers in this study. Perceivers were remunerated with a $5 voucher for their participation. This study was reviewed and approved by the University of Canterbury Human Ethics Committee.

Voice samples. Voice samples from 30 male speakers obtained in Study 1 were used in this study. The speakers ranged in age from 18 to 34 years (*M* = 21.8, *SD* = 3.8). The voice samples consisted of speakers counting from 1 to 10 in English, at a rate of approximately one numeral per second. For all speakers a measure of FA was obtained. For methodological details of speaker recruitment, voice sampling procedures, FA assessment, and analysis of acoustic parameters refer to Study 1.

Apparatus and materials. Voice samples were played to perceivers using the Acoustica audio software program (www.acoustica.com) through a pair of noise attenuating headphones (Sennheiser HD-201) and responses were made on a series of rating scales (see Appendix I).

Procedure. Prior to being included as a perceiver in this study, all participants reported having no conditions that could potentially affect their hearing, and all female participants reported that they experienced a regular menstrual cycle and were not currently using hormonal contraception, breastfeeding, or pregnant. Following the same procedure detailed in Study 1, female perceivers were asked to identify the day on which they expected their next menstrual period to begin, the day on which their last menstrual period began, and the typical length of their cycle. Through this consultation, two listening sessions were scheduled for each female perceiver; one at menstruation (a period of relative low fertility) and one at, or just prior to, ovulation (a period of relative high fertility). The order of the two sessions was determined by each participant's position in the menstrual cycle, with the first session scheduled to coincide with the next appropriate cycle phase (high fertility or low fertility). Thirteen women completed their first session during a phase of high fertility and 17 women completed their first session during a phase of low fertility. For male perceivers, the two listening sessions were scheduled two weeks apart.

Upon arrival at the first listening session perceivers were given an information sheet providing a brief description of the research and outlining their rights as a research participant (see Appendices J & K). Each perceiver was then asked to sign a consent form (see Appendix B) and complete a brief demographic questionnaire (see Appendix I). Perceivers were asked to indicate their sex, age, and sexual orientation. Sexual orientation was assessed by asking perceivers to circle a number on a 7-point Likert scale ranging from 1 = exclusively homosexual to 7 = exclusively heterosexual. All participants gave responses of 6 or above. Perceivers were then seated at a table in a noise isolated testing room and told that they would be played a series of recorded voices through a set of headphones, and that after hearing each voice they should answer the questions on the response sheet (see Appendix I).

During each session, perceivers listened to one of two sets of 15 male voice samples played one at a time. The sets were constructed so that each of the voice samples from the 30 male speakers were heard an equal number of times and so that perceivers heard the same 15 voices during both listening sessions, but with the order of presentation randomised across sessions. Immediately after hearing each voice, perceivers were asked to rate how attractive, sexy, and pleasant they found the voices.[6] Ratings were made on three 7-point Likert scales ranging from 1 = very unattractive/unsexy/unpleasant to 7 = very attractive/sexy/pleasant. Perceivers were also asked whether they recognised the voices. Two perceivers indicated that they thought they recognised a voice and these ratings were consequently excluded from analysis.

Results

Inter-perceiver agreement. Table 13 shows the intra-class correlation coefficients for perceiver ratings of voice attractiveness, pleasantness, and sexiness for each of the two

---

[6] Perceivers were also asked to rate how masculine, feminine, and physically attractive they thought each speaker was. These questions are not directly relevant to the current considerations and are not examined here. Refer to Appendix I for further details of these additional questions.

sets of male voices. As can be seen there was a high degree of concordance among perceiver ratings with all intra-class coefficients .74 or above. Ratings of voice attractiveness, pleasantness, and sexiness were also highly correlated, both for male perceivers (r = .77 to .87, all p < .001) and for female perceivers (r = .87 to .94, all p < .001), and were subsequently averaged to give composite vocal attractiveness scores for each participant.

Table 13. Inter-rater reliability measures for perceiver ratings of male voice attractiveness, pleasantness, and sexiness.

	ICC	95% CI	p-value
Attractiveness			
Set 1	.81	.63 - .92	< .001
Set 2	.75	.51 - .90	< .001
Pleasantness			
Set 1	.82	.66 - .93	< .001
Set 2	.74	.50 - .90	< .001
Sexiness			
Set 1	.80	.61 - .92	< .001
Set 2	.82	.66 - .93	< .001

Voice attractiveness ratings. Voice attractiveness ratings made by male perceivers during the first listening session and during the second listening session were both negatively correlated with male speaker FA, r = -.40, p < .05, and r = - .38, p < .05 respectively. A t-test for a difference between dependent correlations found no significant difference between the two relationships, $t(29) = 0.21$, p > .10. During both listening sessions, male perceivers judged the voices of male speakers that exhibited lower levels of FA to be more attractive. Similarly, ratings of vocal attractiveness made by female listeners during periods of high fertility were found to be negatively correlated with male speakers' FA, r = -.40, p < .05. However, no such significant relationship was found for ratings made during periods of low fertility, r = -.17, p = .38. A t-test for a difference between dependent correlations found these two relationships to be significantly different, $t(29) = 3.69$, $p < 0.05$.

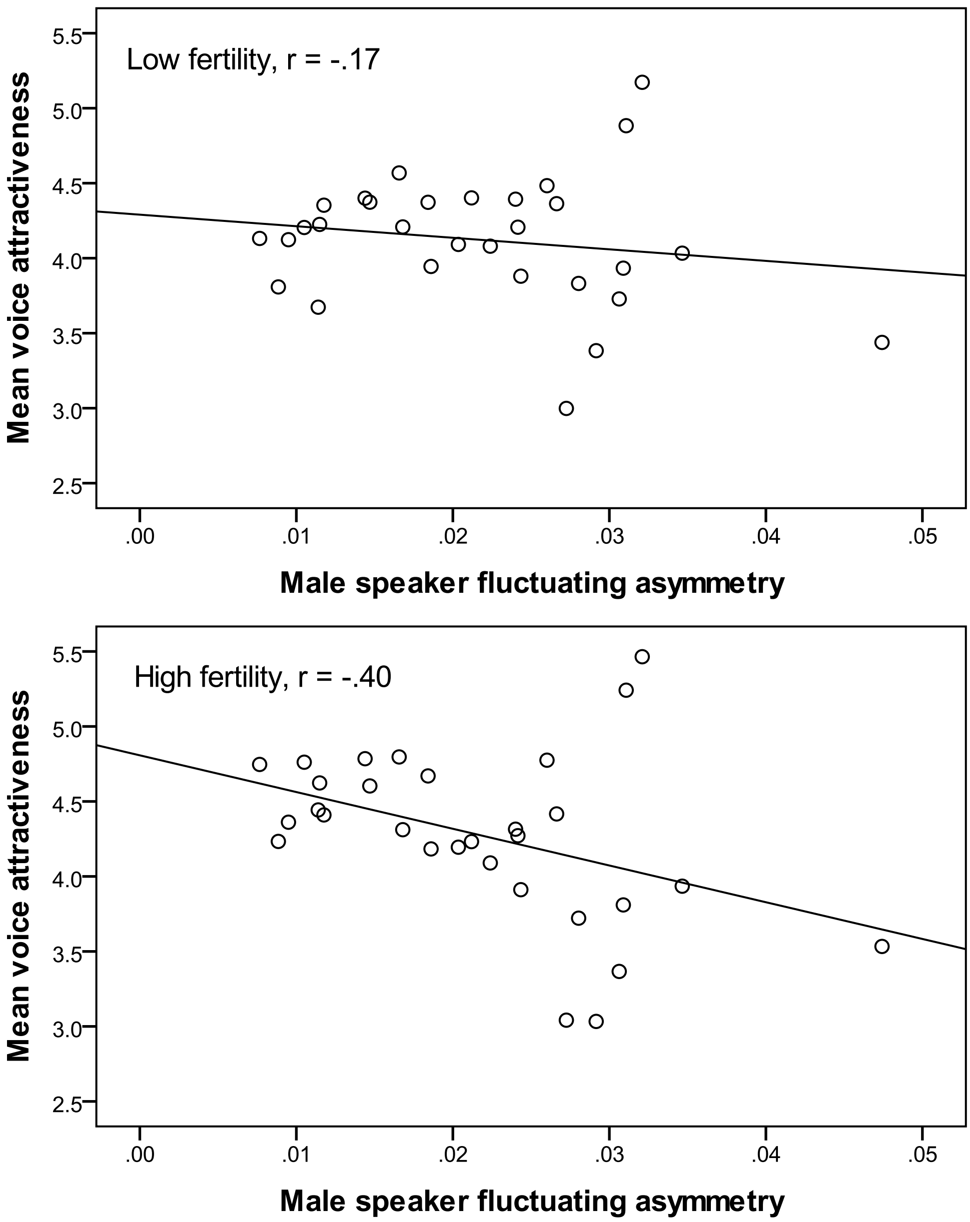

Figure 3. Relationships between male speakers' fluctuating asymmetry and mean voice attractiveness ratings made by female listeners at low fertility (upper panel) and high fertility (lower panel).

As can be seen in Figure 3, these results demonstrate a shift in ratings of voice attractiveness as a function of menstrual cycle phase. At times of high fertility female listeners exhibited voice attractiveness ratings that were more strongly correlated with male speaker FA than at times of low fertility.

Acoustic parameters and attractiveness ratings. Table 14 shows the results of regression analyses for perceiver ratings of male speaker voice attractiveness with acoustic measures as independent predictors. For both male and female perceivers, F_0-mean predicted voice attractiveness ratings, whereby male voices that were judged as more attractive tended to have lower F_0. However, as detailed in Study 1, no acoustic measures (including F_0-mean) were significantly related to male speaker FA measures. Therefore, despite the association between F_0 and male voice attractiveness, the acoustic cues that mediate the relationship between voice attractiveness and FA remain unknown.

Table 14. Regression analyses examining acoustic parameters as predictors of voice attractiveness ratings made by male and female perceivers.

	β	t-value	p-value
Male perceivers			
R^2 = .19, $F(4, 29) = 1.45$, p = .25			
F_0-mean	-0.38	2.00	.05
Jit-mean	-0.05	0.19	.85
Shim-mean	-0.11	0.44	.66
F_d	0.03	0.17	.86
Female perceivers			
R^2 = .22, $F(4, 29) = 1.77$, p = .16			
F_0-mean	-0.43	2.31	.03
Jit-mean	-0.17	0.71	.48
Shim-mean	0.39	1.63	.12
F_d	0.01	0.01	.99

Discussion

Consistent with the hypothesis, perceiver ratings of voice attractiveness co-varied with speaker FA. Perceivers tended to judge the voices of male speakers that exhibited lower indices of body asymmetry to be more attractive. While this association remained stable across different listening sessions for male perceivers, female perceiver ratings of men’s voice attractiveness were found to change across the menstrual cycle. During periods of high fertility, women tended to rate the voices of men with greater body symmetry as more attractive than the voices of men with lesser body symmetry. However, during periods of low fertility, the same women showed no statistically significant tendency to rate the voices of symmetrical men as more attractive. This study provides the first evidence that for female perceivers, the preference for the voices of symmetrical speakers is modulated by cyclically shifting fertility.

These results support the notion that voices are a salient source of socially relevant information. Because FA is a marker of viability and fertility (Moller & Thornhill, 1998), the association between fluctuating asymmetry and vocal attractiveness suggests that male voices may function as an indicator of genetic and phenotypic fitness. It appears that developmental instabilities, as indexed by FA, as well as influencing attraction in the visual (e.g., Perrett et al., 1999) and olfactory domains (e.g., Gangestad and Thornhill, 1998), also influence attraction in the auditory domain.

While previous research has documented an association between male speaker FA and female perceiver judgments of vocal attractiveness (Hughes, et al., 2002; Hughes et al., 2008), the present results demonstrate that this association varies as function of the menstrual cycle. This finding is consistent with other research demonstrating shifts in attraction and sexual preferences across the menstrual cycle (e.g., Feinberg et al., 2006; Gangestad et al., 2004; Gangestad and Thornhill, 1998; Haselton & Miller, 2006; Johnston et al., 2001; Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999; Puts, 2005; Rikowski & Grammer, 1999). Indeed, it is directly analogous with findings in the olfactory domain that female perceivers at high fertility show a stronger preference for the *scent* of symmetrical men than female perceivers at low fertility (e.g., Gangestad & Thornhill, 1998; Rikowski & Grammer, 1999). The explanation typically offered for such systematic variation is that it functions to increase reproductive success by promoting the likelihood of mating with men who possess phenotypic markers of viability and fertility during the most fertile phase of the cycle. Given that FA is one such marker, the present findings are also consistent with such an explanation.

Moreover, while most previous research has focused on the convergence of male and female perceptions of vocal acoustic parameters, the present results suggest that differing adaptive problems for each sex can produce functional differentiation in vocal perception. Both heterosexual male and female perceivers are hypothesized to attend to cues of mate value in male speakers, but to do so for different reasons; males to assess the quality of potential rivals, and females to assess the quality of potential mates. The current results suggest that male perceivers attend to vocal cues of male speaker FA in a consistent manner (i.e., male perceivers evaluate vocal cues of potential same-sex rivals consistently over the different evaluation sessions), while female perceivers attend to the same vocal cues relative to cyclically changing fertility levels.

It should be noted these results cannot readily be explained by either a general increase in acoustic sensitivity or as a side-effect of an increase in sexual desire around ovulation (e.g., Pillsworth, Haselton, & Buss, 2004). Any general increase would likely have increased the vocal attractiveness of all speakers to a comparable extent. Instead, the observed relationship between judgments of vocal attractiveness and speaker FA suggests that listeners are sensitive to acoustic properties of a speaker’s voice that indicate degree of developmental stability. What these properties are remains unclear. No acoustic parameters were found to mediate the relationship between symmetry and voice attractiveness. While F_0-mean was associated with attractiveness ratings, it was not associated with FA. Indeed, contrary to Hughes et al. (2008), who found a measure of amplitude perturbation (Shim-apq11) to be correlated with male speaker FA, the present research found no acoustic measures to be associated with male speaker FA (see Study 1). Such a failure to isolate acoustic properties that can account for the variation in voice attractiveness ratings is not uncommon, with acoustic analyses typically leaving much of the variance in attractiveness ratings unaccounted for (Hughes, et al., 2008; Zuckerman & Miyake, 1993). Nevertheless, it seems that listeners are perceptive to vocal acoustic cues that reflect a speaker's level of developmental stability, and that for women, this perceptivity is mediated by hormonal influences associated with cyclically changing fertility.

Chapter 5

Study 4. Menstrual Cycle Variation in the Attractiveness of Women's Voices

Study 3 demonstrated a shift in female perceivers' attraction to male voices as a function of the menstrual cycle. At times of high fertility relative to low fertility, women showed stronger attraction for the voices of male speakers who exhibited lower levels of fluctuating asymmetry. This finding is consistent with the hypothesis derived from sexual selection theorising that women's mate choice preferences may be adaptively tuned to promote reproductive success by differentially preferring males who possess markers of high genetic fitness at different times of the cycle. By viewing the human voice as such a marker of fitness, another possible influence of hormonal variation on voice attraction can be considered. It may be that hormonal fluctuation across the menstrual cycle influences female vocal production in a way that signals hormonal status and fertility level. Although the acoustic analyses of Study 1 found only limited evidence for vocal changes across the menstrual cycle, the current study examines whether human listeners are perceptive to any changes.

Many mammalian species, including many primate species, exhibit detectable cues to ovulation (Dixson, 1998). Human females however, are typically viewed as lacking such cues (Burt, 1992; Strassman, 1982; Turke, 1984), and a number of hypotheses have been offered as explanation. It has been suggested that this "concealed ovulation" functions to increase paternity uncertainty (Benshoof & Thornhill, 1979), promote male provisioning (Alexander & Noonan, 1979), and reduce infanticide (Hrdy, 1981). Alternatively, it has been suggested that ovulatory cues became "lost" in the transition to bipedalism (Pawlowski, 1999).

Nevertheless, a number of studies have documented a variety of evidence for the existence of ovulatory cues in humans. At times of high fertility relative to low fertility, women's soft tissue features such as ears and breasts may become more symmetrical (Manning, Scutt, Whitehouse, Leinster, & Walton, 1996), women's skin colour becomes lighter (van den Berghe & Frost, 1986), men rate women's body scent as more attractive (Singh & Bronstad, 2001) and judge women's facial attractiveness to be greater (Roberts et al., 2004), women report more mate guarding behaviour from their partners (Gangestad, Thornhill, & Garver, 2002), dress in more conspicuous and attractive clothing (Haselton, Mortezaie, Pillsworth, Bleske-Rechek, & Frederick, 2007), and earn more tips when working as lap dancers (Miller, Tybur, & Jordan, 2007).

While most studies, both human and non-human, have focused on cues in the visual and olfactory domains, there has been growing interest in acoustic cues to ovulation and fertility. Female vocalizations have been found to reveal information about proximity to ovulation in elephants (Leong, Ortolani, Graham, & Savage, 2003), yellow baboons (Semple, McComb, Alberts, & Altman, 2002), and possibly Barbary macaques (Pfefferle, Brauch, Heistrmann, Hodges, & Fischer, 2008; Semple & McComb, 2000). In humans, the larynx is a hormonal steroid target, displaying a similar cytological profile to that of the genitals (Caruso et al., 2000). Estrogen has hypertrophic effects in the larynx, increasing cellular secretion and laryngeal mucosa, while progesterone increases cell viscosity and acidity, producing vocal fold edema (Abitbol, et al., 1999). In line with these physiological effects, there is evidence that circulating hormone levels can affect vocal acoustic characteristics in women. Hormonal changes associated with puberty and menopause influence vocal production (Abitbol, et al., 1999; Caruso et al., 2000), as can hormone replacement therapy (Lindholm, Vilkman, Raudaskoski, SuvantoLuukkonen, & Kauppila, 1997), some forms of hormonal contraception (Amir & Kishon-Rabin, 2004), and hormonal aberrations associated with pre-menstrual syndrome (Abitbol, et al, 1999; Chae, Choi, Kang, Choi, & Jin, 2001). Additionally, there is some evidence that vocally trained and professional singers report decreased vocal performance around menstruation (Ryan & Kenny, 2007). Together, these findings suggest a link between circulating hormone levels and histological effects in vocal production mechanisms that alter women's vocal acoustic properties.

Recent studies have attempted to identify specific acoustic cues underlying vocal changes across the menstrual cycle and to determine whether human perceivers are sensitive to any changes. Pipitone and Gallup (2008) investigated the relationship between vocal attractiveness and conception risk across the cycle, and found that attractiveness ratings increased as the risk of conception increased. No changes were found for women using hormonal contraception. One analysis of women's voice samples recorded during phases of high and low fertility found no differences in acoustic properties when examining single vowel vocalizations, but higher F_0-mean for a spoken sentence recorded at high fertility when compared to the same sentence recorded at low fertility (Bryant & Haselton, 2009). However, most studies examining vocal patterns across the menstrual cycle have found no influence of cycle phase on a range of acoustic parameters, including F_0 (Amir, Kishon-Rabin, & Muchink, 2002; Barnes & Latman, 2011; Chae et al., 2001; Meurer, Garcez, von Eye Corleta, & Capp, 2009; Raj, Gupta, Chowdhury, & Chadha, 2010; Silverman & Zimmer, 1978). One conclusion drawn from these studies was that hormonal variation across the menstrual cycle is not associated with variation in vocal properties (Barnes & Latman, 2011). Study 1 of the present research also assessed women's voices recorded at different stages of the menstrual and found no robust acoustic changes between phases of high and low fertility.

Nevertheless, these studies did not examine perceptual evaluations of women's voices. Given that acoustic analysis of human speech sounds often leaves much of the variance in subjective vocal analysis unaccounted for (Hughes, et al., 2008; Zuckerman & Miyake, 1993), it may be that human perceivers are perceptive to vocal changes across the menstrual cycle in a manner that is not readily duplicated with standard acoustic analysis techniques. Accordingly, the present study assesses whether perceivers differentially evaluate voice recordings made at different stages of the menstrual cycle. Following previous research examining perceptible changes in both visual (Roberts et al., 2004) and olfactory (Singh & Bronstad, 2001) cues across the menstrual cycle, if fertility information is available vocally, it is expected that perceivers would attend to this information in such a way as to preferentially evaluate women's voices recorded during high fertility phases of the menstrual cycle relative to voices recorded during low fertility phases.

Method

Participants. Participants in this study consisted of the same 60 participants (30 male, 30 female) who acted as perceivers in Study 3. All perceivers identified themselves as heterosexual and as having conditions that could potentially affect their hearing. This study was reviewed and approved by the University of Canterbury Human Ethics Committee. All perceivers were provided with a written information sheet (Appendices J and K) and signed a written consent form prior to participation (Appendix B).

Voice samples. Voice samples from 30 female speakers obtained in Study 1 were used in this study. The speakers ranged age from 19 to 36 years (M = 23.9, SD = 5.3). The voice samples played to perceivers consisted of speakers counting from 1 to 10 in English, at a rate of approximately one numeral per second. All speakers provided two voice samples: one recorded during a period of high fertility, and one recorded during a period of low fertility. For methodological details of speaker recruitment, fertility phase estimation, voice sampling procedures, and analysis of acoustic parameters refer to Study 1.

Apparatus and materials. Voice samples were played to perceivers using the Acoustica audio software program (www.acoustica.com) through a pair of noise attenuating headphones (Sennheiser HD-201), and responses were made on a series of rating scales (Appendix I).

Procedure. Perceivers completed listening sessions individually and during each session listened to a set of 15 female vocal recordings through a pair of headphones. The sets were constructed so that they contained a combination of high and low fertility recordings, that they contained only one recording per speaker, that each voice recording was rated an equal number of times, and that the order of presentation for the voice recordings was varied across perceivers. Immediately after hearing each sample, perceivers were asked to indicate how attractive, sexy, and pleasant they thought the voice to be.[7] Ratings were made on 7-point Likert scales ranging from 1 = very unattractive/unsexy/unpleasant to 7 = very attractive/sexy/pleasant. Perceivers were also asked whether they recognised the voices. One perceiver indicated that they thought they had recognised a speaker and the relevant ratings were omitted from analysis.

Results

Inter-perceiver agreement. For each of the two sets of voice samples, intra-class correlation coefficients were used to assess the inter-rater reliability of the three perceptual measures of voice quality – attractiveness, pleasantness, and sexiness. The intra-class coefficients are shown in Table 15, and demonstrate a high level of agreement between perceivers in their ratings of the female voices. The ratings of attractiveness, pleasantness, and sexiness were also strongly correlated, both for male perceivers (r = .83 to .90, all p < 0.001) and for female perceivers (r = .73 to .87, all p < .001), and were subsequently averaged to give a composite voice attractiveness measure for each participant.

---

[7] Perceivers were also asked to rate how masculine, feminine, and physically attractive they thought each speaker was. These questions are not directly relevant to the current considerations and are not examined here. Refer to Appendix I for further details of these additional questions.

Table 15. Inter-rater reliability measures for perceiver ratings of female voice attractiveness, pleasantness, and sexiness.

	ICC	95% CI	p-value
Attractiveness			
Set 1	.77	.56 - .91	< .001
Set 2	.83	.67 - .93	< .001
Pleasantness			
Set 1	.75	.52 - .90	< .001
Set 2	.80	.62 - .92	< .001
Sexiness			
Set 1	.85	.71 - .94	< .001
Set 2	.72	.47 - .89	< .001

Voice attractiveness ratings. Composite voice attractiveness measures were assessed by means of a 2(sex of perceiver: male/female) X 2(speaker menstrual phase: high fertility/low fertility) ANOVA with repeated measures on the second factor. As can be seen in Figure 4, a significant effect of menstrual phase was found. Voice samples recorded at high fertility (M = 4.43, SD = 0.53) were rated as more attractive than voice samples recorded at low fertility (M = 3.92, SD = 0.60), $F(1, 58) = 108.35$, $p < 0.001$, $\eta_p^2 = .65$. No significant difference between the ratings of male and female perceivers was found, $F(1, 58) = 1.34$, $p = .25$, $\eta_p^2 = .02$, nor was there a significant interaction between sex of perceiver and menstrual phase of the speaker, $F(1, 58) = 2.14$, $p = .15$, $\eta_p^2 = .04$.

Acoustic parameters across the menstrual cycle. As detailed in Study 1, four different vocal samples were recorded for each female speaker – spoken numerals, a spoken sentence, vowel sounds, and a sustained vowel sound. For each type of sample, acoustic measures for the high and low fertility phase recordings were compared using paired-samples t-tests. These results are detailed in Table 6 (Study 1) for the spoken numerals and in Appendix Tables D2, E2, and F2 for the spoken sentence, vowel sounds, and the sustained vowel sound respectively. Significant differences between fertility phases were found for the sustained vowel sample only. When producing the sustained vowel /a/, F_0-mean was greater in the high fertility phase recordings than in the low fertility phase recordings, while there was greater F_0 variation, as measured by F_0-SD, in the low fertility recordings compared to the high fertility recordings. None of the other vocal samples, including the spoken numeral count that perceivers heard in the current study, displayed any differences in the measured acoustic parameters between the high and low fertility phase recordings.

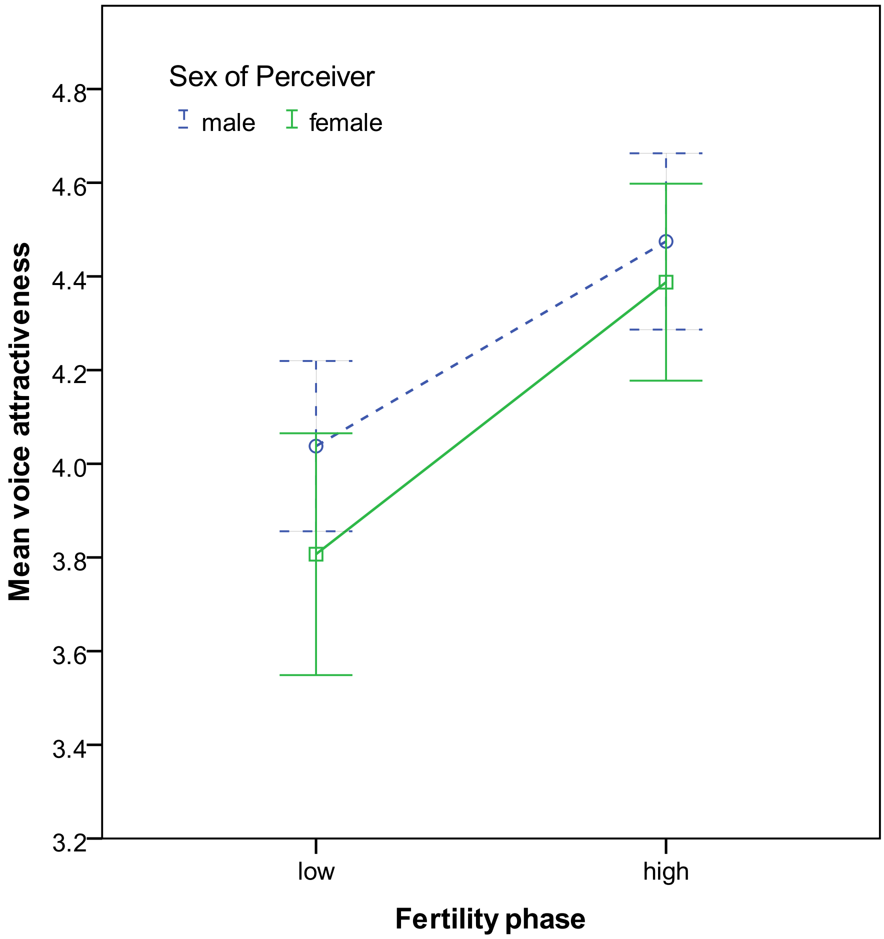

Figure 4. Mean attractiveness ratings for female voices recorded at phases of low and high fertility. Error bars represent 95% confidence intervals.

Acoustic parameters and attractiveness ratings. As displayed in Table 16, regression analyses did not identify any acoustic parameters that accounted for significant variation in female voice attractiveness ratings for either male or female perceivers.

Table 16. Regression analyses examining acoustic parameters as predictors of voice attractiveness ratings made by male and female perceivers.

	β	t-value	p-value
Male perceivers			
R^2 = .14, $F(4, 29)$ = 0.12, p = .97			
F_0-mean	-0.02	0.09	.93
Jit-mean	-0.21	0.59	.56
Shim-mean	0.14	0.40	.69
F_d	0.04	0.18	.86
Female perceivers			
R^2 = .14, $F(4, 29)$ = 1.04, p = .41			
F_0-mean	-0.18	0.91	.37
Jit-mean	-0.37	1.11	.28
Shim-mean	-0.03	0.08	.93
F_d	-0.01	0.02	.98

Discussion

Perceivers in this study were found to differentially evaluate female voice recordings made at different stages of the menstrual cycle. Both male and female perceivers rated female speakers' voices recorded during a period of high fertility to be more attractive than the same speakers' voices recorded during a period of low fertility. This effect was observed in voice samples that displayed no obvious differences in acoustic properties between the high and low fertility phase recordings.

The present finding adds to results that have been reported in other modalities whereby female faces (Roberts et al., 2004) and body scents (Singh & Bronstad, 2001) have been found to be rated as more attractive when assessed at high fertility relative to low fertility. Acoustic cues to cyclic fertility have been documented in other animals (Leong et al., 2003; Semple & McComb, 2000; Semple et al., 2002), but evidence for such cues in humans has been limited, with most studies showing little or no acoustic variation in association with cyclically changing fertility (Amir, Kishon-Rabin, & Muchink, 2002; Barnes & Latman, 2011; Chae et al., 2001; Meurer, Garcez, von Eye Corleta, & Capp, 2009; Raj, Gupta, Chowdhury, & Chadha, 2010; Silverman & Zimmer, 1978). Alongside a previous finding that vocal judgments are correlated with conception risk across the menstrual cycle (Pipitone & Gallup, 2008), the current result of increased vocal attractiveness during periods of high fertility suggests that perceivers are sensitive to acoustically propagated information regarding women's fertility status. Vocal cues to fertility may be driven by cyclic changes in hormone levels affecting vocal fold physiology (Abitbol, et al., 1999), or by such changes affecting neurotransmitter levels and influencing neuromotor laryngeal control (Higgins & Saxman, 1989). Alternatively, given that some aspects of vocal production are only broadly physiologically constrained (Titze, 1994), hormonal influences on mood could be driving vocal changes (Pipitone & Gallup, 2008).

The precise nature of the acoustic information that perceivers are attending to in their evaluations remains unclear. F_0 has previously been found to influence voice attractiveness, with higher pitched women's voices rated as more attractive (Collins & Missing, 2003; Feinberg, DeBruine, Jones, & Perrett, 2008). This preference may reflect positive, pro-social stereotypes associated with high-pitched, feminine sounding women's voices (Zuckerman & Driver, 1989). Alternatively, male preferences for high pitched voices are also potentially adaptive. Women's voice pitch is correlated with a number of reproductive health and developmental indices (Feinberg et al., 2005; Vukovic et al., 2010), such that men choosing female partners with higher pitched, more feminine sounding voices may experience greater reproductive success. However, most research has found no evidence for F_0 changes across the menstrual cycle (Amir, et al., 2002; Barnes & Latman, 2011; Chae et al., 2001; Meurer, et al, 2009; Raj, et al., 2010; Silverman & Zimmer, 1978). One recent study that systematically assessed both hormone levels and acoustic properties on a day-to-day basis across the menstrual cycle did find that for free speech vocal samples there was a significant increase in

F_0 in the days immediately prior to ovulation (Fisher et al., 2011). However, there was also a marked decrease on the day of ovulation – when the probability of conception is just as high as the preceding two days, and the pre-ovulatory increase was not significantly different to other increases in F_0 that occurred during non-fertile cycle phases. Moreover, for a sustained vowel vocal sample in the same study, F_0 was found to be lowest during the ovulatory period. In the present research an increase in F_0 at high fertility was observed for only one of the four types of voice samples analysed (see Appendix F). Moreover, perceivers in the present study differentially evaluated high and low fertility recordings with voice samples that showed no significant difference in F_0. Together, these results suggest that F_0 may be a somewhat unreliable marker of fertility and that the role of F_0 in the acoustic specification of fertility across the menstrual cycle remains equivocal.

Enhanced perceived voice attractiveness around ovulation may be beneficial for both speakers and perceivers. For speakers, enhanced attractiveness could enlarge the pool of potential mates, promote male-male competition, and generally increase the opportunity to choose high-fitness partners. Male preferences for female voices around the time of ovulation may increase the chances of mating with women during phases of high fertility and therefore promote the likelihood of conception. Consistent with previous research (Hughes, et al., 2008; Pipitone & Gallup, 2008), the present study found no differences in ratings of voice attractiveness as a function of perceiver sex. That both male and female perceivers rated high fertility phase recordings as more attractive than low fertility phase recordings, and that they did so to the same degree, suggests that male and female perceivers are equally perceptive to vocal changes across the menstrual cycle. It has been argued that male sensitivity to vocal cues of social and physical dominance in other males was driven by intra-sexual competition (Puts, Gaulin, & Verdolini, 2006), and there is some evidence that women may use voice information to evaluate potential same-sex competitors (Puts, et al., 2011). Similarly, it seems plausible that female sensitivity to vocal cues of fertility in other females could play a role in female intra-sexual behaviour and competition. Previous research has found that during periods of high fertility women report greater interest in men other than their primary partner, are more disposed to flirtatious interactions with men, and may actively seek out such opportunities (Gueguen, 2009; Haselton & Gangestad, 2006; Thornhill & Gangestad, 2008). Accordingly, for women competing for male investment, accurately tracking the fertility status of other women may be a way of monitoring potential rivals and promoting efficient allocation of intra-sexual competitive behaviours such as mate-guarding and competitor derogation.

The current finding of enhanced voice attraction at high fertility is consistent with a number of other findings demonstrating ovulatory cues in humans (e.g., Gangestad, et al., 2002; Haselton, et al., 2007; Manning et al., 1996; Miller, et al., 2007; Roberts et al., 2004; Singh & Bronstad, 2001; van den Berghe & Frost, 1986), and suggests that information on fertility status may be conveyed through multiple channels (Johnstone, 1996). More generally, the results argue against the view that women's ovulation is hidden (Burt, 1992; Strassman, 1982; Turke, 1984). Whether ovulatory cues constitute explicit signals designed to communicate fertility status, or are simply 'leaked' as a by-product of menstrual cycle physiology remains unclear. Given human tendencies for serial monogamy, ovulatory cues may have evolved to be particularly subtle, allowing women to signal fertility status to potential mates, while maintaining a degree of "plausible deniability" in order to minimise a primary partner's sexual jealousy (Miller, et al., 2007).

Chapter 6

Study 5a. The Effect of Incongruent Vocal and Visual Cues to Target Sex on Social Memory

The preceding four studies suggest that not only do human voices reliably signal myriad socially relevant features of speakers, but also that social perceivers are functionally attuned to the information content of voices in such a way as to promote adaptive social behaviour. However, as with much of the previous research considering the social perception of voices, the preceding perceptual studies examined perceiver attunement to vocal cues in isolation from other potentially informative social cues. While social perceivers do sometimes encounter voices in isolation (e.g., telephone conversations) they also frequently encounter voices in conjunction with other sources of information about a social target (e.g., face-to-face conversations). Nevertheless, little empirical or theoretical consideration has been given to how vocal information interacts with other sources of information to influence social perception. For example, despite a growing body of research into voice perception and extensive research into face perception, the amalgamation of voice and face cues in social perception is poorly understood. Given the frequency with which vocal cues are encountered in conjunction with other social cues, establishing how information conveyed in different modalities informs social perception becomes a critical component of understanding how perceivers navigate through the social world.

Several researchers have argued that voices and faces provide highly similar information regarding social targets (e.g., Belin et al., 2011; Feinberg, 2008), and there is evidence to suggest that perceptual evaluations made from voices tend to match those made from faces (e.g., Collins & Missing, 2003; Lander, 2008). However, common anecdotal evidence suggests that perceivers are sometimes quite surprised when their impressions of target individuals derived from vocal information are confounded by the physical appearance of a target. For example, a perceiver may form an impression of a target speaker from a telephone conversation or a radio broadcast, only to have that impression contradicted when meeting the target face-to-face or seeing them on television. Alternatively, a perceiver’s initial impression of a target may be derived from visual information, with subsequent vocal cues being counter to those anticipated.

That such scenarios are found to be surprising is suggestive not only of the fact that perceivers have expectancies about the relationships between a social target’s vocal and visual characteristics, but also that they anticipate these expectancies to have a degree of verisimilitude. These expectancies are presumed to derive from the physiological and social influences that produce reliable relationships – recurrent throughout both individual life-spans and human evolution – between the information content of voice and face signals, and important social dimensions such as age, sex, and attractiveness. While these relationships and the perceptions of them are far from absolute, accounting for the saliency of social perception events resulting from incongruencies between vocal and visual based information concerning a social target requires explanation.

Although previous research has not addressed this issue directly, there is some suggestion that social perceivers have expectations of speaker vocal and physical characteristics co-varying in a reliable manner. Cohen (1974) explored infants’ responses to variation in the congruency between the face and voice of the infants’ mothers and a stranger. Differences in the direction and duration of attention suggested that infants as young as 8 months may have developed expectations of reliable face-voice associations. Moreover, in several infants, incongruity between vocal and visual information appeared to cause distress. More recently it has been shown that infants as young as four months of age are sensitive to face-voice incongruencies in the age of a speaker, with infants preferring congruent experimental displays in which a child's voice is paired with a child's face, or an adult's voice is paired with an adult's face, relative to incongruent crossed-age pairings of a child's voice paired with an adult's face, or an adult's face paired with a child's voice (Bahrick, Netto, & Hernandez-Reif, 1998).

Similarly, recent studies suggest that expectations about the relationships between acoustic and visual information can influence social perception in adults. In a study examining auditory and visual integration in the perception of faces, Smith, Grabowecky, and Suzuki (2007) had participants identify whether androgynous faces were male or female. When the faces were paired with pure tones in the natural speaking range of male F_0, the faces were more likely to be judged as male, but when they were paired with pure tones in the natural range of female F_0, the faces were more likely to be judged as female.

Using natural voices as opposed to pure tones, Latinus, VanRullen, and Taylor (2010) presented perceivers with pairings of target faces and spoken words that were, with regards to target sex, either congruent (i.e., male face/male voice, or female face/female voice) or incongruent (i.e., female face/male voice, or male face/female voice). Irrespective of whether perceivers were asked to discriminate whether the pairings were congruent or incongruent, whether the face was male or female, or whether the voice was male or female, response times were impaired for the incongruent stimuli relative to the congruent stimuli.

In a similar fashion, O'Mahony and Newell (2011) familiarised perceivers with pairings of target faces and voices. They then played perceivers clips in which the familiarised faces and voices were presented either in accordance with the original pairings or in a novel pairing, as well as a control condition comprised of pairings in which the faces and voices were completely unfamiliar. Response times for deciding both whether the face or the voice were familiar were facilitated for previously viewed pairings relative to novel pairings, suggesting that prior experience regarding face and voice associations influences person recognition.

It has been well documented that not all social information is processed equally (Fiske & Taylor, 1991). For example, negative information about a target is given more weight than positive information, and extreme information is given more weight than moderate information (Fiske, 1980). One factor that has received considerable attention in social cognition research is the role of perceiver expectations regarding the target being attended to (Coats & Smith, 2006). Typically in this type of research, participants are presented with behaviours or traits that are associated with a particular individual or group, and are then asked to recall those behaviours. One element that is often of interest in such paradigms is how expectations about a social target can affect a perceiver's recall of information, particularly recall of information as a function of its relevance to the expectation.

Broadly, expectations about social targets could be expected to have one of three influences on memory. Firstly, expectations could lead perceivers to readily recall information that is congruent with their expectations. For example, Rothbart, Evans, and Fulero (1979) found greater recall for behaviours that confirmed a prior induced stereotype than for behaviours that disconfirmed that stereotype. Secondly, expectations may have negligible influence. That is, prior expectations may be easily overridden in the light of current information about a target (e.g., Locksley, Hepburn, & Ortiz, 1982). And thirdly, expectations may lead perceivers to be more likely to recall information that disconfirms those expectations. For example, Hastie and Kumar (1979) established a trait expectation for a target and then presented participants with behaviours attributed to the target that were either, irrelevant to, congruent with, or incongruent with, that expectation. In this case, recall was found to be greater for information that was incongruent with the trait expectation.

Such “incongruency effects” are typically explained as resulting from additional processing devoted to the incongruent information in order to reconcile that information with the prior expectation and form a coherent impression of the target (Coats & Smith, 2006). That is, because incongruent information is counter to the expectancy and therefore surprising, perceivers are motivated to attend to that information more thoroughly. This extra processing of the incongruent information in relation to the target is then thought to facilitate memory for that information.

A similar explanation may account for the saliency of experiences in which information about a target individual obtained from vocal acoustic cues is associated with information obtained from visual appearance cues in manner counter to expectation. It may be that little attention is paid to cases in which voice-based impressions are consistent with information obtained visually, but that cases in which expectancies are contradicted are attended to more thoroughly, with perceivers motivated to reconcile the inconsistency.

While most research examining the influence of incongruent information on social perception and memory has typically focused on behavioural/personality trait information, the current study considers how incongruent vocal and visual information influences social perception. One opportunity for incongruency relates to the vocal and visual cues that specify sex. Voices are sexually dimorphic, with males having characteristically lower and more resonant voices than females (Study 1), and social perceivers are perceptive to these differences (Study 2). Faces are also sexually dimorphic (Ferrario, Sforza, Pizzini, Vogel, & Miani, 1993), and perceivers readily distinguish between males and females on the basis of facial cues (Bruce et al, 1993). This study manipulates the relationship between target vocal and visual characteristics by pairing target faces with opposite sex voices. It is hypothesised that cases in which target characteristics are incongruent (i.e., opposite sex face and voice pairings) will be more memorable than cases in which target characteristics are congruent (i.e., same sex face and voice pairings).

Method

Participants. Twenty-eight participants (13 male and 15 female) from a research participation pool the University of Canterbury were recruited to act as perceivers. Ages ranged from 18 to 56 years (M = 25.2, SD = 9.5). This study was reviewed and approved by the University of Canterbury Human Ethics Committee, and all speakers signed a written consent form (Appendix L) and were provided with a written debriefing sheet (Appendix M).

Stimuli. Four females and four males were recruited to act as target participants. Female targets ranged in age from 21 to 24 years (M = 22.3, SD = 1.3) and male targets from 20 to 26 years (M = 23.3, SD = 2.5). Target participants were remunerated with a $5 voucher for their participation. All target participants reported having New Zealand English as their first language and as being free from any conditions that could potentially affect the sound of their voice (e.g., chronic smoking, hearing impairment, current illness). Each target was seated at a table in a noise-isolated recording room (2.5 m X 3 m) with a unidirectional microphone (Sony ECM-MS907) positioned approximately 5cm from their mouth. Targets were then asked to read aloud, in their normal speaking voice, a series of 24 short statements of neutral content (see Appendix N for the full list of statements). If the target made an error or produced any discontinuities in a statement, they were instructed to simply read that statement again. The spoken statements were recorded using a digital audio recorder (Sony Hi-MD MZ-RH10). A frontal head and shoulders photograph was taken using a digital camera (Sony DSC-P200). Each photograph was taken against the backdrop of a plain, neutral coloured wooden board with targets asked to adopt a neutral expression and to look directly into the camera. The photographs were cropped using Adobe Photoshop to 640 X

480 pixel images in which the face was centred and the eye line of each photograph kept constant.

The 24 statements read by each target were edited into individual sound files using Acoustica audio software. Each of the eight target photographs was then uniquely paired with three of the spoken statements, resulting in 24 such pairings. The pairings were designed so that there were two experimental conditions: congruent and incongruent. Congruent pairings consisted of either a female photograph paired with a female voice recording (9 pairings), or a male photograph paired with a male voice recording (9 pairings). Incongruent pairings consisted of either a female photograph paired with a male voice recording (3 pairings), or a male photograph paired with a female voice recording (3 pairings). No pairing comprised a target photograph with that same target's voice recording. This process was repeated an additional three times (i.e., four sets of voice-face pairings were developed) to ensure variation in the pairings of target faces with target voices, and in the target faces that comprised the incongruent stimuli. Additionally, photographic arrays of the target faces were developed. Each array consisted of two rows of fours faces, with the faces individually numbered one to eight. The voice-face pairings were presented on a 17-inch colour computer monitor and through a pair of noise-attenuating headphones (Sennheiser HD 201), using Microsoft PowerPoint.

Manipulation check. As this study was manipulating the congruency of male and female faces and voices, it was important to establish that perceivers were actually capable of distinguishing the face and voice stimuli on the basis of sex. Four pairings of male and female faces, and four pairings of male and female voices were presented to three volunteers who were asked to identify which of the pair was male and which was female. Across all the evaluations no errors were made, suggesting that perceivers could readily determine the sex of the face and voice stimuli.

Procedure. Perceiver participants were invited to take part in a study that was described as an investigation into impression formation. They were initially asked if they were experiencing any conditions that could affect their hearing or vision. All perceivers reported being free from such conditions. Perceivers were instructed that they would be seeing a series of photographs of individuals, each of which would be paired with a spoken sentence about that individual, and that their task was to watch the entire series and form an impression of the individuals. The experimenter queried the perceivers as to their understanding of the instructions, and then left the room. The instructions for the task were presented again on the computer screen, and the series of voice-face pairings began with a key-press. Perceivers experienced each of the 24 pairings for ten seconds each, with the order of pairings randomized across perceivers.

After experiencing each of the pairings, perceivers completed a 1 minute distracter task in which they were asked to list as many countries of the world as they could. They were then presented with a randomized list of the 24 statements and a numbered photographic array of the target faces, and asked to write the number of the target next to the sentence that each target was paired with (see Appendix O). Perceivers were instructed to place a number by all of the statements and that if they were unsure that they were to guess. Finally, perceivers were asked to indicate if they recognised any of the targets. They were then debriefed as to the purpose of the experiment, and thanked for their participation.

Results

The number of correct congruent and incongruent responses was recorded for each perceiver. A correct response was defined as accurately identifying a previously experienced target face and statement pairing. To control for the greater number of pairings in the congruent condition, the proportion of correct responses in each condition was computed as the basis for analysis. Responses were analysed using a 2(condition: congruent/incongruent) X 2(sex of perceiver: male/female) ANOVA with repeated measures on the first factor. Analysis revealed a significant main effect for condition, $F(1, 26) = 11.3$, $p = .002$, $\eta_p^2 = .30$. As can be seen in Figure 5, perceivers made a higher proportion of correct face-statement pairings in the incongruent condition ($M = 0.71$, $SD = 0.22$) than they did in the in the congruent condition ($M = 0.53$, $SD = 0.13$). No significant difference was found between male and female perceivers, $F(1, 26) = 2.9$, $p = .10$, $\eta_p^2 = .10$, nor was there a significant interaction between perceiver sex and congruency condition, $F(1, 26) = .33$, $p = .57$, $\eta_p^2 = .01$. In line with the hypothesis, the results suggest a recognition advantage for pairings from the incongruent condition. That is, perceivers were more likely to correctly associate a statement with a target face when the statement was made in a voice of the opposite sex to the face, than when the statement was made in a voice of the same sex as the face.

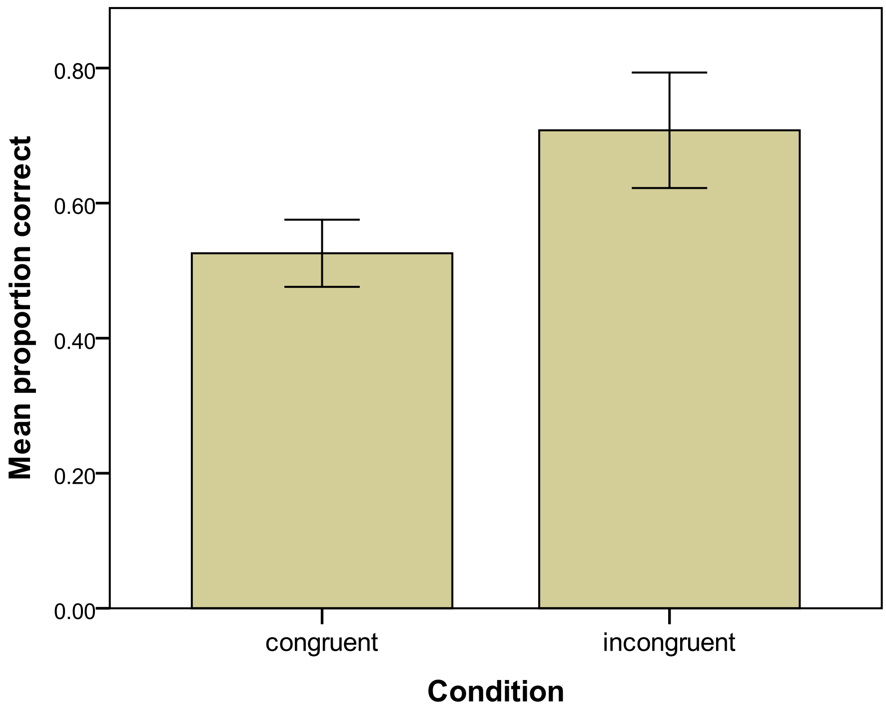

Figure 5. Bar graph depicting the proportion of correct responses in the congruent (same sex) and incongruent (opposite sex) conditions. Error bars represent 95% confidence intervals.

Study 5b. The Effect of Incongruent Vocal and Visual Cues to Target Age on Social Memory

Study 5a demonstrated an incongruency effect whereby perceivers were more able to correctly identify previously experienced face-statement pairings when the statements were made in a voice that was of opposite sex to the face. In doing so, this experiment drew upon the reliably occurring distinctions between both the voice and face characteristics of males and females. While sex differences in voice and face characteristics are typically robust and salient, vocal and visual characteristics also reliably vary with other speaker-specific features in ways that perceivers are sensitive to, and may therefore also serve as a basis for expectations of co-variance in social perception. One such feature is speaker age. There are characteristic changes in voice and face that occur across the life-span such that perceiver estimates of age from vocal (Study 2) and visual (Burt & Perrett, 1995) information accurately reflect the chronological ages of target individuals. The current study replicates and extends the previous study through manipulating the relationship between target vocal and visual characteristics by pairing faces and voices from targets of different ages. It is hypothesised that cases in which target characteristics are incongruent (i.e., different aged face and voice pairings) will be more memorable than cases in which target characteristics are congruent (i.e., same aged face and voice pairings).

Method

The experimental method employed for Study 5b was identical to that of Study 5a, with the following exceptions.

Participants. Twenty-two participants (9 male & 12 female) were recruited from the University of Canterbury to act as perceivers. Ages ranged from 19 to 39 years (*M* = 24.2, *SD* = 5.3).

Stimuli. Eight females, four younger and four older, were recruited to act as target participants. The four younger targets were the same four females used in Study 5b. Their ages ranged from 21 to 24 years (*M* = 22.3, *SD* = 1.3), while the four older targets were aged between 55 and 60 years (*M* = 57.5, *SD* = 2.1). Congruent pairings consisted of either an older female photograph paired with an older female voice recording (9 pairings), or a younger female photograph paired with a younger female voice recording (9 pairings). Incongruent pairings consisted of either an older female photograph paired with a younger female voice recording (3 pairings), or a younger female photograph paired with an older female voice recording (3 pairings). No pairing comprised a target photograph with that same target's voice recording.

Manipulation check. Four pairings of younger and older female faces, and four pairings of younger and older female voices were presented to three volunteers who were asked to identify which of the pair was younger and which was older. Across all the evaluations no errors were made, suggesting that perceivers could readily determine the differences in the ages of the face and voice stimuli.

Results

The number of correct congruent and incongruent responses was recorded for each perceiver. A correct response was defined as accurately identifying a target face and statement pairing. To control for the greater number of pairings in the congruent condition, the proportion of correct responses in each condition was computed as the basis for analysis. Responses were analysed using a 2(condition: congruent/incongruent) X 2(sex of perceiver: male/female) ANOVA with the first factor as a within subjects factor. This revealed a significant main effect for condition, $F(1, 20) = 5.9$, $p = .025$, $\eta_p^2 = .23$. Figure 6 illustrates that the proportion of incongruent face-statement pairings correctly identified was greater in the incongruent condition (M = 0.66, SD = 0.26) than in the congruent condition (M = 0.51, SD = 0.06). ANOVA revealed no significant main effect for perceiver sex, $F(1, 20) = 0.26$, $p = 0.62$, $\eta_p^2 = .01$, nor any significant interaction effect between perceiver sex and condition, $F(1, 20) = 0.57$, $p = 0.46$, $\eta_p^2 = .03$. Consistent with the previous study, the results suggest a recognition advantage for pairings from the incongruent condition. Perceivers were more likely to correctly associate a statement with a target face when the statement was made in a voice of a different age to the face, than when the statement was made in a voice of a similar age to the face.

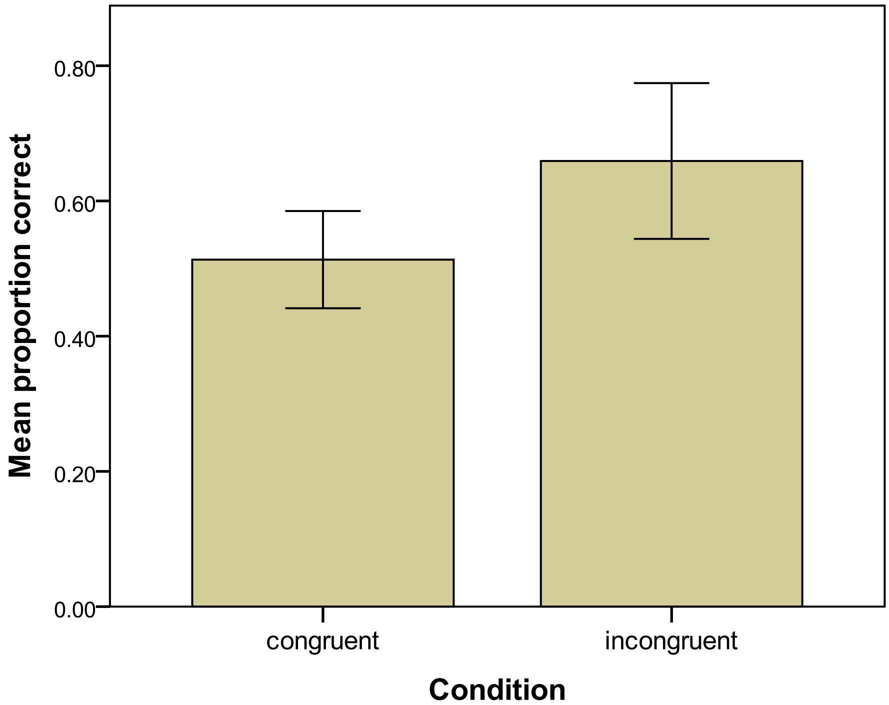

Figure 6. Bar graph depicting the proportion of correct responses in the congruent (same age) and incongruent (different age) conditions. Error bars represent 95% confidence intervals.

Discussion

The results of these two studies suggest a recognition advantage for incongruent target face-voice pairings over congruent target face-voice pairings. In Study 5a, perceivers were more successful in matching target statements with target faces when the statement had initially been presented in the form of opposite sex face-voice pairings than in the form of same sex face-voice pairings. In Study 5b, a similar result was observed with statements presented in incongruently aged face-voice pairings producing a higher proportion of correct responses than statements presented in congruently aged face-voice pairings.

This incongruent face-voice advantage extends previous work examining the well documented influence of incongruent information on person memory to a novel domain. Instead of incongruency between trait expectancies and behavioural/personality descriptions, the present studies explored incongruency between two fundamental features of a social target: vocal acoustic characteristics and facial visual characteristics. Nevertheless, the current results appear amenable to a similar explanation to that typically employed to explain the more conventional incongruency effect in that they point to additional processing being given to expectancy incongruent cases. That is, if perceivers expect particular types of target voices (e.g., female voices) to be concordant with particular types of faces (e.g., female faces), then any violation of this expectancy may warrant additional cognitive effort in order to process the discordance. This additional processing may then facilitate the greater recognition of the target-statement associations that was observed in the incongruent conditions.

This interpretation is dependent on the assumption that perceivers do indeed possess expectations about co-variance in the relationships between a social target's vocal and physical characteristics. This expectancy was not introduced as a part of the experimental protocol and was not explicitly tested for. Nevertheless, the robustness of the relationships between speakers' vocal and facial characteristics and their sex and age, and perceiver sensitivity to these relationships, along with previous research documenting the effects of vocal and visual integration in social perception (e.g., Bahrick, Netto, & Hernandez-Reif, 1998; Latinus, VanRullen, & Taylor, 2010; O'Mahony and Newell, 2011; Smith, Grabowecky, & Suzuki, 2007), and the recognition advantages for the incongruent conditions observed in the current two studies are suggestive that such expectancies may be present. Moreover, unlike research considering the influence of incongruent behavioural/personality information on person memory, in which incongruency effects are typically not observed unless expectancies about the social target are established during the experiment (Stangor & McMillan, 1992), the current studies found incongruency effects for cases in which an expectancy was not explicitly established as part of the study. It may be that broad expectations about the relationships between a social target's physical and vocal characteristics are more fundamental and less easily over-ridden than many expectations based on stereotypical behavioural or disposition related information.

However, caution needs to be exercised in drawing comparisons with previous research. As noted, the current two studies examined the effects of discordant face and voice information rather than discordant trait and behavioural information. That is, in most previous research the expectancy is in terms of some underlying psychological disposition or trait, and perceivers are presented with behavioural information that is inconsistent with that disposition. In the present research it is the face and voice of targets that is inconsistent and the behavioural information is neutral. Thus, instead of assessing the recall of information about a particular target, the measure of interest was recognition of target statement and face pairings, with perceivers displaying superior recognition for pairings when the statement was originally presented in a voice that provided sex and age cues discordant with those provided by the face.

Although it could be argued the current results are not due to the incongruency between target faces and voices as such, but to the incongruent items being a minority in the stimulus set, it is the incongruity that makes these items the minority. That is, in order to be tagged as unusual, perceivers need to notice the incongruity between the paired faces and voices. That perceivers apparently did notice the incongruity suggests that they expect vocal and visual cues specifying age and sex to covary in a reliable manner.

Previous research considering the perceptual integration of acoustic and visual cues has mostly focused on speech perception. Typically studies have found that matched acoustic and visual information facilitates speech perception while mis-matched information has a detrimental effect. For example, speech intelligibility is generally increased when a speaker's face is visible (Belin et al., 2011), while in the well-known "McGurk effect," mismatched acoustic and visual information produces illusory perceptions of spoken sounds (McGurk & MacDonald, 1976). By contrast, the present research investigated the effect of vocal and visual incongruity on social memory. Recognition of pairs of previously encountered face-statement pairings was facilitated when the statement was initially presented in a voice incongruent with the face. This effect was found both when the incongruency was between the sex of the voice and the face (Study 5a), and when the incongruency was between the age of the voice and the face (Study 5b), demonstrating the generality of the finding. That the effect was slightly stronger when the incongruency was in terms of target sex than in terms of target age is perhaps not surprising given the robust dimorphism between male and female voices and the degree to which perceivers appear sensitive to the differences (see Study1 and Study 2). Thus, while perceivers appear to expect congruity in the facial and vocal information that specifies both sex and age, the expectancy may be greater for sex than for age.

Although the current study manipulated the congruity of cues to sex and age, human voices and faces also vary in other dimensions that may serve as a basis for incongruity effects. For example, perceivers readily make judgments about the attractiveness of voices and faces, and previous research has found these judgments tend to be positively correlated, with social targets that are rated as having attractive voices also rated as having attractive faces (e.g., Collins & Missing, 2003). How vocal and visual congruency in terms of attractiveness influences social perception has yet to be examined. Moreover, the present study manipulated vocal and facial incongruity in a somewhat heavy-handed and obvious manner whereby perceivers could readily distinguish between target voices and faces in terms of age and sex. However, both voices and faces vary in a relatively continuous fashion, and more subtle manipulations of the degree of incongruency may allow for the investigation of more naturalistic and ecologically valid variations in vocally and facially conveyed information, as well as allowing for examination of the degree of incongruency that perceivers are sensitive to. Similarly, while the current study used static faces, dynamic faces in the form of video clips time-synchronised with voice recordings would provide more naturalistic voice-face pairings.

Establishing exactly how voice-based expectations influence social behaviour remains an important task for future research. One possible avenue is the influence of vocal acoustic information on stereotype based social attribution and inference. While most research has focused on visual cues to sex, age, race, and the like, as a basis for social categorization, vocal cues may also be salient features by which perceivers categorize social targets. For example, Ko, Judd, and Blair (2006) demonstrated that vocal properties that signal an individual as either male or female can serve as a basis for stereotyped social evaluations. Moreover, because the vocal cues that readily distinguish male from female speakers also vary within each sex, stereotyped social evaluations were also demonstrated to apply to individuals within the categories of male and female.

A further issue is the extent to which vocal and visual incongruency affects the valence of perceiver judgments about a social target. To the extent that perceivers expect acoustic and visual information to naturally co-vary, discordant voice and face information may be seen as unnatural, unhealthy, or disordered, and subject to harsher social evaluations than concordant information. In line with this, Mitchell et al., (2011) found that incongruence in the human realism of faces and voices elicits evaluations of eeriness. In their study, a human with a synthetic voice and a robot with a human voice both resulted in heightened perceptions of eeriness relative to the reverse pairings.

The integration of the information conveyed by faces and voices is an important aspect of social perception. Previous research has mostly considered the effects voice and face interaction in the context of speech perception. The current findings suggest that vocal and visual incongruency has an effect on social memory. Perceivers were more likely to correctly associate previously heard target-related information with a particular target face when that information was originally presented in a voice incongruent with the face, relative to when the information was presented in a voice congruent with the face. To the extent that perceivers expect information conveyed in the face to be concordant with information conveyed in the voice, this recognition advantage may be explained by perceivers attending to the discordant cases more thoroughly in order to reconcile the discrepancy and form a coherent impression of the target. This may account for the saliency of real-world social perception events in which impressions derived from the assessment of a target individual in an initial informational channel are confounded by the information obtained from a secondary channel.

Chapter 7

General Discussion

A functional approach to social perception suggests that perceivers should be attuned to the properties of others for which the correct, or incorrect, detection of would have consequences for adaptive social behaviour. Fundamental physical characteristics are among the most consequential of such properties. Differences in the basic physical composition of individuals afford different behavioural opportunities and the accurate detection of these opportunities is likely to provide direct adaptive benefit to social perceivers. As such, adaptive social behaviour is contingent on the ability to correctly identify the sex, age, and physical size of other individuals, and to make evaluations of their attractiveness that covary with markers of health and fertility. For example, a social perceiver is likely to be better served in attempting to nurture the young, mate with the reproductively mature and healthy, and acquiesce to the physically large, than in attempting the opposite. Critical to such adaptive behaviour is the attunement of perceivers to the information specifying an individual's physical characteristics. Drawing on this functional perspective of social perception, the current thesis makes an argument for the attunement of social perceivers to the vocal specification of speaker physical characteristics. This issue has been addressed in the present research by considering how information specifying physical characteristics can be reliably encoded in vocal signals, whether perceivers are attuned to such information, and whether any attunement is useful for guiding social behaviour in an adaptive manner.

An initial argument was presented suggesting that anatomical and physiological constraints on vocal production mechanisms structure the physical properties of vocal signals in such a way that they contain reliable, and functionally perceivable, information about the sex, age, body configuration, and hormonal profile of speakers. Considerations of the source-filter theory of vocal production (Fant, 1960; Titze, 1994) maintain that a source signal is produced by the mechanical actions of the vocal folds. The vocal folds themselves are directly influenced by the sex, age, and hormonal status of a speaker, and as such, the source signal is believed to be structured in such a way that it reliably encodes information pertaining to these characteristics. The source signal is then filtered by the vocal tract. This filtering process is dependent on the size and shape of the vocal tract, which is related to overall skeletal size. As such, the vocal tract is believed to structure vocal signals in such a way that they reliably encode information pertaining to speaker body size.

In line with this approach to understanding the information content of vocal signals, an initial study investigated relationships between the vocal and physical characteristics of a group of 30 female speakers and 30 male speakers. Voice recordings were made from four different vocal tasks and across two testing sessions. For female speakers these sessions were held at high and low fertility phases of the menstrual cycle, while for male speakers they were held one week apart. Both the vocal and physical measures were comparable to values reported in previous research (e.g., Baken & Orlikoff, 2000; Evans et al., 2006; Gallup et al., 2007; Gonzalez, 2004; Hughes et al., 2004; Manning, 2002; Puts et al., 2011; Rendall et al., 2005). Acoustic analyses found vocal signals to vary between speakers, but to be relatively stable within speakers, showing a strong degree of consistency across different vocal tasks and across different recording sessions. Such findings are assumed to underpin much voice perception research but are seldom given explicit consideration. Robust differences in acoustic properties between male and female speakers were observed, with males having significantly lower fundamental and formant frequency values than females. Previous research has identified consistent acoustic correlates of age (Baken & Orlikoff, 2000), but only two formant measures were associated with age in the present research, and only for female speakers. The restricted age range of speakers in the present research may account for this lack of association. Acoustic correlates of speaker body size appear more difficult to identify, with previous research producing highly inconsistent findings (Bruckert et al., 2006; Collins, 2000; Collins & Missing, 2003; Evans et al., 2006; Gonzalez, 2004; Gonzalez, 2007; Lass & Brown, 1978; Rendall et al., 2005; Sell et al., 2010; Van Dommelen & Moxness, 1995). Central to much of this research has been the notion of formant frequencies as valid indicators of speaker size due to their dependence on the size of the vocal tract. However, the standard formulation of formant structure that has most frequently been examined, F_d, has recently been criticised as providing only limited information regarding the structured relations between formants (Puts et al., 2011). As such, the current research included a recently proposed index, F_p, which has been argued to be a stronger acoustic correlate of speaker size (Puts et al., 2011). Nevertheless, in the present research F_p showed no association with any measure of speaker body size. While as yet unidentified formulations of formant structure could provide better indices of speaker size, it is also possible that characteristics of the source signal may combine with characteristics of the filter to specify speaker size in ways that are not yet well understood (Taylor & Reby, 2010). It was also suggested that due to common developmental paths, physical properties such as WHR, SHR, physical strength, and 2D:4D may covary with vocal parameters. While little systematic research has addressed this issue, in the present research, relationships between these physical characteristics and acoustic parameters were negligible.

Study 2 investigated the ability of perceivers to utilise vocal information to make judgments about the physical characteristics of speakers. After listening to vocal samples of speakers counting from one to ten, perceivers in this study were asked to estimate the speaker’s sex, age, height, and weight. Perceiver estimates displayed a high level of consensus and for the most part accurately reflected the true characteristics of speakers. Perceivers were able to accurately determine the sex of the target speakers with complete accuracy. Across 1200 individual responses, no errors were made in perceiving whether a speaker was male or female. Moreover, with the exception of female speaker weight, mean perceiver ratings of speaker age, height, and weight showed moderate to strong positive correlations with the true values of the speaker characteristics. While most previous research assessing the ability of perceivers to determine speaker physical characteristics from vocal information has averaged across perceiver responses, this approach speaks to the accuracy of a group of perceivers rather than to the accuracy of individual social perceivers. As such, the present research also assessed perceiver accuracy by calculating correlations between estimated and actual values for each individual perceiver. While this method produced relationships of a lower magnitude than pooling perceiver responses, the same pattern of relationships between estimated and actual values was observed with perceiver estimates generally significantly and positively correlating with speaker age and body size measures. Although these relationships were weaker than the pooled estimates, they more accurately reflect the functional nature of social perception, and suggest that at the level of the individual, social perceivers are attuned to the vocal specification of speaker sex, age, and body size.

Study 3 and Study 4 investigated how perceiver judgments of vocal attractiveness are influenced by hormonal and fertility variations in both speakers and perceivers. Previous research has found vocal attractiveness ratings to negatively covary with measures of FA in both male and female speakers (Hughes et al., 2002). In Study 3 this finding was replicated with both male and female perceivers rating the voices of male speakers with lower indices of FA as more attractive. However, while this relationship was found to remain stable across listening sessions for male perceivers, for female perceivers the relationship between male speaker FA and voice attractiveness varied as a function of the menstrual cycle. Vocal attractiveness judgments made by female perceivers were significantly negatively correlated with male speaker FA when made during periods of high fertility, but not when made during periods of low fertility. This study provides the first evidence that for female perceivers, the preference for the voices of symmetrical speakers is modulated by cyclically shifting fertility. This result is directly analogous with findings that female perceivers at high fertility relative to low fertility show a stronger attraction to the body odour of males exhibiting low FA (Gangestad & Thornhill, 1998; Rikowski & Grammer, 1999). This latter effect has been interpreted as evidence of a functional shift in female preferences that promotes the likelihood of mating with, and obtaining genetic investment from, males who possess phenotypic markers of viability and fertility, when the likelihood of conception is greatest (Gangestad & Thornhill, 1998). Applied to the present research, this interpretation suggests that by displaying a stronger attraction to the voices of low FA male speakers during phases of high fertility, female perceiver attunement to the vocal specification of FA is calibrated in such a way as to promote reproductive success.

Study 4 further examined the influence of fertility changes across the menstrual cycle on voice attractiveness, but did so with regards to changes in speakers rather than perceivers. Given the extent to which hormonal factors appear to influence vocal mechanisms, natural hormonal variation across the menstrual cycle may alter female vocal signals in a perceptible manner. Indeed, both male and female perceivers were found to rate female speaker voices recorded during phases of high fertility to be more attractive than voices recorded during phases of low fertility. This finding suggests that perceivers are attuned to vocally propagated information revealing of female speakers' cyclically shifting fertility status. Such an attunement may increase the chances of male perceivers mating with women during phases of high fertility, and may direct female perceivers toward monitoring threats from potential sexual rivals. Taken in conjunction with similar findings showing increased attraction for female faces (Roberts et al., 2004) and body scents (Singh & Bronstad, 2001) when assessed at high fertility relative to low fertility, the present result suggests that cues to cycling female fertility are present in multiple modalities.

Study 2 demonstrated that perceivers were attuned to vocal cues specifying the sex and age of a social target. Studies 5a and 5b sought to extend these findings by considering how vocal cues specifying sex and age interact with visual cues specifying sex and age to influence social perception. By manipulating the degree of congruence in the sex and age information conveyed about a social target through the target's face and voice, it was found that social perceivers exhibited greater recognition for cases in which information provided by the face and voice was incongruent relative to cases in which the information was congruent. To the extent that perceivers expect information conveyed in the faces and voices of social targets to be concordant, it was argued that this incongruency effect could be explained by perceivers attending to the incongruent cases more thoroughly than the congruent cases in order to reconcile the perceived discrepancy and form a coherent impression of the target. It was also suggested this explanation could account for the saliency of perception events in which the cues provided by a social target's visual and acoustic properties are counter to expectation. More generally, this study provides a novel contribution to understanding how information from multiple sources combines to influence social perception.

A number of important themes run throughout these findings. Firstly, the results highlight the functional nature of voice perception. Variation in the physical constitution of different individuals is likely to afford social perceivers differing opportunities for behaviour and interaction. As such, there is considerable adaptive value for social perceivers in being able to accurately identify the physical features of conspecifics that are relevant to guiding adaptive social interactions. Broadly, for a given social perceiver, a person of a particular sex, age, or physical stature may afford courtship, nurturing, or domineering respectively, whereas as a person of a different sex, age, or physical stature may not afford those same behavioural opportunities. Accurately detecting the physical characteristics of social targets is likely to be a critical first step in guiding interactions with that target. Findings of the current research suggest that perceivers are able to utilise vocal cues to make judgments about sex, age, and body stature that accurately reflect the true characteristics of speakers, and are therefore likely to provide a basis from which to guide adaptive social behaviour.

Moreover, the present results point to the functional nature of vocal attractiveness judgments. In line with previous research showing that women's judgments of attractiveness for a variety of male characteristics change across the menstrual cycle (e.g., Feinberg et al., 2006; Gangestad et al., 2004; Gangestad & Thornhill, 1998; Haselton & Miller, 2006; Johnston et al., 2001; Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999; Puts, 2005; Rikowski & Grammer, 1999), the present research found that women favour the voices of male speakers with low indices of FA at times of high fertility relative to times of low fertility. This finding is consistent with the notion that menstrual cycle shifts in attractiveness judgments function to promote the likelihood of mating with men who possess phenotypic markers of viability and fertility during the most fertile phase of the menstrual cycle. Such an explanation has been used in previous voice perception research to account for the finding that women prefer masculinized male voices (i.e., voices lowered in F_0) more so in high fertility phases of the menstrual cycle than in low fertility phases (Feinberg et al., 2006). However, Wells, et al. (2009) have questioned this interpretation on the basis that F_0 is an indicator of testosterone and masculinity, and that more masculine men are less likely to engage in the investment and support of children (e.g., Gangestad & Simpson, 2000; Gray, Kahlenberg, Barrett, Lipson, & Ellison, 2002). That is, cyclic variation in preferences for male F_0, in which there is a decreased preference for low F_0 voices during low-fertility menstrual phases, could reflect a strategy of selecting for indicators of investment potential rather than for indicators of heritable genetic potential. However, this possibility seems unlikely to account for the cyclic change seen in the present research. Unlike the previous research examining menstrual cycle variation in women's vocal judgments, the present research did not examine changes in attractiveness judgments as a function of male speaker F_0, but rather as a function of male speaker FA. Moreover, FA was found to be unrelated to F_0 (see also Hughes et al., 2008), suggesting that judgments of vocal attractiveness covary with speaker FA somewhat independently of F_0. Therefore, regardless of whether FA is related to male investment potential, increased attraction to the voices of male speakers with low indices of FA (a known, heritable marker of phenotypic and genotypic quality) at times of high fertility provides strong support for the notion of adaptive mate choice via the process of “good genes” sexual selection (Gangestad & Thornhill, 1998).

The current results also support a functional account of attractiveness evaluations of female voices. Study 5 found that the voices of naturally cycling women were rated as more attractive when assessed during menstrual phases of high fertility than when assessed during phases of low fertility. Moreover, this effect was seen for ratings made by both male and female perceivers. For male perceivers, preferentially evaluating women’s voices during high fertility phases may increase the chances of mating with women during these times, therefore increasing the likelihood of conception, while for female perceivers, this differential evaluation may be useful in assessing the threat of potential sexual rivals.

A second recurrent theme in the present research is the level of consensus in perceiver judgments. As indicated by high measures of inter-perceiver reliability, perceivers showed strong levels of agreement in their judgments of speaker sex, age, body size, and vocal attractiveness. This suggests that perceivers are attuned to the vocal specification of these features in highly similar ways. In line with this, the judgments made by male and female perceivers were highly similar. For heterosexual social perceivers, same-sex and opposite-sex conspecifics both afford opportunities for either co-operation or competition, whereas only opposite-sex individuals afford opportunities for mating. Given that adaptive problems differ for male and female perceivers, to the extent that traits relevant to intra-sexual competition and traits relevant to inter-sexual attraction are signalled by different vocal features, male and female perceivers may be expected to attend to vocal parameters differentially. For example, Hodges-Simeon et al., (2010) found male and female perceivers to differentially attend to variance in vocal parameters in male voices when making dominance and attractiveness evaluations. However, the present results suggest that for the most part, vocal cues to sex, age, body stature, FA, and female fertility are attended to in highly similar ways by both male and female perceivers. The one exception to this was evidenced in Study 4; differences in the fertility status of female perceivers across time was associated with differences in attractiveness judgments of male voices as a function of speaker FA, whereas male perceiver judgments were found to be consistent over time. It was suggested that this difference was likely to be functional. Female perceivers may benefit from attending to male cues of high heritable fitness more so when their probability of conception is high rather than low, but male perceivers may benefit from a consistent attunement to the fitness cues of potential same-sex rivals. An interesting prospect for future research could be to examine female perceiver judgments of female voices as function of speaker FA. If such judgments were found to remain stable across the menstrual cycle (i.e., female perceivers were consistently attuned to vocal cues of FA in other women, regardless of their own cycle phase) this would lend support to the notion that the present findings of a stronger association between vocal attractiveness ratings and male speaker FA at times of high fertility compared to low fertility may function in the service of adaptive mate choice.

A third recurrent pattern in the present research is the failure to isolate the physical qualities of speakers' voices that are informing perceiver judgments. Indeed, this is a theme common to many studies of voice perception, with acoustic analyses typically leaving much of the variance in the perceptual assessment of human voices unaccounted for (e.g., Bruckert et al. 2006; Collins, 2000; Hughes et al., 2008; Sell et al., 2010; Zuckerman & Miyake, 1993). The approach taken in the current research was to employ considerations of the source-filter theory of vocal production (Fant, 1960; Titze, 1994) and previous empirical work in order to identify acoustic parameters that were likely to legitimately encode information regarding socially important physical and biological characteristics of speakers. The relationships between speaker vocal and physical properties were then assessed, and regression analyses used to identify the acoustic cues accounting for variation in perceiver assessments of the various speaker characteristics. Using this approach it was hoped to be able to identify the acoustic cues, both valid and invalid, that perceivers either do or do not attend to in their assessments. Robust differences in the vocal acoustic properties between males and females were identified, and regression analyses suggested that perceivers used these as legitimate cues in accurately perceiving the sex of speakers. Previous research has identified age-related changes in vocal acoustics (e.g., Baken & Orlikoff, 2000), however in the present research, despite positive relationships between the perceived and actual ages of speakers, no acoustic parameters were found to significantly account for perceptions of age. F_0-mean significantly predicted perceived height, but not when analysed separately for male and female speakers. Moreover, F_0-mean was not significantly related to height for either male or female speakers.

Similarly, the acoustic cues accounting for perceiver evaluations of vocal attractiveness remain unaccounted for. F_0-mean was related to male speakers' voice attractiveness, but ratings of attractiveness appeared to track speaker FA independently of any of the analysed acoustic parameters, including F_0-mean. Despite perceptions of female speakers' vocal attractiveness changing in accordance with cyclic fertility, none of the assessed acoustic parameters accounted for perceiver judgments of attractiveness, nor were any of the parameters in the vocal samples that perceivers listened to found to change between phases of low and high fertility. These findings contribute to the somewhat confusing picture of the acoustic features that constitute attractive voices. For example, Daniel and McCabe (1992) found voices in the mid-range of F_0 to be rated as most attractive for both male and female speakers, suggesting that voices that depart from average are perceived as unattractive. Similarly, Bruckert et al., (2010) found that averaging across individual voices through acoustic morphing increases vocal attractiveness, mirroring a well-known effect in face perception research (e.g., Langlois & Roggman, 1990). However, other research has found perceivers to prefer female voices with higher F_0 to voices with average F_0 (Feinberg, DeBruine, Jones, & Perrett, 2008). Similarly, several studies have reported male voices with lower F_0, and female voices with higher F_0 to be perceived as more attractive (Collins, 2000; Collins & Missing, 2003; Feinberg, DeBruine, Jones & Perret, 2008; Feinberg et al., 2005; Zuckerman & Miyake, 1993), while other studies have reported both male and female voices with lower F_0 to be attractive (Oguchi & Kikuchi, 1997), both male and female voices with higher F_0 to be attractive (Oskenberg, Coleman, & Cannell, 1986), and no relationship at all between F_0 and voice attractiveness for either male or female speakers (Hughes et al., 2008). Feinberg (2008) has suggested that some of these discrepancies in identifying the physical components of vocal attractiveness may be dependent on methodological variations between studies. One such variation may relate to individual perceiver differences and the particular context in which perceiver judgments take place. For example, the acoustic cues attended to by single, heterosexual, naturally cycling women during phases of high fertility, judging male vocal attractiveness in the context of explicit consideration of short-term sexual encounters, may be quite different to the cues attended to by elderly women judging male vocal attractiveness in the context of a telephone counselling service.

Voices are a complex amalgam of multiple features and analyses of these features in isolation, as was largely the case in the current research, may fail to capture important interactive or configural relationships. It seems possible that alternative and as yet undeveloped formulations of indexing acoustic properties may prove useful in identifying the acoustic information that perceivers are attending to in the vocal assessments of social targets (e.g., Puts et al., 2011). Indeed, some researchers have taken the lack of success in the identification of the physical properties that inform voice perception to suggest that human perceivers have a neuro-cognitive architecture that contains sophisticated auditory processing mechanisms that the individual parameter assessments of current acoustic analysis methods do not readily duplicate (e.g., Hughes et al., 2008; Zuckerman & Miyake, 1993).

The findings of the current research should be interpreted in the light of a number of methodological considerations. While menstrual cycle phase for both female speakers and perceivers was estimated using the reverse counting method (Lenton, Landgren, & Sexton, 1984; Thornhill & Gangestad, 1999), direct hormone assays would improve the accuracy of fertility estimates. Indeed, in research that assessed luteinizing hormone and progesterone peaks to verify ovulation, approximately one-third of women do not show hormonal evidence of ovulation having occurred (Gangestad, Thornhill, & Garver, 2002; Haselton, Mortezaie, Pillsworth, Bleske-Rechek, & Frederick, 2007). Nevertheless, in the present research any measurement error in assessing female fertility should work against the likelihood of detecting effects as function of the menstrual cycle. That is, if a female participant was erroneously identified as being in a state of high fertility, any effects, both in vocal production and perception, should be weaker than if fertility status was correctly identified. More accurate determination of fertility may actually increase the strength of the menstrual cycle effects identified in the current research.

The use of hormone assays to determine more precise fertility estimates would also allow more subtle examination of hormonal influences. Feinberg et al. (2006) for example, found that menstrual cycle shifts in preferences for masculine voices were most pronounced for women with lower rather than higher levels of estrogen metabolites, suggesting that vocal attractiveness judgments are contingent on variations in the femininity of female perceivers.

Similar findings have been reported in the face perception literature, with attraction to male facial masculinity influenced by the attractiveness and femininity of female perceivers (Little, Burt, Penton-Voak, & Perrett, 2001; Penton-Voak et al., 2003). Recent research has also found comparable effects in men. Male perceiver attractiveness has been found to influence preferences for female facial femininity, with more attractive men exhibiting stronger preferences for facial femininity than less attractive men in short-term relative to long-term relationship contexts (Burriss, Welling, & Puts, 2011). Such findings speak to the general importance of examining individual difference factors in further elaborating vocal perception phenomena. How factors such as relative mate-value, socio-sexuality, and current relationship status influence perceiver evaluations of vocal cues remains to be explored.

Consideration should also be given to the type of vocal cues used to elicit perceptual evaluations. Most studies investigating vocal perception have speakers vocalise vowel sounds, recite words, or read standard passages, in attempt to control for factors such as semantic content and speech patterns (Hughes & Gallup, 2008). For the perceptual studies in the present research, spoken numeral counts were utilised as vocal stimuli. While this has the advantage of experimental control, there is also need to consider vocalizations that are of greater ecological validity. For example, research has found that the semantic content of vocal stimuli influences both voice production and perception. Bryant and Haselton (2009) found that female speakers produced vocal signals of higher F_0 when speaking at high fertility relative to low fertility, but only when producing vocal samples that had semantically meaningful content and not when producing simple vowel sounds. In terms of perception, male perceivers have been found to show greater preference for female voices raised in F_0 when listening to sentences containing semantic cues of social interest (e.g., “I really like you”), than when listening to sentences containing semantic cues of social disinterest (e.g., “I don’t really like you”) (Jones, Feinberg, DeBruine, Little, & Vukovic, 2008).

Further understanding of the role of vocal signals in social behaviour could also be improved by giving greater consideration to the dynamic, interactive nature of vocal production and perception. The vocal samples in the present research were elicited in a relatively neutral social context. However, vocal signals are frequently produced in the presence of socially meaningful others and often directed toward particular individuals. While many vocal characteristics are tightly constrained, others can be readily manipulated (within certain limits), and it is clear that speakers modify their voices dependent on who they are interacting with. For example, adults tend to raise the F_0 of their voices when talking to children (Trainor & Desjardins, 2002). In adult interaction there is evidence that perceptions of physical dominance influence the modulation of male F_0. In a competitive experimental situation, men who believed they were physically dominant to a rival tended to lower their F_0 when speaking to that rival, while men who believed they were less dominant tended to raise their F_0 (Puts, Gaulin, & Verdolini, 2006). Similarly, there is evidence that both males and females alter their F_0 when directing vocal signals toward opposite sex individuals that they find attractive (Fraccaro et al., 2011; Hughes, Farley, & Rhodes, 2010).

While voices are sometimes encountered in isolation, they are also frequently encountered in combination with other types of social information – most notably the visual information afforded by a speaker’s face. As such, there is a need to consider how vocal information interacts with other sources of information to affect social perception. Feinberg (2008) has argued that faces and voices may signal highly similar speaker traits, and previous research has found some face and voice judgments to be correlated (Collins & Missing, 2003; Feinberg et al., 2008; Lander, 2008). Study 5 in the present research investigated how voice and face information interact to influence social memory. Results suggested that perceivers have expectations that vocal and facial information specifying an individual’s sex and age should co-vary in a reliable manner. Establishing exactly how such expectations influence social behaviour remains an important task for future research.

Together, the above considerations highlight the importance of interaction effects for the present research. It seems likely that particular social contexts (considered in terms of the situation, the speaker, and the perceiver) provide opportunities for both vocal production and perception that other contexts do not. That is, both the information that is available, and the perception of that information, may be differential across social contexts. As such, the current findings may be further refined or extended by considering how individual differences in both speakers and perceivers combine in particular social contexts to influence the processes of voice perception.

One such context may be related to singing. Due to its apparent lack of utility, Darwin (1871) described music, including singing, as one of the most mysterious capacities possessed by humans, but suggested that it could be understood as a courtship display. Several prominent theories have followed Darwin's original position, maintaining that music and singing are sexually selected social displays that function in the service of courtship (e.g., Dutton, 2009; Miller, 2000). In this view, human singing may function, at least in part, as a conspicuous display of vocal quality[8]. That is, because voices appear to honestly signal a range of reproductively relevant traits, some individuals (particularly those with high quality versions of a trait) may benefit from advertising those traits. Moreover, because vocal characteristics are being conspicuously displayed, singing provides a particularly opportune context for motivated perceivers to attend to and evaluate vocal characteristics and may yield higher levels of perceptual accuracy than standard speech. In this sense, singing may represent a

[8] In addition to vocal quality, singing is likely to be informative of a range of features, including aerobic capacity, memory, rhythm, neuro-muscular functioning, motor-control, and creativity.

"social litmus test" (Baron & Misovich, 1993) for both the production and perception of information afforded by vocal signals.

The present research investigated the social perception of human voices. A series of studies explored the relationships between speaker vocal and physical characteristics and the extent to which perceivers attend to the information in vocal signals in a socially adaptive manner. Novel empirical evidence was provided demonstrating that social perceivers are attuned to the vocal specification of speaker physical characteristics. Perceivers were found to use natural variation in the acoustic properties of voices to make consensual and accurate judgments of speaker sex, age, and body size. Attunement to the acoustic specification of such characteristics is likely to proffer considerable advantage to social perceivers. Moreover, perceiver judgments of vocal attractiveness covaried with speaker characteristics in a functional manner. Both male and female perceivers rated the voices of male speakers with lower indices of body asymmetry as more attractive, suggesting that perceivers are attuned to vocal properties that specify a speaker's degree of developmental stability. For female perceivers, this attunement was shown to be functionally modulated by cyclically shifting fertility. Additionally, perceivers were found to rate female voices recorded during menstrual cycle phases of high fertility to be more attractive than voices recorded during phases of low fertility, suggesting that perceivers are attuned the vocal specification of female fertility status. Novel consideration was also given to how the vocal specification of speaker cues combines with the visual specification of such cues to influence social perception. Taken as a whole, the present research suggests that human voices are a highly salient and information rich signal, strategically utilised by both vocalisers and perceivers in the service of adaptive social behaviour.

References

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

Alexander, R. D., & Noonan, K. M. (1979). Concealment of ovulation, parental care, and human social evolution. In N. A. Chagnon, & Irons, W. (Ed.), *Evolutionary biology and human social behavior* (pp. 436-543). North Scituate, MA: Duxbury Press.

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

Amir, O., & Kishon-Rabin, L. (2004). Association Between Birth Control Pills and Voice Quality. *The Laryngoscope, 114*(6), 1021-1026

Amir, O., Kishon-Rabin, & Muchink, C. (2002). The effect of oral contraceptives on voice: Preliminary observations. *Journal of Voice, 16*, 267-273.

Andersson, M. (1994). *Sexual selction*. New Jersey: Princeton University Press.

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters, 3*(6), 682-684.

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

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews, 30*, 319-345.

Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral Brain Science, 32*, 249-311.

Bachorowski, J., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *The Journal of the Acoustical Society of America, 106*, 1054-1063.

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development, 69*, 1263-1275.

Baird, D. D., Weinberg, C. R., Zhou, H., Kamel, F., McConnaughey, D. R., Kesner, J. S., & Wilcox, A. J. (1999). Preimplantation urinary hormone profiles and the probability of conception in healthy women. *Fertility and sterility, 71*, 40-49.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. San Diego: Singular Thomson Learning.

Barnes, L. L., N. (2011). Acoustic measures of hormone affect on female voice during menstruation. *International Journal of Humanities and Social Science, 1*(3), 5-10.

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding Voice Perception. *British Journal of Psychology, 102*(4), 711-725.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129-135.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport, 14*(16), 2105-2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309-312.

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

Benshoof, L., & Thornhill, R. (1979). The evolution of monogamy and concealed ovulation in humans. *Journal of Social and Biological Systems, 2*(2), 95-106.

Berthoud, H., & Neuhuber, W. L. (2000). Functional and chemical anatomy of the afferent vagal system. *Autonomic Neuroscience, 85*, 1-17.

Bhasin, S., Storer, T. W., Berman, N., Callegari, C., Clevenger, B., Phillips, J. (1996). The Effects of Supraphysiologic Doses of Testosterone on Muscle Size and Strength in Normal Men. *New England Journal of Medicine, 335*(1), 1-7.

Bliss-Moreau, E., Feldman Barrett, L., & Owren, M. J. (2010). I like the sound of your voice: Affective learning about vocal signals. *Journal of Experimental Social Psychology, 46*, 557-563.

Braun, A., & Cerrato, L. (1999). Estimating speaker age across languages. *Journal of Voice, 1*, 49-52.

Brockmann, M., Drinnan, M. J., Storck, C., & Carding, P. N. (2011). Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *Journal of Voice, 25*, 44-53.

Brown, W., Morris, R., & Michel, J. (1989). Vocal jitter in young adult and aged female voices. *Journal of Voice, 3*, 113-119.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception, 22*, 131-152.

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 Sciences, 273*(1582), 83-89.

Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters, 5*(1), 12-15.

Burriss, R. P., Welling, L. L. M., & Puts, D. A. (2011). Men's attractiveness predicts their preference for female facial femininity when judging for short-term, but not long term, partners. *Personality and Individual Differences, 50*, 542-546.

Burt, A. (1992). 'Concealed Ovulation' and Sexual Signals in Primates. *Folia Primatologica, 58*(1), 1-6.

Burt, D. M., & Perrett, D. I. (1995). Perception of Age in Adult Caucasian Male Faces: Computer Graphic Manipulation of Shape and Colour Information. *Proceedings: Biological Sciences, 259*(1355), 137-143.

Caffi, C., & Janney, R. W. (1994). Toward a pragmatics of emotive communication. *Journal of Pragmatics, 22*(3–4), 325-373.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition, 56*(1), 30-35.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial Structure Is a Reliable Cue of Aggressive Behavior. *Psychological Science, 20*(10), 1194-1198.

Caruso, S., Roccasalva, L., Sapienza, G., Zappalá, M., Nuciforo, G., & Biondi, S. (2000). Laryngeal cytological aspects in women with surgically induced menopause who were treated with transdermal estrogen replacement therapy. *Fertility and sterility, 74*(6), 1073-1079.

Cerrato, L., Falcone, M., & Paoloni, A. (2000). Subjective age estimation of telephone voices. *Speech Communication, 31*, 107-112.

Chae, S. W., Choi, G., Kang, H. J., Choi, J. O., & Jin, S. M. (2001). Clinical analysis of voice change as a parameter of premenstrual syndrome. *Journal of Voice, 15*(2), 278-283.

Chan, R. W., & Titze, I. R. (2006). Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics *Journal of the Acoustical Society of America, 119*, 2351-2362.

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

Clark, A., S., & Henderson, L. P. (2003). Behavioral and physiological responses to anabolic-androgenic steroids. *Neuroscience and Biobehavioral Reviews, 27*, 413-436.

Coats, S., & Smith, E. R. (2006). Distinctiveness and memory: a comparison of the social and cognitive literatures. In R. R. Hunt, & Worthen, J. B. (Ed.), *Distinctiveness and memory* (pp. 313-336). New York: Oxford University Press.

Cohen, J. R., Crystal, T. H., House, A. S., & Neuberg, E. P. (1980). Weighty voices and shaky evidence: A critique. *Journal of the Acoustical Society of America, 68*(6), 1884-1886.

Cohen, S. E. (1974). Developmental differences in infants' attentional responses to face-voice incongruity of mother and stanger. *Child Development, 45*, 1155-1158.

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

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

Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition, 44*(3), 197-126.

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences, 27*(4), 801-804.

Daly, N. W., P. (2001). Pitching it differently in New Zealand English: speaker sex and intonation patterns. *Journal of Sociolinguistics, 5*, 85-96.

Daniel, H. J., & McCabe, R. B. (1992). Gender differences in the perception of vocal sexiness. In J. M. G. van der Dennen (Ed.), *The nature of the sexes: The Sociobiology of Sex Differences and the "Battle of the Sexes."* (pp. 55-62). Groningen: Origin Press.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual Differences in the Ability to Recognise Facial Identity Are Associated with Social Anxiety. *PLoS ONE, 6*(12), e28800.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 1980*, 1174-1176.

Deterding, D. (1997). The Formants of Monophthong Vowels in Standard Southern British English Pronunciation. *Journal of the International Phonetic Association, 27*, 47-55.

Dixson, A. F. (1998). *Primate sexuality: Comparative studies of the prosimians, monkeys, apes and human beings*. Oxford: Oxford University Press.

Dutton, D. (2009). *The Art Instinct: Beauty, Pleasure, and Human Evolution*. New York: Bloomsbury.

Ekman, P., Friesen, W. K., & Scherer, K. (1976). Body movements and voice pitch in deceptive interaction. *Semiotica, 16*, 23-27.

Ellis, D. S. (1967). Speech and social status in America. *Social Forces, 45*, 431-437.

Endres, W. B., W., & Flosser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America, 49*, 1842-1847.

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*, 160-163.

Fairbanks, G. (1960). *Voice and articulation drillbook*. New York: Harper & Bros.

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

Feinberg, D. R. (2008). Are human faces and voices ornaments signalling common underlying cue to mate value? *Evolutionary Anthroplogy, 17*, 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 relative role of feminity and averageness in aesthetic judgments of women's voices. *Perception, 37*, 615-623. doi: 10.1068/p5514

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E. (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., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561-568.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, D. I., & Perrett, D. I. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior, 49*, 215-222.

Ferdenzi, C., Lemaître, J.-F., Leongómez, J. D., & Roberts, S. C. (2011). Digit ratio (2D:4D) predicts facial, but not voice or body odour, attractiveness in men. *Proceedings of the Royal Society B: Biological Sciences*.

Ferrario, V. F., Sforza, C., Pizzini, G., Vogel, G., & Miani, A. (1993). Sexual dimorphism in the human face assessed by euclidean distance matrix analysis. *Journal of Anatomy, 183*, 593-600.

Fink, B., Grammer, K., Mitteroecker, P., Gunz, P., Schaefer, K., Bookstein, F. L., & Manning, J. T. (2005). Second to fourth digit ratio and face shape. *Proceedings of the Royal Society B: Biological Sciences, 272*, 1995-2001.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Fitch, W. T. (1994). *Vocal tract length perception and the evolution of language.* Unpublished doctoral dissertation, Brown University.

Fitch, W. T. (1997a). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America, 102*, 1213-1222.

Fitch, W. T. (1997b). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America, 102*(2), 1213-1222.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences, 4*, 258-267.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106*, 1511-1522.

Fitch, W. T., & Hauser, M. D. (2002). Unpacking "honesty": vetebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, Fay, R. R., & Popper, A. N. (Ed.), *Acoustic Communication* (pp. 65-137). New York: Springer.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society B: Biological Sciences, 268*, 1669-1675.

Fitch, W. T., Neubauert, J., & Herzel, H. (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour, 63*, 407-418.

Folstad, I., & Karter, A. J. (1992). Parasites, Bright Males, and the Immunocompetence Handicap. *American Naturalist, 139*(3), 603-622.

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

Gallup, A. C., White, D. D., & Gallup, G. G. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *28*(6), 423-429.

Gallup, G. G., & Cameron, P. A. (1992). Modality specific metaphors: is our mental machinery "colored" by a visual bias? *Metaphor and Symbolic Activity, 7*, 93-98.

Gamboa, J., Jimenez-Jimenez, F. J., Mate, M. A., & Cobeta, I. (2001). Voice disorders caused by neurological diseases. *Review Neurology, 33*, 153-168.

Gangestad, S. W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proceedings Royal Society of London B, 265*, 927-933.

Gangestad, S. W., Simpson, J. A., Cousins, A. J., Garver-Apgar, C. E., & Christensen, P. N. (2004). Research Report Women's Preferences for Male Behavioral Displays Change Across the Menstrual Cycle. *Psychological Science, 15*(3), 203-207.

Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: Examination of a functional hypothesis. *Journal of Personality and Social Psychology, 62*, 688-698.

Gangestad, S. W., Thornhill, R., & Garver, C. E. (2002). Changes in women's sexual interests and their partner's mate–retention tactics across the menstrual cycle: evidence for shifting conflicts of interest. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 269*(1494), 975-982.

Garn, S. M., Burdi, A. R., Babler, W. J., & Stinson, S. (1975). Early prenatal attainment of adult metacarpal–phalangeal rankings and proportions. *American Journal of Physical Anthropology, 43*, 327-332.

Ghazanfar, A. A., & Hauser, M. D. (1999). The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends in Cognitive Science, 3*, 377-384.

Giles, N. L., & Powsland, N. F. (1975). *Speech style and social evaluation*. New York: Academic 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), 1166-1183.

Gonzalez, J. (2003). Estimation of speakers' weight and height from speech: A re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills, 96,* 297-304.

Gonzalez, J. (2006). Research in acoustics of human speech sounds: correlates and perception of speaker body size. *Recent Research and Development in Applied Physics, 9,* 1-15.

Gonzalez, J. (2007). Correlations between speakers' body size and acoustic parameters of the voice. *Perceptual and Motor Skills, 105*, 215-220.

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

Gradol, D., & Swan, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech, 26,* 351-366.

Grafen, A. (1990). Biological Signals as Handicaps. *Journal of Theoretical Biology, 144*(4), 517-546.

Gray, P. B., Kahlenberg, S. M., Barrett, E. S., Lipson, S. F., & Ellison, P .T. (2002). Marriage and fatherhood are associated with lower testosterone in males. *Evolution and Human Behavior, 23*,191-201.

Green, S. M. (1981). Sex differences and age gradations in vocalizations of Japanese and lion-tailed monkey. *American Zoologist, 21*, 165-184.

Grossman, C. J. (1985). Interactions between gonadal steroids and the immune system. *Science*, *227*, 257-261.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, *65*(6), 852-858.

Gueguen, N. (2009). The receptivity of women to courtship solicitation across the menstrual cycle: A field experiment. *Biological Psychology*, *80*, 321-324.

Harries, M., Hawkins, S., Hacking, J., & Hughes, I. A. (1998). Changes in the male voice at puberty: vocal fold length and its relationship to the fundamental frequency of the voice. *Journal of Laryngology and Otology, 112*, 451-454.

Harris, T. R., Fitch, W. T., Goldstein, L. M., & Fashing, P. J. (2006). Black and white colobus roars as a source of both honest ans exaggerated information about body mass. *Ethology*, *112*, 911-920.

Haselton, M. G., & Miller, G. F. (2006). Women's fertility across the cycle increases the short-term attractiveness of creative intelligence. *Human Nature, 17*, 50-73.

Haselton, M. G., Mortezaie, M., Pillsworth, E. G., Bleske-Rechek, A., & Frederick, D. A. (2007). Ovulatory shifts in human female ornamentation: Near ovulation, women dress to impress. *Hormones and Behavior, 51*, 41-45.

Hastie, R., & Kumar, P. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25-38.

Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Higgins, M. B., & Saxman, J. H. (1989). Variations in vocal frequency perturbation across the menstrual cycle. *Journal of voice : official journal of the Voice Foundation, 3*(3), 233-243.

Hirano, M., Kurita, S., Yukizane, K., & Hini, S. (1989). Asymmetry of the laryngeal framework: a morphological study of cadaver larynges. *Annals of Oyology, Rhinology and laryngology, 98*, 135-140.

Hodges-Simeon, C., Gaulin, S., & Puts, D. (2010). Voice Correlates of Mating Success in Men: Examining "Contests" Versus "Mate Choice" Modes of Sexual Selection. *Archives of Sexual Behavior, 40*(3), 551-557.

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronological age in males. *Journal of Speech and Hearing Research, 15*, 155-159.

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

Horii, Y. (1980). Vocal Shimmer in Sustained Phonation. *J Speech Hear Res, 23*(1), 202-209.

Hrdy, S. B. (1981). *Thw woman that never evolved*. Cambridge, MA: Harvard University Press.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *Journal of the Acoustical Society of America, 106*, 1532-1542.

Hughes, S., Harrison, M. A., & Gallup, G, G. (2009). Sex-specific body configurations can be estimated from voice samples. *Journal of Social, Evolutionary, and Cultural Psychology, 3*, 343-355.

Hughes, S., Pastizzo, M., & Gallup, G. (2008). The Sound of Symmetry Revisited: Subjective and Objective Analyses of Voice. *Journal of Nonverbal Behavior, 32*(2), 93-108.

Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior, 25*(5), 295-304. doi: 10.1016/j.evolhumbehav.2004.06.001

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*, 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*(3), 173-180.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H. (1997). Vocal identification of speaker and emotion activates differerent brain regions. *NeuroReport, 8*(12), 2809-2812.

Ingemann, F. (1968). Identification the speaker's sex form voiceless fricative. *Journal of the Acoustical Society of America, 88*, 642-654.

Jamison, C. S., Meier, R. J., & Campbell, B. C. (1993). Dermatoglyphic asymmetry and testosterone levels in normal males. *American Journal of Physical Anthropology, 90*, 185-198.

Jenkins, J. S. (1998). The voice of the Castrato. *Lancet, 351*, 1877-1880.

Jochle, W. (1973). Coitus-induced ovulation. *Contraception, 7*, 523-564.

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*, 192-194.

Johnson, K. A. (2005). Speaker normalization in speech perception. In D. B. Pisoni, & Remez, R. (Ed.), *The Handbook of Speech Perception* (pp. 363-389). Oxford: Blackwell.

Johnson, W., Gangestad, S. W., Segal, S., & Bouchard, T. J. (2008). Heritability of fluctuating asymmetry in a human twin sample: The effect of trait aggregation. *American Journal of Human Biology, 20*, 651-658.

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22*(4), 251-267.

Kahane, J. (1987). Connective tissue changes in the larynx and their effects on voice. *Journal of Voice, 1*, 27-30.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109-2128.

Kasperk, C., Helmboldt, A., Borcsok, I., Heuthe, S., Cloos, O., Niethard, F. (1997). Skeletal site-dependent expression of the androgen receptor in human osteoblastic cell populations. *Calcified Tissue International, 61*(6), 464-473.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98*(2), 155-163.

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, Ke., & Wang, Z. (2003). Effetcs of experience on fetal voice recognition. *Psychological Science, 14*, 220-224.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the Voice Reveals: Within- and Between-Category Stereotyping on the Basis of Voice. *Personality and Social Psychology Bulletin, 32*(6), 806-819.

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology, 38*(6), 618-625.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98-104.

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

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115-121.

Lass, N., Tecca, J., Mancuso, R., & Black, W. (1979). The effect of phonetic complexity on speaker race and sex identification. *Journal of Phonetics, 7,* 105-118.

Lass, N. J., & Brown, W. S. (1978). Correlational study of speakers' heights, weights, body surface areas and speaking fundamnetal frequencies. *Journal of the Acoustical Society of America, 63,* 1218-1220.

Lass, N. J., & Davis, M. (1976). An investigation of speaker height and weight identification. *Journal of the Acoustical Society of America, 60*(3), 700-703.

Lass, N. J., Beverly, A. S., Nicosia, D. K., & Simpson, L. A. (1977). An investigation of speaker height and weight identification: II. Direct estimations. *The Journal of the Acoustical Society of America, 61*(S1), S70-S70.

Lass, N. J., Brong, G. W., Ciccolella, S. A., Walters, S. C., & Maxwell, E. L. (1980). An investigation of speaker height and weight discriminations by means of paired comparison judgments. *Journal of Phonetics, 8*, 205-212.

Lass, N. J., & Colt, E. G. (1980). A Comparative-Study of the Effect of Visual and Auditory Cues on Speaker Height and Weight Identification. *Journal of Phonetics, 8*(3), 277-285.

Lass, N. J., Dicola, G. A., Beverly, A. S., Barbera, C., Henry, K. G., & Badali, M. K. (1979). The effect of phonetic complexity on speaker height and weight identification. *Language and Speech, 22*(4), 297-309.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America, 59*(3), 675-678.

Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neuroscience, 11,* 1-13.

Lenton, E. A., Landgren, B. M., & Sexton, L. (1984). Normal variation in the length of the luteal phase of the menstrual cycle: Identification of the short luteal phase. . *British Journal of Obstetrics and Gynaecology, 91,* 685-689.

Leong, K. M., Ortolani, A., Graham, L. H., & Savage, A. (2003). The use of low-frequency vocalizations in African elephant(Loxodonta africana) reproductive strategies. *Hormones and Behavior, 43*(4), 433-443.

Lindholm, P., Vilkman, E., Raudaskoski, T., SuvantoLuukkonen, E., Kauppila, A. (1997). The effect of postmenopause and postmenopausal HRT on measured voice values and vocal symptoms. *Maturitas, 28,* 47-53.

Linville, S. E. (1998). Acoustic Correlates of Perceived versus Actual Sexual Orientation in Men's Speech. *Folia Phoniatrica et Logopaedica, 50*(1), 35-48. doi: 10.1159/000021447

Linville, S. E., & Fisher, H. B. (1985). Acoustic characteristics of women's voices waith advancing age. *Journal of Gerontology, 40,* 324-330.

Livshits, G., & Kobyliansky, E. (1991). Fluctuating asymmetry as a possible measure of developmental homeostasis in humans - a review. *Human Biology, 63*(441-466).

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of baserate fallacy. *Journal of Experimental Social Psychology, 18*, 23-42.

Ludlow, C. L., Sedory Holzer, S. E., & Fujita, M. (1997). Correlations among intrinsic laryngeal muscles during speech gestures. In S. Kiritani, Hirose, H., & Fujisaki, H. (Ed.), *Speech production and language* (pp. 9-32). New York: Mouton de Gruyter.

Manning, J. T. (2002). *Digit ratio: a pointer to fertility, behavior, and health*. New Brunswick: Rutgers University Press.

Manning, J. T., Scutt, D., Whitehouse, G. H., Leinster, S. J., & Walton, J. M. (1996). Asymmetry and the menstrual cycle in women. *Ethology and Sociobiology, 17*(2), 129-143.

Masataka, N. (1994). Lack of correlation between body size and frequency of vocalizations in young female. *Folia Primatologica, 63,* 115-118.

McComb, K. E. (1991). Female choice for high roaring rates in red deer, Cervus elaphus. *Animal Behaviour, 41*, 79-88.

McGurk, H., & MacDonald, J. D. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Meurer, E. M., Garcez, V., von Eye Corleta, H., & Capp, E. (2009). Menstrual cycle influences on voice and speech in adolescent females. *Journal of Voice, 23*(1), 109-113.

Miles, L., & Johnston, L. (2007). Detecting Happiness: Perceiver Sensitivity to Enjoyment and Non-Enjoyment Smiles. *Journal of Nonverbal Behavior, 31*(4), 259-275.

Miller, G. (2000a). Evolution of human music through sexual selection. In R. Dunbar, Knight, C., & Power, C. (Ed.), *The origins of music* (pp. 329-360). Boston: MIT Press.

Miller, G. (2000b). *The mating mind: how sexual choice shaped the evolution of human nature*. New York: Doubleday.

Miller, G., Tybur, J. M., & Jordan, B. D. (2007). Ovulatory cycle effects on tip earnings by lap dancers: economic evidence for human estrus? *28*(6), 375-381.

Mitani, J. C., & Gros-Louis, J. (1995). Species and sex differences in the screams of chimpanzees and bonobos. *International Journal of Primatology, 16,* 393-411.

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2*, 10-12.

Moller, A. P., & Swaddle, J. P. (1997). *Asymmetry, developmental stability and evolution*. Oxford: Oxford University Press.

Moller, A. P., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: A meta-analysis. *American Naturalist, 151*(2), 174-192.

Mondloch, C. J. (2012). Sad or fearful? The influence of body posture on adults' and children's perception of facial displays of emotion. *Journal of Experimental Child Psychology, 111*, 180-196.

Moore, J. F., & Thompson, C. L. (1965). Comments on physiology of hoarseness. *Archives of Otolaryngology, 81*, 97-102.

Morton, E. S. (1977). On the occurrence and significance of motivational structural rules in some bird and animal sounds. *American Naturalist, 111*, 855-869.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia, 39*(10), 1047-1054.

Need, A. G., Durbridge, T. C., & Nordin, B. E. (1993). Anabolic steroids in postmenopausal osteoporosis. *Wiener Medizinische Wochenschrift, 143*, 392-395.

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

Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological Impairments in the Recognition of Faces, Voices, and Personal Names. *Brain and Cognition, 44*(3), 342-366.

Newman, S. R., Butler, J., Hammond, E. H., & Gray, S. D. (2000). Preliminary report on hormone receptors in the human vocal fold. *Journal of Voice, 14*, 72-81.

Nieman, G. S., & Applegate, J. A. (1990). Accuracy of listener judgments of perceived age relative to chronological age in adults. *Folia Phoniatrica, 42*, 327-330.

O'Mahony, C., & Newell, F. N. (2011). Integration of faces and voices, but not faces and names, in person recognition. *British Journal of Psychology, 103*, 73-82.

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Human Development, 18*, 27-36.

Oguchi, T., & Kikuchi, H. (1997). Voice and interpersonall attraction. *Japanese Psychological Research, 39*, 56-61.

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences, 107*(30),

Oskenberg, L., Coleman, L., & Cannell, C. F. (1986). Interviewers' voice and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*, 97-111.

Palmer, A. R. (1994). Fluctuating asymmetry analyses: a primer. In T. Markow (Ed.), *Developmental stability: its origins and evolutionary implications* (pp. 355-364). Dordrecht: Kluwer Academic.

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

Pawłowski, B. (1999). Loss of Oestrus and Concealed Ovulation in Human Evolution: The Case against the Sexual-Selection Hypothesis. *Current Anthropology*, *40*(3), 257-276.

Penton-Voak, I. S., & Perrett, D. I. (2000). Female preference for male faces changes cyclically: Further evidence. *Evolution and Human Behavior*, *21*(1), 39-48.

Penton-Voak, I. S., Perrett, D. I., Castles, D. L., Kobayashi, T., Burt, D. M., Murray, L. K. (1999). Menstrual cycle alters face preference. *Nature, 399*(6738), 741.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, *117*(6),

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*(3), 367-374.

Pfefferle, D., Brauch, K., Heistermann, M., Hodges, J. K., & Fischer, J. (2008). Female Barbary macaque (Macaca sylvanus) copulation calls do not reveal the fertile phase but influence mating outcome. *Proceedings of the Royal Society B: Biological Sciences, 275*(1634), 571-578.

Pillsworth, E. G., Haselton, M. G., & Buss, D. M. (2004). Ovulatory shifts in female sexual desire. *Journal of Sex Research, 41*, 55-65.

Pipitone, R. N., & Gallup, G. G. (2008). Women's voice attractiveness varies across the menstrual cycle. *29*(4), 268-274.

Planalp, S. (1998). Communicating emotion in everday life: Cues, channels, and processes. In P. A. Anderson, & Guerrero, L. K. (Ed.), *Handbook of communication and emotion* (pp. 29-48). New York: Academic Press.

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research, 9*, 273-277.

Puts, D., Hodges, C., Cardenas, R., & Gaulin, S. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*, 340-344.

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

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2011). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*.

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.

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.

Raj, A., Gupta, B., Chowdhury, A., & Chadha, S. (2010). A Study of Voice Changes in Various Phases of Menstrual Cycle and in Postmenopausal Women. *Journal of voice : official journal of the Voice Foundation, 24*(3), 363-368.

Ramig, L., & Ringel, R. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech and Hearing Research, 26*, 22-30.

Rastatter, M., & Jacques, R. (1990). Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatrica, 42*, 312-319.

Reby, D., & McComb, K. (2003). Vocal Communication and Reproduction in Deer. *Advances in the Study of Behavior, 33*, 231-264.

Rendall, D., Vokey, J. R., Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1208-1219.

Rendall, D. K., S. Ney, C. & Lloyd, P. (2005). Pitch (Fo) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117*, 944-955.

Reynolds, D. J., & Gifford, R. (2001). The sounds & sights of intelligence: A lens model channel analysis. *Personality and Social Psychology Bulletin, 27*, 187-200.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*, 199-226.

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.

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behaviour, 22*, 31-46.

Riede, T., & Fitch. W. T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog. *Journal of Experimental Biology, 202*, 2859-2869.

Rikowski, A., & Grammer, K. (1999). Human Body Odour, Symmetry and Attractiveness. *Proceedings: Biological Sciences, 266*(1422), 869-874.

Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C. (2004). Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271*(Suppl 5), S270-S272.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirmng events: memory processes and the maintenance of of social stereotypes. *Journal of Experimental Social Psychology, 15*, 343-355.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Edpidemiology, 1,* 43-46.

Roughgarden, J., Oishi, M., & Akçay, E. (2006). Reproductive Social Behavior: Cooperative Games to Replace Sexual Selection. *Science, 311*(5763), 965-969.

Ryan, M., & Kenny, D. T. (2009). Perceived effects of the menstrual cycle on young female singers in the Western classical tradition. *Journal of Voice, 23*, 99-108.

Ryan, M. J., & Keddy-Hector, A. (1992). Directional Patterns of Female Mate Choice and the Role of Sensory Biases. *The American Naturalist, 139*, S4-S35.

Ryan, W., & Burk, K. (1974). Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders, 7*, 181-192.

Saez, S. J., & Martin, P. M. (1976). Androgen receptors in human pharyngo-laryngeal mucosa and pharyngo-laryngeal epithelium. *Journal of Steroidal Biology, 7*, 919-921.

Sapir, E. (1927). Speech as a personality trait. *American Journal of Sociology, 32*(6), 892-905.

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant Centralization Ratio: A Proposal for a New Acoustic Measure of Dysarthric Speech. *Journal of Speech, Language and Hearing Research, 53*, 114-125.

Sasaki, C. T., Levine, P. A., Laitman, J. T., & Crelin, E. S., Jr. (1977). Postnatal Descent of the Epiglottis in Man: A Preliminary Report. *Arch Otolaryngol, 103*(3), 169-171.

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*, 1179-1185.

Scherer, K. R. (1972). Judging personality from voice: a cross cultural approach to an old issue in interpsersonal perception. *Journal of Personality, 40*, 191-210.

Scherer, K. R., & Zei, R. (1988). Vocal indicators of affective disorders. *Psychotherapy and Psychosomatics, 49*, 179-186.

Schwartz, M. F., & Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America, 44*, 1736-1737.

Scukanec, G. P., Petrosino, L., & Squibb, K. (1991). Formant frequency characteristics of children, young adult, and aged female speakers. *Perceptual and Motor Skills, 73*, 203-208.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*(1699), 3509-3518.

Semple, S., & McComb, K. (2000). Perception of female reproductive state from vocal cues in a mammal species. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 267*(1444), 707-712.

Semple, S., McComb, K., Alberts, S., & Altmann, J. (2002). Information content of female copulation calls in yellow baboons. *American Journal of Primatology, 56*(1), 43-56.

Silverman, E.-M., & Zimmer, C. H. (1978). Effect of the Menstrual Cycle on Voice Quality. *Arch Otolaryngol, 104*(1), 7-10.

Singh, D. (1993). Adaptive Significance of Female Physical Attractiveness - Role of Waist-to-Hip Ratio. *Journal of Personality and Social Psychology, 65*(2), 293-307.

Singh, D. (1995). Female health, attractiveness, and desirability for relationships: Role of breast asymmetry and waist-to-hip ratio. *Ethology and Sociobiology, 16*, 465-481.

Singh, D., & Bronstad, P. M. (2001). Female Body Odour Is a Potential Cue to Ovulation. *Proceedings: Biological Sciences, 268*(1469), 797-801.

Smith, E. L., Grabowecky, M. & Suzuki, S. (2007). Auditory-visual crosmodal integration in perception of face gender. *Current Biology, 17*, 1680-1685.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social-development literatures. *Psychological Bulletin, 111*, 42-61.

Stirrat, M., & Perrett, D. I. (2010). Valid Facial Cues to Cooperation and Trust. *Psychological Science, 21*(3), 349-354.

Strassmann, B. I. (1981). Sexual selection, paternal care, and concealed ovulation in humans. *Ethology and Sociobiology, 2*(1), 31-40.

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology, 35*, 345-350.

Tallerman, M., & Gibson, K. (2011). *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America, 96*, 2101-2107.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology, 280*(3), 221-236.

Thornhill, R., & Gangestad, S. W. (2008). *The evolutionary biology of human female sexuality*. New York: Oxford University Press.

Thornhill, R., & Gangestead, S. W. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior, 20*, 175-201.

Tigges, M., Mergell, P., Herzel, H., Wittenberg, T., & Eysholdt, U. (1997). Observation and modelling of glottal biphonation. *Acustica, 83*, 707-714.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85*, 1699-1707.

Titze, I. R. (1994). *Principles of voice production*. New Jersey: Prentice Hall.

Tovee, M. J., Reinhart, S., Emery, J. L., & Cornelissen, P. L. (1998). Optimum body-mass index and maximum sexual atttactiveness. *The Lancet, 352*, 548.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review, 9*(2), 335-340.

Traunmuller, H. (1997). Perception of speaker sex, age, and vocal effort. *Phonum, 4*, 183-186.

Turke, P. W. (1984). Effects of ovulatory concealment and synchrony on protohominid mating systems and parental roles. *Ethology and Sociobiology*, *5*(1), 33-44.

van den Berghe, P. L., & Frost, P. (1986). Skin color preference, sexual dimorphism and sexual selection: A case of gene culture co-evolution? *Ethnic and Racial Studies, 9*(1), 87-113.

Van Dommelen, W. A. (1993). Speaker height and weight identification: reevaluation of some old data. *Journal of Phonetics, 21,* 337-341.

Van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech, 38*, 267-287.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*(5), 829-834.

Van Valen, L. (1962). A study of fluctuating asymmetry. *Evolution, 10*, 139-146.

von Kriegstein, K., Egera, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17,* 48-55.

Vukovic, J., Feinberg, D. R., DeBruine, L. M., Smith, F. G., & Jones, B. C. (2010). Women's voice pitch is negatively correlated with health risk factors. *Journal of Evolutionary Psychology, 8,* 217-225.

Walker, J. (1994). Craniofacial changes. In J. Pinkham, R., Casamassimo, P. S., McTigue, D. J., Fields, H. W., & Nowak, A. (Ed.), *Pediatric Dentistry: Infancy Through Adolescence*. Philadelphia: Saunders.

Wells, T. J., Dunn, A. K., Sergeant, M. J. T., & Davies, M. N. O. (2009). Multiple singals in human mate selection: A review and framework for integrating facial and vocal signals. *Journal of Evolutionary Psychology, 7,* 111-139.

Whiteside, S. P. (1998). Indentification of a speaker's sex: a study of vowels. *Perceptual and Motor Skills, 86*, 579-584.

Wilcox, A. J., Dunson, D. B., Weinberg, C. R., Trussell, J., & Baird, D. D. (2001). Likelihood of conception with a single act of intercourse: providing benchmark rates for assessment of post-coital contraceptives. *Contraception, 63*(4), 211-215.

Wilcox, K. A., & Horii, Y. (1980). Age changes in vocal jitter. *Journal of Gerontology, 35*, 184-198.

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

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech: a pilot studt. *Journal of Speech, Language and Hearing Research, 46*, 689-701.

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. London: Oxford University Press.

Zebrowitz, L. A. (2006). Finally, Faces Find Favor. *Social Cognition, 24*(5), 657-701.
Zebrowitz, L. A., Hall, J.A., Murphy, N.A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin, 28*, 238-249.

Zheng, Z., & Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1108312108

Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior, 17*, 119-135.

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13*(2), 67-82.

Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The Vocal Attractiveness Stereotype - Replication and Elaboration. *Journal of Nonverbal Behavior, 14*(2), 97-112.

Appendices

Appendix A: Information sheet for female participants in Study 1

Information Sheet

Perception of Voices

You are invited to participate in the research project *Perception of Voices.*

The aim of this project is to examine the judgements people make about a person based on hearing that person's voice.

Your involvement in this project will be to provide a sample of your voice for others to listen to, and have measurements of your physical characteristics taken in order to assess how accurate other people are in estimating these characteristics based solely on your voice.

You will be asked to speak into a microphone and have your voice recorded as you count from 1 to 10 in English, say the letters A, E, I, O, U, and read a short passage. Measurements of your fingers, wrists, elbows, shoulders, hips and waist will be taken, and your height, weight and age will be recorded. The characteristics of your voice will be analysed using computer software and your voice recordings will be played to others.

Because hormone levels may affect the quality of your voice, you will be asked to complete two recording sessions at specific times across the menstrual cycle. To help achieve this timing, you will be asked a few questions about your menstrual cycle. Each recording session will take around 20 minutes, and as remuneration for your time and participation you will receive a $10 voucher.

Should you decide to participate you have the right to withdraw from the project at any time, including withdrawal of any information provided. The results of this project may be published but you can be assured of complete confidentiality.

The project is being carried out by Brad Miles as part of the requirements for PhD study, and is being conducted under the supervision of Professor Lucy Johnston. They can be contacted at the addresses below and will be happy to discuss any concerns that you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Brad Miles
Department of Psychology
University of Canterbury
3642987 x3409
brad.miles@pg.canterbury.ac.nz

Dr Lucy Johnston
Department of Psychology
University of Canterbury
3642967
lucy.johnston@canterbury.ac.nz

Appendix B: Information sheet for male participants in Study 1

Information Sheet

Perception of Voices

You are invited to participate in the research project *Perception of Voices.*

The aim of this project is to examine the judgements people make about a person based on hearing that person's voice.

Your involvement in this project will be to provide a sample of your voice for others to listen to, and have measurements of your physical characteristics taken in order to assess how accurate other people are in estimating these characteristics based solely on your voice.

You will be asked to speak into a microphone and have your voice recorded as you count from 1 to 10 in English, say the letters A, E, I, O, U, and read a short passage. Measurements of your fingers, wrists, elbows, shoulders, hips and waist will be taken, and your height, weight and age will be recorded. The characteristics of your voice will be analysed using computer software and your voice recordings will be played to others.

Because you voice may change over time, you will be asked to complete two recording sessions held one week apart. Each recording session will take around 20 minutes, and as remuneration for your time and participation you will receive a $10 voucher.

Should you decide to participate you have the right to withdraw from the project at any time, including withdrawal of any information provided. The results of this project may be published but you may be assured of complete confidentiality.

The project is being carried out by Brad Miles as part of the requirements for PhD study and is being conducted under the supervision of Professor Lucy Johnston. They can be contacted at the addresses below and will be happy to discuss any concerns that you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Brad Miles
Department of Psychology
University of Canterbury
3642987 x3409
brad.miles@pg.canterbury.ac.nz

Professor Lucy Johnston
Department of Psychology
University of Canterbury
3642967
lucy.johnston@canterbury.ac.nz

Appendix C: Participant consent form

CONSENT FORM

Perception of Voices

I have read and understood the description of the above named project. On this basis I agree to participate in the project, and I consent to the publication of the results of the project with the understanding that anonymity will be preserved.

I understand also that I may at any time withdraw from the project, including withdrawal of any information that I have provided.

Name (please print):

Signature:

Date:

Appendix D: Acoustic analyses results for the spoken sentence vocal sample

Table D1. Comparison of male speakers' acoustic measures for each measurement session.

	Session 1		Session 2				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	111	9.64	110	8.2	1.00	.33	0.18
F_0-SD (Hz)	14.68	11.80	12.13	6.05	1.15	.29	0.21
F_0-min (Hz)	94.57	9.15	93.93	8.74	0.58	.57	0.15
F_0-max (Hz)	179.03	56.49	162.87	46.12	1.45	.16	0.26
F_1 (Hz)	605	55.54	572	48.92	3.32	.002	0.61
F_2 (Hz)	1639	79.82	1675	85.69	2.19	.04	0.40
F_3 (Hz)	2708	88.40	2738	131.71	1.56	.13	0.28
F_4 (Hz)	3747	213.46	3794	177.37	1.14	.27	0.21
Jit-local (%)	2.67	0.84	2.37	0.52	1.87	.07	0.34
Jit-rap (%)	1.20	0.60	1.00	0.33	1.69	.10	0.31
Jit-ppq5 (%)	1.29	0.57	1.18	0.42	1.08	.29	0.20
Shim-local (%)	12.18	2.43	12.33	2.68	0.34	.74	0.06
Shim-apq3 (%)	4.34	1.12	4.16	1.33	0.72	.10	0.13
Shim-apq5 (%)	7.01	2.58	6.64	1.72	1.24	.22	0.23
Shim-apq11 (%)	12.85	5.42	14.11	3.88	1.42	.17	0.26
HNR	10.97	2.34	10.90	2.54	0.13	.90	0.02

Table D2. Comparison of female speakers' acoustic measures for each measurement session.

	Low fertility		High fertility				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	197.77	16.54	195.83	13.58	1.12	.27	0.20
F_0-SD (Hz)	28.17	11.94	31.47	11.97	1.08	.29	0.20
F_0-min (Hz)	133.87	28.54	123.83	26.64	1.35	.19	0.25
F_0-max (Hz)	356.37	90.43	356.37	90.43	1.20	.24	0.22
F_1 (Hz)	615	105.08	601	29.36	0.79	.44	0.14
F_2 (Hz)	1802	91.38	1799	56.74	0.29	.78	0.05
F_3 (Hz)	2767	215.53	2825	106.24	1.30	.21	0.24
F_4 (Hz)	3929	85.53	3959	142.18	0.99	.33	0.18
Jit-local (%)	2.25	0.48	2.15	0.49	0.95	.35	0.17
Jit-rap (%)	1.09	0.30	1.04	0.27	0.86	.40	0.16
Jit-ppq5 (%)	1.01	0.23	1.01	0.23	0.48	.63	0.09
Shim-local (%)	9.08	1.76	8.92	1.40	0.48	.64	0.09
Shim-apq3 (%)	3.34	1.30	3.04	0.69	1.21	.23	0.22
Shim-apq5 (%)	4.26	0.85	4.21	0.77	0.25	.80	0.05
Shim-apq11 (%)	8.64	1.32	8.71	1.44	0.25	.80	0.05
HNR	14.72	1.63	14.42	1.70	1.17	.25	0.21

Table D3. Comparison of anthropometric and acoustic measures between male and female participants.

	Males		Females				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	110.4	8.08	196.8	14.37	28.70	< .001	7.7
F_0-SD (Hz)	13.41	7.16	29.82	8.55	8.06	< .001	2.09
F_0-min (Hz)	94.25	8.42	128.85	18.65	9.26	< .001	2.56
F_0-max (Hz)	170.95	41.50	343.2	65.56	12.16	< .001	3.22
Jit-local (%)	2.52	0.55	2.20	0.39	2.66	.01	0.69
Jit-rap (%)	1.10	0.37	1.06	0.23	0.47	.64	0.12
Jit-ppq5 (%)	1.23	0.42	1.02	0.19	2.50	.02	0.68
Jit-mean (%)	1.62	0.42	1.43	0.27	2.09	.04	0.55
Shim-local (%)	12.25	2.24	9.00	1.28	6.89	< .001	1.85
Shim-apq3 (%)	4.25	1.04	3.19	0.78	4.47	< .001	1.16
Shim-apq5 (%)	6.83	2.03	4.24	0.65	6.66	< .001	1.93
Shim-apq11 (%)	13.48	4.03	8.67	1.14	6.28	< .001	1.86
Shim-mean (%)	9.21	1.90	6.27	0.88	7.68	< .001	2.11
F_1 (Hz)	589	44.58	608	57.74	1.41	.16	0.37
F_2 (Hz)	1657	69.26	1800	69.34	8.02	< .001	2.07
F_3 (Hz)	2723	99.03	2796	118.29	2.59	.01	0.67
F_4 (Hz)	3770	159.32	3944	82.69	5.31	< .001	1.44
F_d (Hz)	1060	54.47	1112	24.51	4.74	< .001	1.29
F_p	-0.32	0.58	0.33	0.51	4.65	< .001	1.2
HNR	10.93	1.93	14.57	1.50	8.15	< .001	2.12

Table D4. Correlations between physical and acoustic measures for male speakers. N = 30.

	Age	Height	Weight	BMI	Grip	SHR	2D:4D	FA
F_0-mean	-.17	-.20	-.10	.02	.27	-.20	.01	-.20
F_0-SD	-.04	.00	.02	.01	-.02	-.34	-.34	-.11
F_0-min	-.38	-.15	-.07	.03	.19	-.14	-.19	.38*
F_0-max	.04	.27	.23	.05	.10	-.42*	-.17	-.19
Jit-loc	.14	-.03	.07	.12	-.23	.09	.08	-.16
Jit-rap	.08	-.14	.15	.34	-.19	.09	-.14	-.15
Jit-ppq5	-.09	-.05	.10	.18	-.31	.00	-.11	-.11
Jit-mean	.05	-.07	.11	.21	-.26	.07	-.04	-.15
Shim-loc	-.08	.01	-.08	-.14	-.16	-.19	.18	-.23
Shim-apq3	-.03	-.11	-.07	-.13	-.08	-.20	-.10	-.16
Shim-apq5	-.19	-.07	-.03	.01	-.14	-.14	.06	-.25
Shim-apq11	.06	.31	.02	-.25	.26	-.15	.30	-.28
Shim-mean	-.05	.13	-.03	-.17	.04	-.20	.21	-.30
F_1	-.11	.11	-.01	-.11	-.08	.07	.15	.02
F_2	.08	-.07	-.04	.01	-.07	.26	.21	.17
F_3	.25	-.26	.09	.10	.02	.51**	.18	.32
F_4	.37*	-.05	.20	.33	.22	.48*	-.17	.15
F_d	.39*	-.08	.20	.35	.23	.45*	-.21	.14
F_p	.21	-.10	.05	.15	.06	.46*	.10	.18
HNR	.04	.21	.11	-.01	.11	.09	-.07	.03

*p < .05, **p < .01

Table D5. Correlations between physical and acoustic measures for female speakers. N = 30.

	Age	Height	Weight	BMI	Grip	WHR	2D:4D	FA
F_0-mean	.11	.00	-.13	-.15	-.19	.01	-.28	.24
F_0-SD	.02	.08	-.07	-.10	.09	.05	.19	.09
F_0-min	-.12	.18	.01	-.07	.23	-.13	.05	.20
F_0-max	.13	.09	.06	.05	-.02	-.14	.19	-.13
Jit-loc	.31	.14	-.09	-.06	-.12	-.14	.13	.23
Jit-rap	.27	-.14	-.15	-.13	-.06	-.04	.16	.22
Jit-ppq5	.23	-.13	-.11	-.09	-.07	-.03	.16	.24
Jit-mean	.28	-.14	-.11	-.09	-.09	-.09	.15	.23
Shim-loc	-.01	-.16	-.25	-.23	.16	.20	.12	.34
Shim-apq3	.12	-.11	-.25	-.25	.18	.24	.05	.32
Shim-apq5	-.02	-.13	-.24	-.23	.19	.24	.17	.38*
Shim-apq11	-.11	-.09	-.06	-.03	.23	-.02	.23	.14
Shim-mean	-.01	-.14	-.21	-19	.21	.16	.16	.31
F_1	-.09	-.28	-.03	.08	-.31	.01	-.42*	-.03
F_2	.19	-.12	.01	.05	-.05	.13	-.03	-.30
F_3	.29	.16	.19	.16	.21	-.23	-.03	-.36
F_4	-.09	-.18	-.31	-.33	-.18	.27	.00	.20
F_d	-.03	.20	-.33	-.43*	.04	.29	.33	.24
F_p	.13	-.15	-.02	.03	-.09	.04	-.18	-.22
HNR	.01	.16	.07	.02	-.05	-.05	-.15	-.39*

*p < .05

Appendix E: Acoustic analyses results for the vowel sound vocal samples

Table E1. Comparison of male speakers' acoustic measures for each measurement session.

	Session 1		Session 2				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	110.4	10.96	108.8	9.14	1.09	.28	0.20
F_0-SD (Hz)	10.14	5.96	9.73	4.85	0.28	.78	0.05
F_0-min (Hz)	94	9.54	94.87	8.31	0.80	.43	0.15
F_0-max (Hz)	142.1	38.09	145.9	25.21	0.41	.68	0.08
F_1 (Hz)	592	97.33	591	103.25	0.03	.97	0.08
F_2 (Hz)	1888	108.16	1874	85.23	0.56	.58	0.19
F_3 (Hz)	2864	161	2743	149.98	5.17	.001	0.18
F_4 (Hz)	3666	192.52	3656	177.03	0.32	.75	0.13
Jit-local (%)	2.10	0.91	2.00	0.62	0.45	.66	0.04
Jit-rap (%)	0.99	0.66	0.86	0.30	1.04	.31	0.14
Jit-ppq5 (%)	0.93	0.51	0.84	0.30	1.00	.32	0.01
Shim-local (%)	10.36	2.93	9.99	2.91	0.69	.49	0.01
Shim-apq3 (%)	3.63	1.03	3.56	1.27	0.21	.84	0.10
Shim-apq5 (%)	4.97	1.18	5.21	1.77	0.77	.45	0.94
Shim-apq11 (%)	9.28	2.44	9.32	3.39	0.07	.95	0.06
HNR	12.08	3.10	12.58	2.56	1.16	.26	0.21

Table E2. Comparison of female speakers' acoustic measures for each measurement session.

	Low fertility		High fertility				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	204	20.2	202	13.9	0.94	.35	0.17
F_0-SD (Hz)	22	11.1	23.3	11.82	0.53	.6	0.10
F_0-min (Hz)	151.3	29.1	148.6	33.5	0.34	.74	0.06
F_0-max (Hz)	291.2	73.8	290.4	58.86	0.04	.97	0.01
F_1 (Hz)	610	81.3	585	64.0	1.70	.10	0.02
F_2 (Hz)	1993	106.71	1972	85.0	1.02	.32	0.22
F_3 (Hz)	2875	147.7	2852	126.7	1.33	.19	0.04
F_4 (Hz)	3929	150.43	3929	150.43	0.06	.95	0.29
Jit-local (%)	1.54	0.62	1.54	0.62	0.09	.93	0.15
Jit-rap (%)	3.80	13.99	1.52	0.28	1.20	.24	0.14
Jit-ppq5 (%)	0.72	0.32	0.70	0.16	0.22	.83	0.02
Shim-local (%)	5.72	1.79	6.23	1.54	1.57	.13	0.31
Shim-apq3 (%)	2.72	0.97	2.58	0.85	0.83	.41	0.19
Shim-apq5 (%)	3.16	1.08	3.04	0.96	0.74	.46	0.24
Shim-apq11 (%)	4.76	1.76	4.74	1.34	0.11	.92	0.01
HNR	16.81	2.55	16.06	3.28	1.40	.17	0.25

Table E3. Comparison of anthropometric and acoustic measures between male and female participants.

	Males		Females				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	110	9.22	203	15.38	28.48	< .001	7.58
F_0-SD (Hz)	9.93	3.63	22.67	9.42	6.91	< .001	1.95
F_0-min (Hz)	94.43	3.63	149.95	23.00	12.41	< .001	4.17
F_0-max (Hz)	144	20.16	290.82	46.46	15.88	< .001	4.41
Jit-local (%)	2.04	0.62	1.53	0.36	3.925	< .001	1.05
Jit-rap (%)	0.922	0.37	2.26	6.98	1.05	.30	0.36
Jit-ppq5 (%)	0.88	0.35	0.71	0.18	2.45	.02	0.66
Jit-mean (%)	1.28	0.43	1.50	2.31	0.51	.62	0.16
Shim-local (%)	10.18	2.52	5.98	1.41	7.97	< .001	2.14
Shim-apq3 (%)	3.59	0.83	2.65	0.78	4.56	< .001	1.18
Shim-apq5 (%)	5.09	1.23	3.10	0.93	7.05	< .001	1.84
Shim-apq11 (%)	9.30	2.34	4.75	1.41	9.13	< .001	2.43
Shim-mean (%)	7.04	1.55	4.12	1.06	8.52	< .001	2.24
F_1 (Hz)	591	85.89	597	60.21	0.32	.75	0.08
F_2 (Hz)	1881	69.7	1982	77.99	5.32	< .001	1.37
F_3 (Hz)	2804	141.8	2864	128.36	1.71	.09	0.44
F_4 (Hz)	3661	165.82	3929	127.09	7.04	< .001	1.83
F_d (Hz)	1023	42.51	1111	40.79	8.13	< .001	2.10
F_p	-0.56	0.73	0.11	0.47	4.19	< .001	1.11
HNR	12.33	2.58	16.44	2.53	6.22	< .001	1.61

Table E4. Correlations between physical and acoustic measures for male speakers. N = 30.

	Age	Height	Weight	BMI	Grip	SHR	2D:4D	FA
F_0-mean	-.03	-.08	-.05	-.01	.32	-.11	-.02	-.12
F_0-SD	.20	.11	.10	.04	-.18	-.25	-.25	.23
F_0-min	-.23	-.06	-.01	.04	.23	-.22	-.06	-.28
F_0-max	-.02	.22	.05	-.12	-.25	-.38*	-.18	-.16
Jit-loc	-.03	-.24	-.15	-.01	-.36*	.05	-.29	.01
Jit-rap	-.17	-.22	-.18	-.06	-.40*	-.04	-.40*	-.03
Jit-ppq5	.02	-.21	-.17	-.05	-.34	.16	-.23	.16
Jit-mean	-.06	-.23	-.17	-.03	-.37*	.05	-.31	.04
Shim-loc	-.24	-.10	-.11	-.07	-.17	-.18	-.33	-.29
Shim-apq3	-.11	-.09	-.13	-.11	-.03	-.11	-.06	-.30
Shim-apq5	-.14	.02	.07	.06	-.05	-.10	-.18	-.32
Shim-apq11	-.18	.20	.22	.13	-.06	-.17	-.26	-.13
Shim-mean	-.21	.03	.04	.02	-.11	-.17	-.28	-.27
F_1	.16	.01	.15	.20	.05	.50**	.18	.24
F_2	.08	-.14	.09	.24	-.02	.14	.14	.37*
F_3	.18	-.12	.07	.20	-.01	.53**	.14	.38*
F_4	.30	-.03	.13	.21	.15	.63**	.16	.27
F_d	.27	-.04	.06	.14	.16	.49**	.09	.19
F_p	.20	*.07	.12	.23	.04	.53**	.19	.35
HNR	.17	.36*	.17	-.06	.12	.12	.25	-.01

*p < .05, **p < .01

Table E5. Correlations between physical and acoustic measures for female speakers. N = 30.

	Age	Height	Weight	BMI	Grip	WHR	2D:4D	FA
F_0-mean	.16	-.09	-.09	-.07	-.33	-.06	-.11	.07
F_0-SD	.12	.07	-.02	-.04	-.03	.05	.10	-.21
F_0-min	-.13	.05	-.07	-.14	-.06	.00	.09	.08
F_0-max	.32	.32	.29	.21	-.14	-.26	.02	-.39*
Jit-loc	.01	-.01	.01	.04	.00	-.09	-.03	-.05
Jit-rap	.39*	.28	.32	.23	-.13	-.21	-.05	-.08
Jit-ppq5	-.16	-.03	-.08	-.08	.03	.10	-.09	.01
Jit-mean	.39	.28	.32	.23	-.13	-.21	-.05	-.08
Shim-loc	-.17	-.17	-.12	-.06	-.03	.6	.10	.06
Shim-apq3	-.19	-.23	-.19	-.12	-.09	.14	-.10	.13
Shim-apq5	-.18	-.27	-.27	-.20	-.12	.20	.02	.15
Shim-apq11	-.09	-.15	-.20	-.18	-.01	-.01	.07	.13
Shim-mean	-.16	-.21	-.20	-.15	-.06	.08	.06	.12
F_1	.09	-.09	-.18	-.2	-.27	.10	-.08	.18
F_2	.02	.04	-.05	-.10	.17	-.13	-.11	-.22
F_3	.33	-.12	-.15	-.14	-.12	.09	-.12	-.28
F_4	.02	-.10	-.31	-.41*	-.14	.44*	-.13	-.07
F_d	-.03	.04	-.24	-.33	-.01	.41*	-.09	-.16
F_p	.20	-.07	-.25	-.31	-.13	.17	-.16	-.17
HNR	.28	.08	.10	.09	-.08	-.05	.22	-.07

*p < .05

Appendix F: Acoustic analyses results for the sustained vowel vocal sample

Table F1. Comparison of male speakers' acoustic measures for each measurement session.

	Session 1		Session 2				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	109.5	9.61	112.47	16.92	1.16	.26	0.21
F_0-SD (Hz)	2.27	1.38	3.42	4.04	1.59	.12	0.29
F_0-min (Hz)	102.43	7.75	104.30	14.09	0.78	.44	0.14
F_0-max (Hz)	120.13	14.16	130	30.09	1.62	.12	0.30
F_1 (Hz)	681	73.74	649	133.97	1.10	.28	0.20
F_2 (Hz)	1331	143.56	1333	108.18	0.07	.95	0.01
F_3 (Hz)	2528	179.26	2540	173.90	0.37	.72	0.07
F_4 (Hz)	3640	244.81	3572	184.82	1.34	.19	0.24
Jit-local (%)	0.91	0.95	0.72	0.64	0.90	.37	0.16
Jit-rap (%)	0.43	0.45	0.37	0.27	0.71	.48	0.13
Jit-ppq5 (%)	0.37	0.18	0.43	0.34	0.86	.39	0.16
Shim-local (%)	4.75	2.15	5.03	3.16	0.42	.68	0.08
Shim-apq3 (%)	2.19	0.85	2.36	1.42	0.64	.53	0.12
Shim-apq5 (%)	2.51	0.82	3.11	1.84	1.80	.08	0.33
Shim-apq11 (%)	3.64	1.22	4.24	2.78	1.07	.29	0.20
HNR	15.98	5.84	16.32	6.04	0.34	.73	0.06

Table F2. Comparison of female speakers' acoustic measures for each measurement session.

	Low fertility		High fertility				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	199.3	25.15	207.13	28.04	3.74	< .001	0.68
F_0-SD (Hz)	6.13	8.26	3.34	2.66	2.26	.03	0.41
F_0-min (Hz)	184.83	29.39	190.03	34.63	1.41	.17	0.26
F_0-max (Hz)	215.37	22.44	218.73	28.32	1.22	.23	0.22
F_1 (Hz)	791.7	75.34	798.87	101.81	0.54	.15	0.10
F_2 (Hz)	1565	110.77	1549	118.02	1.90	.53	0.35
F_3 (Hz)	2758	303.09	2701	340.39	0.76	.25	0.14
F_4 (Hz)	3874	252.79	3810	305.55	1.47	.49	0.27
Jit-local (%)	0.49	0.18	0.72	0.79	1.49	.52	0.27
Jit-rap (%)	0.36	0.41	0.31	0.14	0.64	.70	0.12
Jit-ppq5 (%)	0.27	0.09	0.30	0.11	1.16	.78	0.21
Shim-local (%)	4.00	2.04	3.74	1.71	0.70	.59	0.13
Shim-apq3 (%)	2.20	1.32	2.25	1.06	0.65	.07	0.12
Shim-apq5 (%)	2.34	1.34	2.25	1.06	0.39	.45	0.07
Shim-apq11 (%)	2.85	1.39	2.78	1.18	0.28	.15	0.05
HNR	20.48	3.38	19.51	3.96	1.46	.16	0.27

Table F3. Comparison of anthropometric and acoustic measures between male and female participants.

	Males		Females				
Trait	*M*	*SD*	*M*	*SD*	*t* (29)	*p*	*d*
F_0-mean (Hz)	111	11.83	203.2	26.00	17.68	< .001	4.88
F_0-SD (Hz)	2.85	2.27	4.74	5.13	1.85	.07	0.51
F_0-min (Hz)	103.37	9.28	187.43	30.48	14.45	< .001	4.23
F_0-max (Hz)	125.07	16.69	217.05	24.40	17.04	< .001	4.48
Jit-local (%)	0.82	0.58	0.60	0.39	1.71	.09	0.45
Jit-rap (%)	0.40	0.28	0.33	0.22	1.01	.32	0.26
Jit-ppq5 (%)	0.40	0.18	0.28	0.08	3.15	< .001	0.89
Jit-mean (%)	0.54	0.33	0.40	0.17	1.95	.06	0.53
Shim-local (%)	4.89	2.00	3.87	1.55	2.20	.03	0.57
Shim-apq3 (%)	2.28	0.91	2.12	0.95	0.67	.51	0.17
Shim-apq5 (%)	2.81	1.10	2.29	1.00	1.90	.06	0.49
Shim-apq11 (%)	3.94	1.48	2.82	1.11	3.32	< .001	0.87
Shim-mean (%)	3.48	1.24	2.75	1.14	2.29	.03	0.59
F_1 (Hz)	665	74.72	795	81.87	6.44	< .001	1.66
F_2 (Hz)	1332	102.97	1557	111.98	8.11	< .001	2.09
F_3 (Hz)	2537	153.04	2729	250.25	3.65	< .001	0.97
F_4 (Hz)	3606	166.19	3842	253.84	4.26	< .001	1.12
F_d (Hz)	980	62.56	1016	90.04	1.76	.08	0.46
F_p	-0.47	0.35	0.48	0.47	8.83	< .001	2.33
HNR	16.15	5.29	19.99	3.20	3.41	< .001	0.90

Table F4. Correlations between physical and acoustic measures for male speakers. N = 30.

	Age	Height	Weight	BMI	Grip	SHR	2D:4D	FA
F_0-mean	-.19	-.20	-.42*	-.40*	.00	-.22	-.18	.29
F_0-SD	.04	-.09	.14	.26	.15	.34	.00	.20
F_0-min	-.22	-.33	-.34	-.19	.20	-.32	-.20	.22
F_0-max	.08	-.22	-.18	.04	.24	.11	.10	.27
Jit-loc	-.04	-.13	.20	.38*	-.21	.15	-.23	.21
Jit-rap	-.08	-.14	.17	.36*	-.26	.01	-.28	.14
Jit-ppq5	.09	-.18	.01	.16	-.33	.22	.22	.15
Jit-mean	-.03	-.15	.16	.35	-.25	.13	-.25	.18
Shim-loc	-.05	-.21	.00	.17	-.08	.12	-31	.02
Shim-apq3	-.05	-.15	-.04	.08	-.02	.02	-.32	-.02
Shim-apq5	.09	.22	-.11	.04	-.07	.10	-.32	-.08
Shim-apq11	.25	-.21	-.07	.08	-.06	.24	.03	-.02
Shim-mean	.07	-.22	-.05	.12	-.07	.14	-.24	-.02
F_1	-.26	.01	-.29	-.39*	.01	-.08	.05	.12
F_2	.30	-.25	-.11	.08	-.03	.22	-.04	.41*
F_3	.08	-.35	-.20	.02	-.12	.31	-.13	.35
F_4	.18	-.27	-.05	.19	-.09	.44*	-.34	.23
F_d	.26	-.23	.08	.32	-.08	.42	-.32	.16
F_p	.09	-.36	-.27	-.04	-.11	.35	-.20	.47**
HNR	.19	.07	-.10	-.19	-.11	-.02	.26	.24

*p < .05, **p < .01

Table F5. Correlations between physical and acoustic measures for female speakers. N = 30.

	Age	Height	Weight	BMI	Grip	WHR	2D:4D	FA
F_0-mean	.27	-.10	-.04	.01	-.08	-.11	-.24	-.22
F_0-SD	-.19	.12	-.08	-.16	-.05	-.07	-.15	.30
F_0-min	.16	-.27	-.05	.08	-.14	-.06	-.11	-.27
F_0-max	.28	-.03	.06	.09	-.09	-.17	-.24	-.21
Jit-loc	.01	.31	.19	.07	.08	.16	.16	.05
Jit-rap	-.16	-.24	-.22	-.19	-.28	.12	-.18	.04
Jit-ppq5	-.11	.06	-.02	-.07	.15	.21	.18	.04
Jit-mean	-.08	.14	.04	-.04	-.04	.21	.07	.06
Shim-loc	-.27	-.03	-.30	-.36*	.04	.46**	-.01	.31
Shim-apq3	-.26	.03	-.21	-.29	.03	.44*	-.07	.27
Shim-apq5	-.28	-.01	-.23	-.29	.01	.45	-.06	.27
Shim-apq11	-.32	-.03	-.28	-.34	.01	.47**	.00	.32
Shim-mean	-.28	-.01	-.26	-.32	.02	.46*	-.03	.30
F_1	.04	-.32	-.23	-.11	-.11	-.07	-.08	-.04
F_2	.27	.21	.21	.13	.06	-.07	.01	-.20
F_3	.02	.07	.16	.16	.07	-.07	.06	.16
F_4	-.22	.07	.00	-.06	-.22	.07	-.11	-.05
F_d	-.22	.16	.07	-.02	-.17	.08	-.10	-.03
F_p	.01	.02	.06	.05	-.09	-.05	-.03	-.18
HNR	.14	-.06	-.09	-.06	-.03	.18	.12	-.20

*p < .05, **p < .01

Appendix G: Information sheet for participants in Study 2

Information Sheet

Perception of Voices

You are invited to participate in the research project *Perception of Voices.*

The aim of this project is to examine the judgements people make about a person based on hearing that person's voice.

Your involvement in this project will be to listen to a number of recorded voices and answer a series of questions about each of the voices and what characteristics you think the people speaking might possess.

Should you decide to participate you have the right to withdraw from the project at any time, including withdrawal of any information provided.

The results of this project may be published but you are assured of complete anonymity and confidentiality.

The project is being carried out by Brad Miles as part of the requirements for PhD study and is being conducted under the supervision of Professor Lucy Johnston. They can be contacted at the addresses below and will be happy to discuss any concerns that you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Brad Miles
Department of Psychology
University of Canterbury
3642987 x7704
brad.miles@pg.canterbury.ac.nz

Professor Lucy Johnston
Department of Psychology
University of Canterbury
3642967
lucy.johnston@canterbury.ac.nz

Appendix H: Participant response sheet for Study 2

Response Sheet

Perception of Voices

a) Age:

b) Sex:

c) Do you have any problems with your hearing?

1.

a) Do you think the speaker is **MALE** or **FEMALE**? (please circle one)

b) How old do you think the speaker is? __________

c) Please place a mark on the horizontal line below indicating how tall you think the speaker is:

Very Short |---------------------------------|---------------------------------| Very Tall

d) Please place a mark on the horizontal line below indicating how heavy you think the speaker is:

Very Light |---------------------------------|---------------------------------| Very Heavy

e) Do you think the speaker had an accent? **YES / NO** (please circle one)

f) Do you recognise the speaker? **YES / NO** (please circle one)

Appendix I: Participant response sheet for Study 3 and Study 4

Response Sheet

Perception of Voices

a) Age:

b) Sex: male / female

c) Please indicate your sexual orientation by circling a number on the scale below:

exclusively homosexual	1	2	3	4	5	6	7	exclusively heterosexual

d) Do you have any problems with your hearing?

Please answer the following questions by circling a number on the scales provided.

a) How attractive do you think the voice sounds?

Very unattractive 1 2 3 4 5 6 7 Very Attractive

b) How pleasant do you think the voice sounds?

Very unpleasant 1 2 3 4 5 6 7 Very Pleasant

c) How sexy do you think the voice sounds?

Very unsexy 1 2 3 4 5 6 7 Very Sexy

d) How feminine do you think the speaker is?

Not feminine 1 2 3 4 5 6 7 Very feminine

e) How masculine do you think the speaker is?

Not masculine 1 2 3 4 5 6 7 Very masculine

f) How physically attractive do you think the speaker is?

Very unattractive 1 2 3 4 5 6 7 Very Attractive

g) Do you recognise the speaker? **YES / NO** (please circle one)

Appendix J: Information sheet for female participants in Study 3 and Study 4

Information Sheet

Perception of Voices

You are invited to participate in the research project *Perception of Voices.*

The aim of this project is to examine the judgements people make about a person based on hearing that person's voice.

Your involvement in this project will be to listen to a number of recorded voices and answer a series of questions about each of the voices and what characteristics you think the people speaking might possess.

Because judgments may change over time and in association with hormone levels, you will be asked to complete two listening sessions at specific times across the menstrual cycle. To help achieve this timing, you will be asked a few questions about your menstrual cycle. Each listening session will take around 20 minutes, and as remuneration for your time and participation you will receive a $10 voucher.

Should you decide to participate you have the right to withdraw from the project at any time, including withdrawal of any information provided. The results of this project may be published but you can be assured of complete confidentiality.

The project is being carried out by Brad Miles as part of the requirements for PhD study, and is being conducted under the supervision of Professor Lucy Johnston. They can be contacted at the addresses below and will be happy to discuss any concerns that you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Brad Miles
Department of Psychology
University of Canterbury
3642987 x3409
brad.miles@pg.canterbury.ac.nz

Dr Lucy Johnston
Department of Psychology
University of Canterbury
3642967
lucy.johnston@canterbury.ac.nz

Appendix K: Information sheet for male participants in Study 3 and Study 4

Information Sheet

Perception of Voices

You are invited to participate in the research project *Perception of Voices.*

The aim of this project is to examine the judgements people make about a person based on hearing that person's voice.

Your involvement in this project will be to listen to a number of recorded voices and answer a series of questions about each of the voices and what characteristics you think the people speaking might possess.

Because judgments may change over time, you will be asked to complete two listening sessions held two weeks apart. Each listening session will take around 20 minutes, and as remuneration for your time and participation you will receive a $10 voucher.

Should you decide to participate you have the right to withdraw from the project at any time, including withdrawal of any information provided. The results of this project may be published but you can be assured of complete confidentiality.

The project is being carried out by Brad Miles as part of the requirements for PhD study, and is being conducted under the supervision of Professor Lucy Johnston. They can be contacted at the addresses below and will be happy to discuss any concerns that you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Brad Miles
Department of Psychology
University of Canterbury
3642987 x3409
brad.miles@pg.canterbury.ac.nz

Dr Lucy Johnston
Department of Psychology
University of Canterbury
3642967
lucy.johnston@canterbury.ac.nz

Appendix L: Participant consent form for Study 5

CONSENT FORM

Impression Formation Study

I have understood the description of the above named project. On this basis I agree to participate in the project, and I consent to the publication of the results of the project with the understanding that anonymity will be preserved.

I understand also that I may at any time withdraw from the project, including withdrawal of any information that I have provided.

Name (please print):

Signature:

Date:

Appendix M: Participant debriefing sheet for Study 5

Debriefing sheet

Impression Formation Study

Thank you for participating in the research project *Impression Formation Study*.

The aim of this project is to examine some of the features that influence the way in which we process information about other people.

Your involvement in this project has been to watch an audio/visual slide-show depicting a series of photographs of target individuals along with spoken sentences. Your task was to watch the slideshow and form an impression of the people shown. Some of slides contained faces and voices that were matched (e.g., young face/young voice or female face/female voice) and some contained faces and voices that were mis-matched (e.g., old face/young voice or female face/male voice). After viewing the slideshow you were asked to recall which face was paired with which sentence.

Previous research has shown that mis-matched or incongruent information is sometimes better recalled than matched or congruent information. The current research aims to determine whether information presented when faces and voices are mis-matched is more memorable than information presented when faces and voices are matched.

You have the right to withdraw from the project at any time, including withdrawal of any information provided. The results of this project may be published but you may be assured of complete confidentiality.

The project is being conducted by Brad Miles as part of the requirements of PhD study, and under the supervised of Professor Lucy Johnston. They can be contacted at the addresses below and will be happy to discuss any concerns that you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Brad Miles
Department of Psychology
University of Canterbury
3642987 x7704
brad.miles@pg.canterbury.ac.nz

Prof. Lucy Johnston
Deputy Vice-Chancellor's Office
University of Canterbury
lucy.johnston@canterbury.ac.nz

Appendix N: List of target statements used in Study 5

1. I play tennis
2. I’m wearing new shoes
3. I have a dental appointment
4. I’m going on holiday
5. I’m wearing woollen socks
6. I drive a red car
7. I own a blue bike
8. I drink coffee
9. I go to the gym twice a week
10. I have a friend named Sam
11. I donate to charity
12. I watch the news on TV
13. I have a younger sister
14. I eat toast for breakfast
15. I have a dog named Max
16. I read mystery novels
17. I went to New York last year
18. I have family in Scotland
19. My sister’s name is Rosie
20. My lucky number is four
21. My favourite food is ice-cream
22. I play the piano
23. My favourite flowers are tulips
24. I have two brothers

Appendix O: Participant response sheet for Study 5

Impression Formation Study: Response Sheet

Age:

Sex: **MALE / FEMALE**

Do you have any problems with your hearing or vision? **YES / NO**

Please indicate which target individual was associated with each of the following statements by writing that target's number in the space next to the statement.

_____ I drive a red car
_____ I have a younger sister
_____ I'm wearing new shoes
_____ I have two brothers
_____ My favourite food is ice-cream
_____ I watch the news on TV
_____ I read mystery novels
_____ I have a friend named Sam
_____ I'm going on holiday
_____ I play tennis
_____ I go to the gym twice a week
_____ My favourite flowers are tulips
_____ I have family in Scotland
_____ I play the piano
_____ I donate to charity
_____ I have a dog named Max
_____ I have a dental appointment
_____ I'm wearing woollen socks
_____ I eat toast for breakfast
_____ I went to New York last year
_____ My lucky number is four
_____ I own a blue bike
_____ My sister's name is Rosie
_____ I drink coffee

If you recognised any of the targets, please write their number(s) here: ________________