Contextual musicality: vocal modulation and its perception in human social interaction

Juan David Leongómez
July 2014

Thesis submitted for the degree of PhD

Department of Psychology, School of natural Sciences

UNIVERSITY OF STIRLING

ACKNOWLEDGEMENTS

I would first like to thank Dr S. Craig Roberts, whose support, knowledge, ideas, and encouragement have been vital since the moment I decided to dedicate my life to research. He supported me since I left my comfort zone of music, and ventured into the uncharted territory that science appeared to me back in 2008 when Craig accepted me onto an MSc in Evolutionary Psychology at Liverpool University. He has continued to support me through this as well as my PhD, and I cannot stress enough the positive impact and inspiration he has had on this thesis.

I want to express my most sincere gratitude to Professor Phyllis C. Lee and Professor Stuart Semple, who acted as examiners during the viva voce examination, for their insightful comments and questions that were not only vital in shaping this final version of the thesis but further inspired me beyond the main scope of the works presented.

Additionally, thanks are due to Dr Anthony C. Little, my second supervisor, and his student (and my friend) Viktoria R. Mileva, who have both been instrumental in helping me develop interesting ideas and translating them into well-designed experiments. Also, I want to thank all my collaborators in Prague, Jakub Binter, Petra Stolařová, and Lydie Kubicová, who have become more than just colleagues. A big shout out to Dr Jan Havlíček and Dr Kateřina Klapilová, without whose friendship, support and encouragement this thesis would be of a much lower standard. Some parts

of this thesis would have been impossible, and much less fun, without the help of Dr Alice Murray, whose collaboration was crucial.

To all of the participants who endured some of the most boring experiments known to man, which are here included, you deserve not only my gratitude, but a round of applause, and my sincerest apologies. The same is true of Eugenio Valderrama, from the Royal Botanic Garden Edinburgh, and his outlandish skills using R, to whom I owe thanks for some of the figures here included. I also want to express my gratitude to all the anonymous reviewers who, despite giving me a hard time, had an enormous positive influence on large portions of this work.

All my friends in the department have made my life in Stirling an amazing experience, and one that I will never forget; Vicky, Simone, Caroline, Mary, Kelly, Ailsa, Michelle, Eoin, Adam, Alice, Herbert, Phil, Steve, and all those whose friendships have been central to my life in Stirling for the last four years, I love you all. Everyone in the Psychology Department, who have created the friendliest of environments, deserves credit for making this a truly enjoyable experience.

To my family, and particularly my mom and dad, whose love and support I simply cannot even attempt to express, all the love and a million thanks. My friends Eugenio, Santi, Hele, Maria, Mafe, have all brought bit of Colombia here with them, essential to keep me from running back.

I also want to express my most sincere gratitude to the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), for funding my research, as well as the people from LASPAU, for dealing with all my questions.

Finally, I want to make an especial recognition to Carl Sagan. In particular his book and TV series Cosmos marked my life as a child and triggered my curiosity and passion for science, and his book Shadows of Forgotten Ancestors revealed to me the wonders and explanatory power of evolution. Without his work, his honesty, curiosity, and his amazing educational work, I probably would not be here.

I’m really out of words to say what you all really deserve, so I hope you are all happy with my feeble attempt.

ABSTRACT

Music and language are both deeply rooted in our biology, but scientists have given far more attention to the neurological, biological and evolutionary roots of language than those of music. Because of this, and probably partially due to this, the purpose of music, in evolutionary terms, remains a mystery. Our brain, physiology and psychology make us capable of producing and listening to music since early infancy; therefore, our biology and behaviour are carrying some of the clues that need to be revealed to understand what music is “for”. Furthermore, music and language have a deep relationship, particularly in terms of cognitive processing, that can provide clues about the origins of music.

Non-verbal behaviours, including voice characteristics during speech, are an important form of communication that enables individual recognition and assessment of the speaker’s physical characteristics (including sex, femininity/masculinity, body size, physical strength, and attractiveness). Vocal parameters, however, can be intentionally varied, for example altering the intensity (loudness), rhythm and pitch during speech. This is classically demonstrated in infant directed speech (IDS), in which adults alter vocal characteristics such as pitch, cadence and intonation contours when speaking to infants. In this thesis, I analyse vocal modulation and its perception in human social interaction, in different social contexts such as courtship and authority ranking relationships. Results show that specific vocal modulations, akin to those of IDS, and perhaps music, play a role in communicating courtship intent.

Based on these results, as well the body of current knowledge, I then propose a model for the evolution of musicality, the human capacity to process musical information, in relation to human vocal communication. I suggest that musicality may not be limited to specifically musical contexts, and can have a role in other domains such as language, which would provide further support for a common origin of language and music. This model supports the hypothesis of a stage in human evolution in which individuals communicated using a music-like protolanguage, a hypothesis first suggested by Darwin.

TABLE OF CONTENTS

CHAPTER 1:

GENERAL INTRODUCTION

Darwin (1871) portrayed music as one of the most mysterious human abilities, because it is a human universal with no obvious function. It is, in fact, a phenomenon that seems to be present in all human cultures, the roots of which can be traced for a few tens of thousands of years, back to the earliest known musical instruments. However, music does not depend on flutes or man-made musical instruments to exist, as humans are equipped and able to sing, use objects and even our own bodies as drums, and to dance, from very early stages of ontogenetic development. By the moment when the earliest known flute –a complex and differentiated instrument– was made (see Hahn & Münzel, 1995), construction abilities and sound experiences must already have been improved and developed for some time. Moreover, before these abilities started to be improved, the social role of music ought to have become relevant –at least by its playful qualities and as an object of pleasure– and, even before, the physiological structures necessary for musical thought and perception must have been developed, referring us to a point even further back in time.

When and why did music appear? With the exception of instruments, music itself cannot be recovered from archaeological sites, making the answer to this question a complicated work. However, we do know that our brain, physiology and psychology make us capable of producing and listening to music. Today, our biology and behaviour are carrying some of the clues that need to be revealed to understand what music is "for".

A number of the recent studies around the evolutionary origins of music have focused on attempts to explain music as a spandrel –for example, Pinker's incendiary description of music as an "evolutionary cheesecake" (Pinker, 1997)– or described as an adaptation (see Brown, 2000b; Miller, 2000). However, the existence of music is still to be understood and, as with any human universal, its study would provide essential answers to our understanding of the human being. But how can research on this area be addressed?

An important area of study from which research on biomusicology can benefit is language: similarly to music, language is deeply rooted into our biology (Hauser, Chomsky, & Fitch, 2002; Mithen, 2006), but historically it has obtained far more attention than music from scientists. What is more, several brain processing areas are shared between music and language (Fitch, 2006b; Masataka, 2009; Mithen, 2006) and there is evidence suggesting that the human brain does not treat language and music as different kinds of stimuli, at least during early stages of ontogenetic development (Koelsch & Siebel, 2005). The deep relationship between language and music in terms of shared neural resources is supported by strong evidence, principally from neuroimaging research (e.g. Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Koelsch et al., 2003; Patel, 2003; Schön, Magne, & Besson, 2004).

These connections between language and music can be beneficial for the study of music in two different dimensions:

First, and more obvious, biomusicology can benefit from the advances made by scientific research on language, which can provide a model for understanding the evolution of music (Fitch, 2006b). Second, the connections and similarities between both phenomena provide evidence for the idea of an ancient link in their evolutionary paths. This is an idea often called musical protolanguage, or music-like protolanguage, which was first suggested by Darwin (1871) and that has been revived by several researchers in many different forms (e.g. Brown, 2000b; Dunbar, 2003b; Fitch, 2006a; Marler, 2000; Merker, 2000a, 2000b, 2001; Mithen, 2006). As Fitch (2006a) points out, the possible link between music and language can, in fact, be tested “by exploring and comparing the cognitive, neural, and genetic mechanisms underlying musical and linguistic abilities in modern humans”.

An important and largely unexplored area is the means by which music and language transmit information. In an extensive review of the empirical research involving the acoustic clues implicated in the expression of emotions in music and vocal expression –the nonverbal aspects of speech– Juslin & Laukka (2003) found many parallels between the acoustic cues involved in the expression of five general emotional categories: anger, fear, happiness, sadness, and tenderness, most of which are related to the tempo or speech rate, the micro-structural regularity or irregularity, and to pitch and loudness variation, suggesting that, in fact, music and vocal expression are at least partially based on the same codes.

This suggests an even more interesting hypothesis, and one that can be tested: musicality, the human capacity to process musical information, may not be limited to specifically musical contexts, and can have a role in other domains such as language, which would provide further support for a common origin of language and music. The main purpose of this thesis is precisely that: to test whether vocal elements typically associated with music, particularly those related to pitch modulation and perception, are present in speech.

This work has five experimental chapters, which are divided into three sections:

The first three experimental chapters (Chapters 2, 3 and 4), which constitute the main body of the dissertation, are related vocal modulation, and its perception, in mate choice contexts. In Chapter 2, using a novel methodology, vocal responses to attractive or unattractive potential partners or competitors were recorded and analysed (Study 1) in two model languages (English and Czech), and then the responses to these recordings by naive listeners were tested (Study 2), across both model languages. Results from this chapter show specific vocal modulations, particularly in fundamental frequency (F_0), that seem to be specific to courtship scenarios, and that can in fact be perceived by listeners, even in the absence of verbal content. In Chapter 3, using a similar scenario to record voices in a courtship scenario, the effects of male body odour and one of its components (androstadienone) on vocal responses were tested. These results are consistent with those of Chapter 2, and show that people respond differently to potential partners in the presence of male body odour and/or androstadienone, and that these differences in response, although subtle, are manifest in the vocal parameters of the speakers. Finally, in Chapter 4, associations between pitch discrimination and partner choice were tested, using an original online experiment to test pitch discrimination, which was then compared to partner satisfaction as well as partners' parenting skills and investment, measured using well established tests. Chapter 4 produced unexpected, yet interesting results, showing that pitch discrimination skills change across the menstrual cycle, peaking during high conception risk days of the menstrual cycle.

Chapter 5, which constitutes the second section, analysed vocal modulation in a different context, not directly related to mate choice: authority ranking relationships. To do this, a novel scenario based on job interviews was used, in which participants responded to putative employers manipulated to appear relatively dominant, prestigious, or average. Results show that vocal modulation occurs depending on the characteristics of the target listener, and that this modulation is dependent on self-perceived status of the speaker. Vocal modulations in this context are, however, different from those found in mate choice scenarios.

The last experimental section, presented in Chapter 6, is a methodological experiment testing the effects of using a mechanical scale for the manipulations of voice frequency in perceptual studies, as opposed to psychoacoustic scales (which are based on perceived pitch). To do this, male and female voices were manipulated by both a mechanical (Hz) and a psychoacoustic (mel) scale, and presented to a panel of raters; changes in perceptions of four domains (attractiveness, masculinity/femininity, body size, age) were measured and compared between the two types of manipulations. The results from this chapter suggest that, although perceptual differences produced by using one type of scale over the other are subtle, they are quantifiable and could bias conclusions regarding voice cues. This is important as the majority of studies that use pitch manipulations to study vocal cues, have used mechanical scales.

To conclude this thesis, the general discussion (Chapter 7) proposes a hypothetical model for the evolution of musicality and its role on complex human vocal communication, based on the results of the experimental chapters, as well as the most current body of knowledge related to this issue. This chapter discusses the problems of the scientific study of music, and its complex relation with language, as well as the most important similarities between complex vocal communication in humans and non-human animals, to then highlight the most important theories for the origins of the capacity for music. The model, which integrates several theories, presents a radically different view of musicality, not limited to specifically musical scenarios, in which this capacity originally evolved as a means to aid parent-infant communication and bonding, and even today plays a role not only in music, but also in infant-directed speech (IDS) as well as some specific contexts in language; the seemingly musical elements found in IDS appear to be akin to those found in language during courtship interactions (Chapters 2,3 and 5), and seem to represent what I have called *contextual musicality*.

CHAPTER 2:
VOCAL MODULATION DURING COURTSHIP

2.1 Introduction

The human voice is remarkably variable. Aside from communication through verbal content, paralinguistic elements of the voice during speech enable individual recognition and assessment of the speaker’s physical characteristics such as sex (Puts, Apicella, & Cárdenas, 2012), body size (Feinberg, Jones, Little, Burt, & Perrett, 2005; Xu, Lee, Wu, Liu, & Birkholz, 2013), physical strength (Sell et al., 2010), femininity (Feinberg, 2008; Feinberg, Jones, DeBruine, et al., 2005), attractiveness (Feinberg, Jones, Little, et al., 2005; Feinberg, Jones, DeBruine, et al., 2005; Xu et al., 2013), conception risk (Pipitone & Gallup, 2008), and sexual maturity (Mulac & Giles, 1996). In humans, perceived attractiveness and mate quality can be manipulated by artificially lowering the pitch of male voices or artificially increasing it in female voices, commensurate with sex-typical vocal properties (Collins, 2000; Feinberg, Jones, Little, et al., 2005). In fact, there is evidence for increased reproductive success in traditional societies for both low-pitched males (Apicella, Feinberg, & Marlowe, 2007), and high-pitched females (Atkinson et al., 2012).

In addition, vocal parameters can be intentionally varied, for example altering the intensity (loudness), rhythm and pitch. The classic example of such intentional modulation is infant directed speech (IDS) (Falk, 2005; Ferguson, 1977), in which adults alter vocal characteristics such as pitch, cadence and intonation contours when speaking to infants. Infants prefer these altered signals over adult-directed speech (Fernald & Kuhl, 1987) and it has been suggested that IDS aids human acquisition of vocal language (Burnham, Kitamura, & Vollmer-Conna, 2002) and might underpin the origins of musicality (Dissanayake, 2000; Trehub, 2003). In human and animal social interactions, modulations of the intensity of speech or vocalisations are often associated with hostility (Collias, 1960; Kudo, 1987) and dominance (Ohala, 1982; Tusing & Dillard, 2000), and changes in intensity contribute to emotional expression (Baker, 2001). Regarding pitch modulations, men lower their voices during competitive interactions when they perceive themselves as physically dominant (Puts, Gaulin, & Verdolini, 2006), and while women have been found to increase voice pitch when directing speech towards attractive faces (Fraccaro et al., 2011), both men and women have also been found to lower their voice pitch when speaking to attractive targets of the opposite sex (Hughes, Farley, & Rhodes, 2010). This suggests that, while more evidence for specific types of modulation is needed (e.g. in the case women responding to attractive opposite-sex stimuli), modulations do actually occur. Similar subtle modulation in voices might be expected in courtship contexts. In fact, there is evidence of vocal differences between speech directed towards romantic partners and same-sex friends which can be detected by listeners (Farley, Hughes, & LaFayette, 2013), and intentional voice manipulations make female voices, but not male voices, sound more attractive (Hughes, Mogilski, & Harrison, 2013; see also Fraccaro et al., 2013). Indeed, such modulations occur in other species including frogs (Ryan, 1980), koalas, *Phascolarctos cinereus* (Charlton, Ellis, Brumm, Nilsson, & Fitch, 2012), fallow deer, *Dama dama* (Charlton & Reby, 2011), red deer, *Cervus elaphus* (Reby et al., 2005; Reby, Charlton, Locatelli, & McComb, 2010), and birds. For example, in the zebra finch, *Taeniopygia guttata*, males sing more rapidly to females than when they sing alone, producing syllables with lower spectral variability (Kao & Brainard, 2006).

Studies aiming to measure the effects that acoustic parameters have on human communication are hampered by the confounding influence of verbal content. To address this issue, many studies record voices enunciating vowel sounds or speaking standard sentences, or measure responses to voices with artificially manipulated vocal parameters (e.g. Feinberg, Jones, Little, et al., 2005; Puts, Hodges, Cárdenas, & Gaulin, 2007). These methodologies have provided important insights into the role that vocal parameters play in human communication. Similarly, to study vocal modulation, and unlike research on animals or IDS (where infants understand little or none of the semantic content), it is necessary to control the confounding influence that verbal content may play. Some studies have used scripted speech (e.g. Fraccaro et al., 2011; Hughes et al., 2010), therefore eliminating prosodic variation in vocal acoustic parameters. Although challenging, testing free, unscripted speech is ideal, as standard sentences may not accurately reflect the levels of natural vocal variation; standardised sentences likely limit the kind of spontaneous paralinguistic variation found in normal free speech, as well as the nuance and range of paralinguistic modulation known at least to occur in IDS, which is characterised by an extreme range of pitches, typically starting from a high pitch and containing many glissandos. Finally, while some studies have successfully tested natural vocal variation during speech (e.g. Hodges-Simeon, Gaulin, & Puts, 2010, 2011), apparent paralinguistic modulation in one language may be underpinned by specific parameters of that language (e.g. rhythm, intonation, and use of specific phonemes). Here I circumvented these issues (i.e. the confounding influence of verbal content, using unscripted speech, and the potential effects of one language in paralinguistic modulation) by adopting a cross-language design involving two model languages.

Based on evolutionary theory and the current knowledge of human voices, I hypothesized that males and females would modulate their acoustic parameters (study

1), depending on the sex and attractiveness of the target, to affect the way in which they would be perceived. Because speech intensity is associated with hostility and dominance (Collias, 1960; Kudo, 1987; Ohala, 1982) I expected participants to speak with increased intensity in responses to same-sex targets, in comparison to opposite-sex targets. Furthermore, because emotional expressiveness is attractive (Sprecher, 1989), and changes in intensity improve emotional expression (Baker, 2001), I expected participants to speak with increased variability in intensity when responding to opposite-sex targets, and especially when those targets were attractive. In addition, based on the body of knowledge produced by studies testing perception of manipulated pitch, I predicted that women, and especially men, would emphasise sex-specific vocal characteristics when responding to attractive individuals of the opposite sex (i.e. lowering F_0 in men, and increasing it in females), and that both sexes would increase F_0 variability, in order to sound more attractive to those attractive targets. Additionally, I predicted that these modulations would be detectable by naive listeners (study 2), and that speakers would sound more attractive when speaking to attractive versus unattractive targets.

2.2 Study 1

First, I tested the possibility that individuals might alter vocal parameters in speech directed at potential romantic partners or competitors depending on the attractiveness of the listener. Recorded voice samples from speakers of two different languages were used to avoid the possibility that apparent paralinguistic modulation in one language might be reinforced by specific parameters of that language (e.g. rhythm, intonation, and use of specific phonemes).

2.2.1 *Materials and Methods*

2.2.1.1 *Participants*

I recruited 110 heterosexual participants who were students at the Universities of Liverpool and Stirling (UK) or Charles University (Czech Republic). Of these, 30 were English speaking males (mean age ± SD = 22.6 ± 4.17), 30 English speaking females (21.8 ± 3.96), 25 Czech speaking males (22.8 ± 2.30), and 25 Czech speaking females (21.8 ± 1.84) not suffering from voice hoarseness or nasal congestion. No participant suffered from speech impediments, and all were fluent in English. All participants signed a written consent form.

2.2.1.2 *Target Videos*

The stimuli were selected from a group of 40 videos, of about 20 seconds length (mean length ± SD = 19.3 ± 2.60), half of which pictured men (mean age ± SD = 22.5 ± 2.41) and half women (22.1 ± 1.65). Individuals were visible from the waist upwards before a white background and were filmed having been asked to introduce themselves to an attractive person of the opposite sex. Each video was rated for attractiveness on a 1 to 7 scale, by an independent panel of 24 opposite-sex raters. From these, the 3 most attractive and the 3 least attractive male and female videos were selected (12 in total). Individuals seen in the videos did not take part in any of the other experiments.

2.2.1.3 *Experimental Procedure*

After the experiment had been explained and written informed consent obtained, participants were shown the twelve target videos, played without sound to avoid possible effects of pitch convergence (Gregory, Green, Carrothers, Dagan, & Webster, 2001), and asked to record a response message to each one. Participants were told that these messages would be presented to opposite-sex participants who would judge them as a potential date: in the case of responses to opposite-sex targets, participants believed their messages were going to be presented to the target they were responding to (i.e. the person in the video), while in the case of responses to same-sex targets, participants were told that their responses were going to be presented to all the opposite-sex targets. Participants were instructed to either explain whether and why they would like to date the person in the video (for opposite-sex targets) or why they should be chosen over the person in the video for a date (for same-sex targets). This scenario was based on a study which produced demonstrable effects on mate preferences (Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004).

After recording their response to each presented target video, participants were debriefed. In total, 1304 recordings were obtained (4 recordings were not collected because the participant recognised the target, and 12 were discarded because of background noise that affected audio quality), with length ranging from 6 to 46 seconds (mean ± SD = 14.70 ± 7.24 s). Additional details are provided in Appendix A.

2.2.1.4 Data Analysis

Each recording was acoustically analysed using Praat© 5.2 to obtain data on intensity (dB) and F_0 (Hz). Values were obtained every 10 ms. F_0 was measured using a noise-resistant autocorrelation method, between 75 and 300 Hz for male voices, and 100 and 500 Hz for female voices. Since recordings were of free speech, I did not analyse formant frequencies as these would be affected by the amount and duration of particular vowels. For intensity, only time points for which the Praat algorithm produced a value of pitch were used; this was done to control for any background noise during silent periods and to ensure that intensity scores were unaffected by differences in pause length or number. Finally, I checked that there were no significant differences in length of recordings after viewing attractive and unattractive targets, or depending on target sex.

Means and standard deviations were then obtained for intensity and F_0, and minimum F_0 for males, for each of the 1304 recordings (descriptive statistics of acoustic measures and length of the recordings are presented in Table A1, in Appendix A). For these values, mean scores were calculated for each participant according to the attractiveness and sex of the target; because each participant responded to three targets of each sex/attractiveness combination, values used in the analysis were the mean of their three responses to same-sex attractive, same-sex unattractive, opposite-sex attractive, and opposite-sex unattractive targets. These were analysed using repeated-measures generalised linear models (GLM) for each parameter (with Bonferroni-adjusted $\alpha = 0.0125$ because I performed 4 analyses), using sex and language of the participant as between-subjects factors, and sex and target attractiveness as within-subjects factors. I report the within-subjects effects involving attractiveness in Table 2.1, reflecting the experimental design; the full models are provided in Table A2 in Appendix A. Post-hoc pairwise comparisons (*t*-tests) were conducted for significant effects of target attractiveness. All tests are two-tailed. Additional details are provided in Appendix A.

2.2.2 Results and Discussion

Analysis revealed that variability in F_0 (F_0 SD) was particularly sensitive to change in social context compared with the other three parameters (Table 2.1). There was a significant main effect of target attractiveness, such that F_0 SD increased after viewing attractive compared with unattractive targets. There were also two significant interactions: between target attractiveness, target sex and participant sex (in which men, but not women, raised F_0 SD after viewing attractive individuals in the opposite-sex condition, Fig. 2.1d), and between target attractiveness and target sex (in which F_0 SD was highest after viewing attractive individuals in the same-sex condition). These interactions indicate that men's F_0 SD was higher in the opposite-sex condition, while women spoke with more variability after viewing attractive romantic competitors (Fig. 2.1d). Previous studies have noted that women are particularly sensitive to attractiveness of perceived competitors, seeking to increase their perceived attractiveness to potential partners relative to other women (Buss & Dedden, 1990; Fisher, 2004), and the differences in F_0 variability that women show after watching same-sex (but not opposite-sex) targets, could be reflecting this. Each of these effects indicate that individuals tended to speak with increased variability in F_0 when motivation was high – in response to perceived attractiveness of potential dates or when competing for a date against an attractive rival. Such variability might serve as a marker of social interest, or help to capture attention of the listener, or could more simply reflect general autonomic arousal in the speaker.

Table 2.1. Context-dependent variation in vocal parameters

Effect	**Vocal parameter**							
	Mean intensity		**Intensity SD**		**Mean F_0**		**F_0 SD**	
Within-subject Effect	*F*	p	*F*	p	*F*	p	*F*	p
TA	0	0.99	**8.18**	**0.005**	4.88	0.029	**68.15**	**<0.001**
TA x PS	0.51	0.476	0.03	0.864	0.83	0.364	0.16	0.687
TA x TS	1.37	0.244	1.98	0.162	0.34	0.563	**9.85**	**0.002**
TA x PS x TS	1.71	0.194	0.19	0.661	3.49	0.065	**17.45**	**<0.001**
TA x L	1.28	0.261	0.01	0.921	**7.27**	**0.008**	2.11	0.15
TA x PS x L	1.41	0.239	1.98	0.163	2.22	0.139	0.02	0.9
TA x TS x L	0.46	0.5	4.08	0.046	0.92	0.339	1.28	0.26
TA x PS x TS x L	1.72	0.193	0.35	0.558	1.01	0.317	0.01	0.921

TA = Target Attractiveness, PS = Participant sex, TS = Target sex, L = Language. Results are from repeated-measures generalized linear models (d.f. = 1, 106 in each case) for each vocal parameter, with Bonferroni adjustment for multiple tests ($\alpha = 0.0125$). Significant effects are in bold. For all results, see table A1 in Appendix A.

Furthermore, there was striking similarity in these patterns of F_0 SD across the two languages. Post hoc tests showed that differences in F_0 SD during responses to attractive and unattractive individuals of the same or opposite sex (shown in Fig. 2.1d) occurred in almost identical patterns in English and Czech speakers. This is further illustrated by the absence of any significant interaction involving target attractiveness and language (Table 2.1, lower panel).

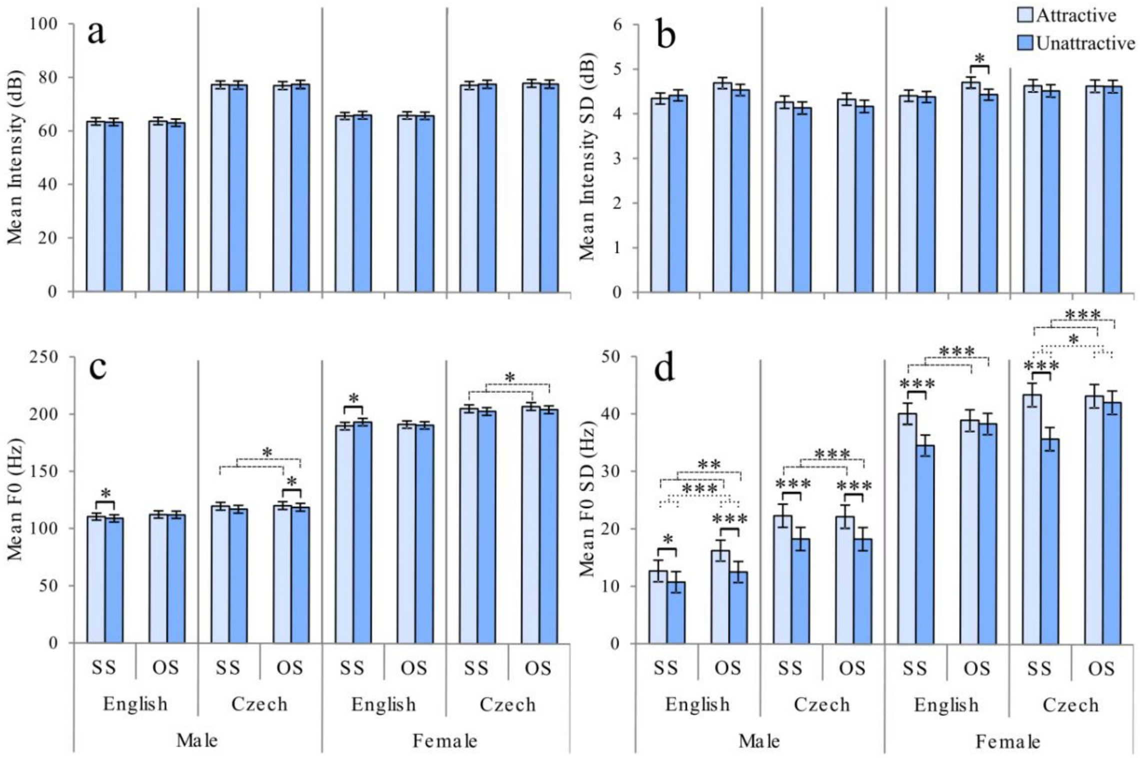


Figure 2.1. Modulation of acoustic parameters in speech towards same and opposite-sex targets, split by attractiveness category of the targets (attractive: light blue bars; unattractive: dark blue bars) and sex of the stimuli (SS: same-sex; OS: opposite-sex). (a) Mean Intensity; (b) Intensity SD; (c) Mean F0; (d) F0 SD. Standard deviation (SD) for intensity and F0 were used as a measure of variability. Bars represent mean ± 1 s.e.m. For interactions, dashed lines represent an effect of target attractiveness (attractive, unattractive); dotted lines represent an effect of target sex (same, opposite). Post-hoc tests, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. For detailed results, see Table A3 in Appendix A.

In contrast, there were few context-dependent differences in the other vocal parameters and no similar consistency across languages (Table 2.1). There were no significant differences in mean intensity. For variability in intensity (intensity SD), there was a significant main effect of target attractiveness, such that participants changed their intensity levels more to attractive individuals, but post hoc tests revealed that this effect was driven mainly by English speakers in the opposite-sex condition (Fig. 2.1b). There was also a near significant (after Bonferroni correction) interaction between target attractiveness, target sex and language, in which higher intensity SD occurred after viewing opposite-sex attractive individuals in English but not Czech speakers. Finally, for mean F_0, there was a significant interaction between target attractiveness and language, in which Czech (but not English) speakers spoke with high mean pitch after viewing attractive individuals; the main effect of target attractiveness was not significant (after Bonferroni correction) but tended towards higher pitch after viewing attractive individuals.

On the basis of previous studies testing perception of manipulated pitch, I had expected that men might lower mean F_0 when speaking to attractive opposite-sex targets because modulation might serve to emphasise sex-typical characteristics (Hughes et al., 2010), but there were no significant interaction effects involving participant or target sex, and the only significant effects for F_0 corresponded to higher, not lower, pitch in the attractive condition (in Czech speakers; Fig. 2.1c). However, absence of this expected result can be explained upon recognizing the relative importance of F_0 variability (Table 2.1; Fig. 2.1): there was a positive correlation between F_0 SD and mean F_0 ($r = 0.46$, $p < 0.001$; Fig. 2.2a). This suggests that increased variability in F_0 results in higher mean F_0, and that the observed tendency towards higher mean pitch may therefore emerge as a consequence of increasing F_0 variability, rather than being a directly modulated parameter.

Despite this, men's minimum $F_0$was significantly lower, in both Czech and English samples, when responding to attractive (M = 82.36 Hz, SD = 6.47) versus unattractive women (M = 86.20 Hz, SD = 9.13) (paired-samples t-test: t54 = 5.41, p < 0.001; Fig. 2.2b) and, in contrast to the relationship between F_0SD and mean F_0, F0 variability and minimum F_0were not significantly correlated (r = -0.11, n = 55, p =

0.44). This suggests that F_0variability and minimum $F_0$are independent parameters which might provide different cues of mating intent and mate quality.

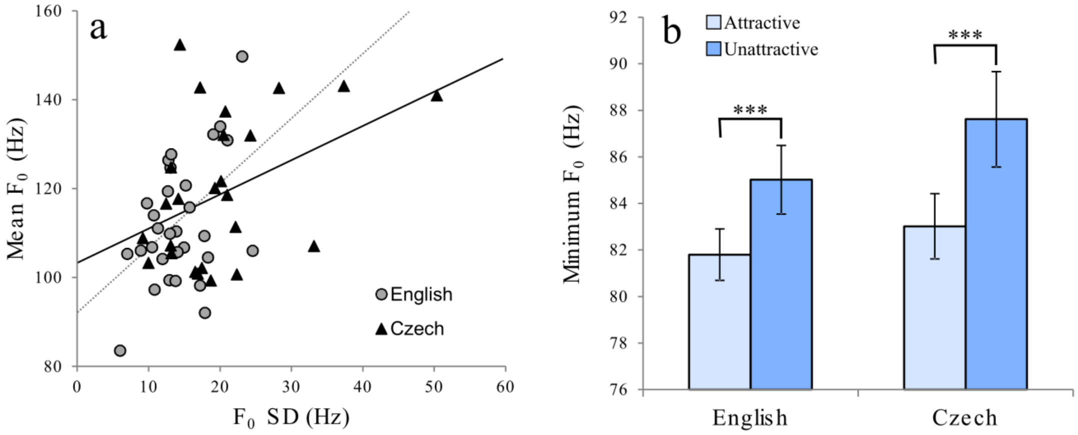


Figure 2.2. Relationships between pitch parameters. (a) Correlation between mean F_0 and F_0 variability (F_0 SD) for men's responses to women (English: $r = 0.45$, n = 30, $p = 0.012$; Czech: $r = 0.41$, n = 25, $p = 0.041$; all individuals: $r = 0.46$, n = 55, $p < 0.001$). (b) Men's minimum F_0 in responses to opposite-sex targets (attractive: light blue bars; unattractive: dark blue bars). Bars represent mean ± 1 s.e.m. ***$p < 0.001$.

2.3 Study 2

Findings from study 1 indicate that paralinguistic parameters vary depending on the attractiveness of the target, but did not test the perception of this modulation. For it to be functionally relevant and have an effect on mate choice, it must be perceptually detectable and influence proceptivity towards the speaker. Study 2 aimed to investigate whether this is indeed the case.

2.3.1 *Materials and Methods*

In order to test whether paralinguistic modulation is detectable and context-specific, while ruling out influence of verbal content, I presented subsets of 10 pairs of recordings from each language group to naive listeners (English participants who do not speak Czech, and vice versa) in a series of forced-choice tests. Recordings were judged for attractiveness by both opposite- as well as same-sex listeners. To test context-specificity, I conducted a confirmatory test, in which recordings were rated for friendliness instead of attractiveness. Additionally, and to test whether differences in judgement are dependent on F_0 modulation, low-pass filtered versions of the recordings were rated for both attractiveness and friendliness in separate tests.

2.3.1.1 Participants

For the test using original voice recordings rated for attractiveness, the final sample included 123 participants judging opposite-sex recordings, and 98 judging same-sex recordings. Opposite-sex listeners were 24 men (mean age ± SD = 29.2 ± 9.29) and 35 women (27.3 ± 8.89) in the English sample, and 24 men (26.5 ± 7.11) and 40 women (26.9 ± 5.30) in the Czech sample. For same-sex listeners, the equivalent participant numbers were as follows: 25 (24.4 ± 2.93), 32 (24.4 ± 2.95), 20 (23.2 ± 4.88), and 21 (24.6 ± 6.13), respectively. Informed consent was obtained from all subjects.

For the test using the same original voice recordings rated for friendliness, 131 heterosexual participants were recruited. Here, listeners were presented with both same- and opposite-sex recordings (with order fully randomised). The final sample included 108 participants: 23 men (mean age ± SD = 32.7 ± 11.78) and 44 women (30.4 ± 14.79) in the Czech sample, and 15 men (33.3 ± 9.38) and 26 women (28.2 ± 10.17) in the English sample.

For the tests assessing low-pass filtered voice recordings, 174 heterosexual participants were recruited. Again, listeners were presented with both same- and opposite-sex recordings, and because filtering renders speech unintelligible, I relaxed selection for participants who understood a little of the other language. The final sample included 82 participants judging the recordings in terms of attractiveness, and 92 judging on friendliness. For attractiveness judgements there were 22 men (mean age ± SD = 25.6 ± 3.16) and 21 women (24.8 ± 4.18) in the English sample, and 11 men (25.9 ± 5.89) and 28 women (24.8 ± 6.11) in the Czech sample. Equivalent participant numbers judging friendliness were 25 (26.8 ± 6.70), 30 (25.9 ± 4.12), 20 (24.6 ± 5.33), and 17 (23.0 ± 4.62), respectively.

Additional details regarding the exclusion criteria for these tests are provided in Appendix A.

2.3.1.2 *Audio Samples*

I used the recorded responses of the first 10 tested participants from each sex/language combination to the most attractive and the most unattractive females as, in the voice recordings, there was significant variation in F_0 SD for both male and female participants (notice that all participants were told their recordings would be presented to opposite-sex participants to be judged as a potential date). Separate tests were also composed using responses subjected to low-pass filtering (Burnham et al., 2002) using Praat$^{©}$ 5.2 with an upper cut-off of 400 Hz (i.e. removing all frequencies above the cut-off level), and standardised to approximately 9 seconds in length (mean ± SD = 8.98 ± 2.28). Low-pass filtering retains variation in fundamental frequency in the voice samples, including minimum F_0, but removes all spectral information above the cut-off point (including most formants) and renders speech unintelligible. Additional details are provided in Appendix A.

2.3.1.3 *Experimental Procedure*

Rating tests were conducted online and presented to participants in their native language. Participants were presented with each pair of recordings of the opposite linguistic group, in a different randomised order for each listener. Within each pair, the same voice was directed towards an attractive and an unattractive individual. For the original voices, the research was described as a study of vocal preferences in a foreign language. For the low-pass filtered voices, participants were asked to imagine that they were listening to somebody speaking in a nearby room (because filtered recordings sounded somewhat like this). In all tests, participants were asked to select the recording that sounded either more attractive (i.e. "please listen to both recordings and select the one you think sounds more attractive") or friendly (i.e. "please listen to both recordings and select the one you think sounds more friendly") from each pair.

2.3.2 *Results*

First, I compared the extent to which listeners preferred recordings directed towards an attractive target with the level expected by chance (0.5) using one-sample t-tests. In response to the original unfiltered voices, the recording directed towards attractive individuals was chosen as more attractive by opposite-sex naive listeners in every case (Fig. 2.3a): English men speaking to attractive women were preferred by Czech women more often than expected by chance ($t_9 = 15.05$, $P < 0.001$), and the same effect was found for English-speaking women and Czech-speaking men and women ($t_9 = 14.57$, $P < 0.001$; $t_9 = 20.77$, $P < 0.001$; $t_9 = 8.72$, $P < 0.01$, respectively). The same was true of judgments based on the filtered recordings (Fig. 2.3c): opposite-sex listeners preferred recordings directed towards an attractive target at levels above chance, in each language/sex combination (English men: $t_9 = 3.49$; Czech men: $t_9 = 3.64$; English women: $t_9 = 3.50$; Czech women: $t_9 = 5.21$; $P < 0.01$ in every case). I also asked independent groups of listeners to select the recording that sounded friendlier, rather than more attractive, from both the original (Fig. 2.3b) and low-pass filtered (Fig. 2.3d) recordings. In these tasks, judgments of neither opposite-sex nor same-sex listeners differed significantly from chance, except in one case, where original recordings of English speaking males were rated by Czech females ($t_9 = 3.44$; $P < 0.01$; Fig. 2.3b). However, in this one case, the strength of preference was lower than in the mate choice context.

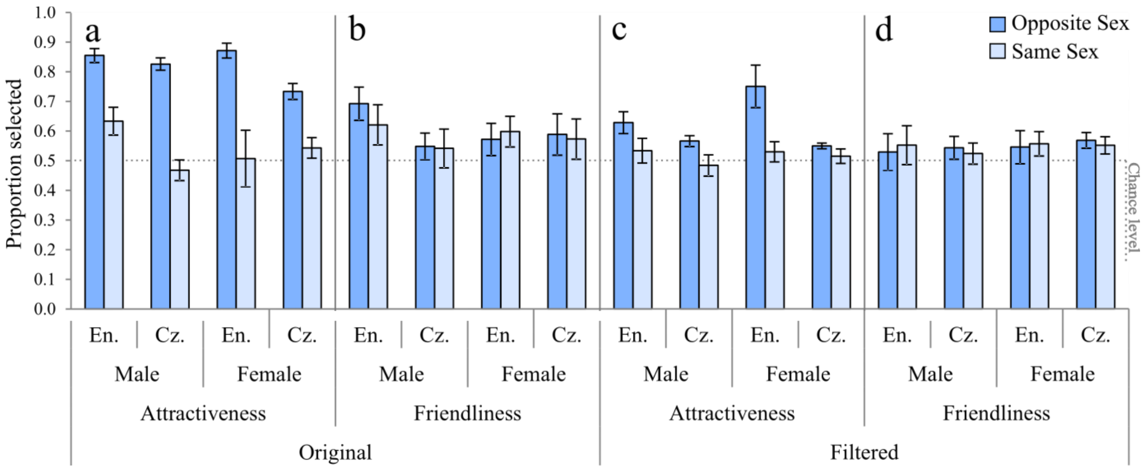


Figure 2.3. Mean proportion of recordings towards attractive targets that were selected as more attractive (a,c) or friendly (b,d) by naive listeners. (a) Original recordings, selected as more attractive; (b) original recordings, selected as friendlier; (c) low-pass filtered recordings selected as more attractive; (d) low-pass filtered recordings selected as friendlier. The horizontal axis represents the type of recordings used (original, filtered), the context (whether recordings were judged for attractiveness or friendliness), as well as the gender and language (En.: English; Cz.: Czech) of the speakers in the recordings. In every case, the recordings were rated by judges (opposite-sex: dark blue bars; same-sex: light blue bars) from the other linguistic group (i.e. English raters who do not speak Czech, and vice versa). The dotted horizontal line represents a chance level of 0.5. Bars represent mean ± 1 s.e.m. Descriptive statistics are presented in Table A4 in Appendix A.

To compare these effects directly, I used generalised linear models (GLM) (with Bonferroni-adjusted $\alpha = 0.025$ because I performed 2 analyses), with Rater Sex (same, opposite), and Context (attractiveness, friendliness), as within-subjects factors, and Language (Czech, English) and Gender (male, female) as between-subject factors. I tested whether judges preferred responses to attractive individuals depending on the context (attractiveness, friendliness), and sex of the raters (same, opposite), when presented with original, unaltered recordings. I found significant main effects of both context and rater sex (Fig. 2-3a,b) on the proportion of responses to attractive individuals selected as more attractive or friendly, such that the proportion was significantly higher when recordings were rated for attractiveness than for friendliness ($F_{1,36} = 10.27$, $p < 0.001$) and by opposite-sex compared to same-sex raters ($F_{1,36} = 63.19$, $p < 0.001$). Furthermore, there was a significant interaction between context and rater sex ($F_{1,36} = 50.93$, $p < 0.001$; Fig. 2.3a,b): in every case, recordings directed towards attractive individuals were chosen as more attractive by a higher proportion of naive opposite-sex listeners (Fig. 2.3a), but this was not the case when the recordings were rated for friendliness (Fig. 2.3b). In contrast, neither the language nor the gender of the speakers (nor the interaction between these) had a significant effect on the proportion selected ($F_{1,36} = 3.29$; $F_{1,36} = 0.47$; $F_{1,36} = 1.11$, respectively; $p > 0.05$ in all cases). For full results, see Table A5 in Appendix A.

Finally, I tested whether these preferences were preserved after stripping the voices of most acoustic information, but retaining F_0. To do this, the recordings were subjected to low-pass filtering with an upper cut-off of 400 Hz (Burnham et al., 2002). Filtered recordings were then presented in two-alternative forced-choice tasks (as with the original, unfiltered recordings) to listeners from the other language group. Again, rater sex had a significant main effect on the proportion of responses to attractive individuals selected as more attractive or friendly (Fig. 2.3c,d), with opposite-sex raters selecting a higher proportion of responses to attractive targets than same-sex raters ($F_{1,36} = 8.15$, $p < 0.01$). Moreover, and similarly to ratings of original recordings, there was a significant interaction between context and rater sex ($F_{1,36} = 5.81$, $p = 0.021$; Fig. 2.3a,b), in which recordings directed towards attractive individuals were chosen as more attractive by a higher proportion of naive opposite-sex listeners, and neither the language nor the gender of the speakers, or their interaction, had a significant effect on the proportion selected ($F_{1,36} = 3.06$; $F_{1,36} = 0.27$; $F_{1,36} = 0.15$, respectively; $p > 0.025$ in all cases). Full results are presented in Table A4 in Appendix A. The strength of preference for recordings directed towards attractive individuals was reduced in comparison to the original, unfiltered voices, indicating that other acoustic parameters also contribute to vocal judgments, but this test nonetheless suggests that modulation of F_0 is sufficient to influence proceptivity in naive opposite-sex listeners. Together, these tests indicate that listeners respond proceptively to pitch information contained within these recordings, but only within the context of mate choice.

2.4 General Discussion

Although previous results suggest that voice pitch plays a role in human courtship (Fraccaro et al., 2011; Puts et al., 2006), my cross-language experimental design provides new insights into the specific nature and mechanisms of paralinguistic modulation involved in courtship. While the two languages (English and Czech) are both European, they lie on separate branches of the Indo-European family with several millennia of largely independent development (Gray, Atkinson, & Greenhill, 2011) and are sufficiently distinct to ensure that semantic content cannot be understood by monolingual listeners. I thus believe that similarities in paralinguistic modulation, and their influence on proceptivity, provide evidence for robust context-dependent sensitivity across languages, but confirmatory studies in other languages and language families are now called for. At least within the two tested languages, however, modulation of F_0 occurred flexibly within a human courtship context in both men and women, and was sufficient to influence proceptivity towards the speaker independently of listeners' understanding of verbal content. Furthermore, the acoustic analysis revealed that variability in F_0 was especially sensitive to manipulation of social context and varied across social contexts in strikingly similar ways across languages.

Hormonal contraceptive use has been shown to affect evolutionary relevant preferences in potential partners (e.g. Feinberg, DeBruine, Jones, & Little, 2008; Puts et al., 2006; Roberts, Gosling, Carter, & Petrie, 2008), and could potentially explain why modulation in F_0 SD in female participants was apparent in responses to other women, but not to men. Future research specifically controlling for hormonal contraceptive use should be conducted to explore this possibility.

Previous studies of the influence of F_0 variability on attractiveness judgments have produced mixed results. Across individuals, higher F_0 variability has been found to be negatively associated (Hodges-Simeon et al., 2010), or not significantly associated (Riding, Lonsdale, & Brown, 2006), with attractiveness, leading Hodges-Simeon *et al.* (2010) to conclude that further study was needed to determine whether these different findings result from individual differences or contextual variation. Here, my within-subjects design leads me to suggest that, even if individual differences in F_0 variability do not robustly cue attractiveness, speakers do increase variability in F_0 during free speech towards individuals to whom they are attracted.

With the exception of some recent studies (e.g. Hodges-Simeon et al., 2010; Riding et al., 2006) it has generally been assumed that mean F_0 is the key parameter influencing listeners' perception and, specifically, that females prefer low-pitched males. The opposite effect, however, has been reported for red deer, a sexually dimorphic species in which females prefer males with higher F_0 (Reby et al., 2010), questioning the assumption of a general female preference for low-pitched males in mammals. Individual differences in mean F_0 are an important cue for mate quality and attractiveness, but my results suggest that F_0 variability (rather than mean F_0) may be the critical parameter underpinning vocal modulation in human courtship and competition over mates. Men generally tend to speak towards the lower limit of their pitch range (for information regarding human vocal range, see Honorof & Whalen, 2005; Keating & Kuo, 2012), potentially driving mean pitch upwards when they increase variability. Interestingly, men also reached a lower minimum F_0 when responding to attractive women, but minimum F_0 is not correlated with variability, unlike mean F_0. This raises the intriguing possibility that, as low-pitched vocal sounds are physiologically constrained (unlike high pitches; e.g. *falsetto*) (Lieberman & Blumstein, 1988; see also Fitch & Hauser, 1995), producing a low pitch at some point during an interaction might provide sufficient indication of physical masculinity (Puts et al., 2007) while freeing men to 'play' with their pitch, potentially providing independent cues of both mating intent and mate quality. Furthermore, because low-pitched masculine voices might be associated with aggression (Puts et al., 2012) and because masculinity is often associated with negative attributions (Hodges-Simeon et al., 2010), such modulation could potentially enable men to signal both their masculinity and lack of threat simultaneously, thereby moderating the effect of such negative attributions.

These ideas are consistent with previous suggestions that modulation of F_0 is a general mechanism to signal low or high threat in social interactions (see Hodges-Simeon et al., 2010, 2011; Puts et al., 2012). For example, increased F_0 variability has been associated with positive traits such as dynamism, femininity and aesthetic inclinations in male speakers (Addington, 1968) or simply friendliness, because adults tend to exaggerate this trait when speaking to infants (Trainor, Austin, & Desjardins, 2000). Alternatively, decreased F_0 variability occurs in competitive contexts (Hodges-Simeon et al., 2010, 2011) and is associated with higher aggressiveness in both foraging and industrial societies (Puts et al., 2012). Thus it could be argued that my results support this, more general, hypothesis - that modulation in speakers' F_0 variability might influence attractiveness assessment indirectly, by increasing perceived friendliness and low threat. However, my perceptual studies suggest this is unlikely: responses to attractive targets were preferred consistently only by opposite-sex listeners, and only when rated for attractiveness (Fig. 2.3).

Finally, such variability in paralinguistic prosody in courtship contexts has implications for ideas about the evolution of musicality in humans. One important part of musicality is the ability to process the pitch variations that produce the contours and, ultimately, a melody (Peretz & Coltheart, 2003; Peretz & Hyde, 2003). Others have argued that IDS (in which adults alter vocal characteristics such as pitch, cadence and intonation contours when speaking to infants) could be an important component in the development of musicality (Trehub, 2003), in view of characteristic patterns of vocal modulation by mothers and its detection by infants (Fernald & Kuhl, 1987), as well as the effects of IDS on infants’ arousal, focus on the mother and strengthening of mother-infant bonds.

Applying the same logic, I suggest that production of similar kinds of vocal modulation during courtship, and its detection and influence on proceptivity, could also be precursors for the development of musicality. In contrast to IDS, vocal modulation in courtship can also help to explain why music and singing is so prevalent in adulthood (Brown, 2000a; Fitch, 2006b), and why serenading is so prevalent both historically and cross-culturally. My results thus introduce a new line of support for the hypothesis of an evolutionary origin of music through sexual selection, as first suggested by Darwin (1871).

CHAPTER 3: CONTEXTUALISING COURTSHIP: ODOUR EFFECTS ON VOCAL MODULATION

3.1 Introduction

In Chapter 2, we saw how speakers subtly altered vocal parameters, particularly variation in pitch, according to experimentally manipulated courtship contexts. In this chapter, I extend this approach by testing vocal modulation in response to the presence of biologically relevant odours.

In recent years, numerous studies have shown that mere presence of odours can bring about a number of psychological changes in people in a range of different contexts. For example, ambient odours can influence people’s mood and creativity (Knasko, 1992), and ambient odours that are perceived to be more associated with one or other gender alter gender-congruent shopping behaviour (Spangenberg, Sprott, Grohmann, & Tracy, 2006). Furthermore, subliminal presence of citrus scent, an odour associated with cleanliness, can influence hygienic behaviour (Holland, Hendriks, & Aarts, 2005), and odours associated with faeces and vomit appear to trigger behaviour associated with disgust and avoidance, including more positive attitude towards safe sex (Tybur, Bryan, Magnan, & Hooper, 2011) and more conservative attitudes towards sexual behaviour (Adams, Stewart, & Blanchar, 2014).

Such effects are not limited to ambient fragrances and those associated with disease risk, but also involve bodily odours and their influence on mating behaviour. For example, subliminal presence of male axillary odour alters ratings of men’s faces by women (Thorne, Neave, Scholey, Moss, & Fink, 2002), and manipulation of men’s axillary odour by use of artificial fragrances alters their self-perceived confidence, and this influences attractiveness judgments by women even in the absence of olfactory information (Roberts et al., 2009). Furthermore, specific components of male axillary odour have been shown to have effects on mood and women’s prosocial behaviour towards men. For example, many studies focus on a group of naturally occurring steroids, the 16-androstenes, particularly the compound androstadienone. Previous researchers have demonstrated effects on women of androstadienone exposure on positive mood (Jacob & McClintock, 2000) and changes in attractiveness judgements such that presence of androstadienone led to higher attractiveness ratings (Saxton, Lyndon, Little, & Roberts, 2008).

In light of these results, and those presented in Chapter 2, I set out here to test whether presence of male axillary odour, and androstadienone in particular, would also influence vocal modulation in courtship contexts. I used the same measures of vocal parameters as used in Chapter 2 (mean and variability (SD) of voice fundamental frequency (F0), and similar measures of intensity), and the same experimental procedure, to test men’s and women’s vocal changes in the presence and absence of the odour in responses to opposite-sex targets.

3.2 Materials and Methods

3.2.1 Participants

I recruited 80 heterosexual participants who were students at the University of Stirling, half of which were men (mean age ± SD = 20.48 ± 0.41) and half women (20.50 ± 0.49). Participants were not suffering from voice hoarseness or nasal congestion at the time of testing. To ensure participants had a normal sense of smell, all participants were asked to do a brief screening test, in which they had identify 12 odorants in a multiple choice task with 4 alternatives for each odorant (the Sniffin’

Sticks^{TM} Screening 12 test, www.burghart-mt.de); only data from participants who could identify at least 9 of the smells were included in the analysis.

3.2.2 Target Videos

I used videos that were selected as stimuli for a previous study (Chapter 2). These stimuli were selected from a group of 40 videos: 20 of males (mean age ± SD = 22.5 ± 2.41) and 20 of females (22.1 ± 1.65), of 20 seconds length. Their task was presented as: "Please introduce yourself to an attractive person of the opposite sex". Each video was rated for attractiveness by 24 opposite sex persons. 12 videos were selected based on their mean attractiveness ratings: 6 of males and 6 of females (3 most attractive and the 3 least attractive in both cases).

3.2.3 Odour Stimuli

Body odour samples were collected from 12 men (mean age 21.4 ± 1.9). Each wore a cotton pad in each armpit for one night. They were instructed to wash with unperfumed soap before going to bed, to avoid spicy foods, and to place the pads into the provided sealable bags on waking. These are standard and well-used procedures for axillary odour perception studies (Havlíček, Roberts, & Flegr, 2005; Roberts et al., 2005, 2008). Each odour sample was then frozen immediately – freezing does not alter axillary odours (Lenochova, Roberts, & Havlíček, 2009; Roberts et al., 2008).

Male odours were then rated for pleasantness by a separate group of people (5 males and 5 females), who gave each pad a score out of 7 points, which ranged from -3 (very unpleasant) to 3 (very pleasant). The top four odours were pooled to create a “high quality” (HQ) male smell, while the bottom four odours formed the “low quality” (LQ) male smell. Between-individual differences in attractiveness of body odour, when averaged across a number of different raters, likely reflect a measure of absolute quality such as psychosocial dominance (Havlíček et al., 2005) or low fluctuating asymmetry (Gangestad, 2003; Rikowski & Grammer, 1999), rather than a relative measure of mate compatibility based on MHC, because the latter effect will differ between different odour donor/rater pairs. Differences in Mean ratings of pleasantness given by each rater to the composite odours in the HQ category (*M* = 0.35, *SD* = 0.57) were significantly higher than those given in the LQ category (*M* = -1.35, *SD* = 0.27) (paired-samples *t*-test: $t_9 = 10.52$, $p < 0.001$). Note also that use of composite samples (i.e. pooling odours of 4 men in each category) further avoids the potential confounding influence of differences in genetic similarity between sniffer and odour donor (see e.g. Roberts et al., 2008; Wedekind, Seebeck, Bettens, & Paepke, 1995).

Each cotton pad was shredded into little pieces, and samples for each session were created mixing equal parts from each odour of the HQ or LQ category. These final samples were frozen in individual sealable bags.

3.2.4 Experimental Procedure

Participants were randomly divided into four experimental groups corresponding to the odour they were to be exposed to (HQ, LQ), and whether Androstadienone (AND) was added to the odour (group A: HQ; group B: HQ + AND: group C: LQ; group D: LQ + AND). Each participant was asked to attend two sessions (experimental and control), between 7 and 14 days apart. Participants were exposed to the odour only in the experimental condition, and sessions were counterbalanced so that for half of the men and women in each group, the control took place in the first session, and for half in the second.

For use during each experimental session, the appropriate odour sample was removed from the freezer; at this point, when testing participants in the B and D groups, 250 µM of AND were added to the odour sample. The sample was placed in the cubicle where the experiment was going to take place 15 minutes before the session, in a small plastic container wrapped in clean aluminium foil. Odour samples were left in the cubicle during the duration of the experimental session and removed afterwards, leaving the cubicle open and empty for no less than 15 minutes before placing new odour samples to test other participants. For control sessions, clean pieces of cotton pads were placed in the same manner, so that participants could not differentiate between the control and experimental sessions.

Sessions were conducted in small, quiet testing cubicles with artificial light and no windows. During the experiment, participants were alone in the cubicle, sitting in front of a laptop, with the plastic container placed directly on the desk between the participant and the laptop, so that the odour sample was about 25 cm below the participant's nose.

The procedure from here on followed the methods described in Chapter 2. The study was presented to participants as an experiment on selection of potential mates and relationship formation, looking into the relative importance of attractiveness, self-confidence and body language on male and female preferences, as well as to understand the effect that different odours have on these psychological mechanisms. The odours used in the experiment remained undisclosed until participants were fully debriefed.

During the experiment, participants were shown the six target videos of the opposite sex, and asked to record a response message to each one of them, using a head mounted microphone. They were told that these messages would be presented to opposite-sex participants who would judge them as a potential date. Based on a study which produced demonstrable effects on mate preferences (Gangestad et al., 2004), participants were instructed to explain whether and why they would like to date the person in the video.

To avoid possible effects of pitch convergence (Gregory et al., 2001), all videos were played without sound. Participants were told that "at this stage" (to maintain the illusion that they might meet the judges) they had to base their responses only on visual characteristics of the person in the video (e.g. attractiveness, body language and clothing style). Additionally, the laptop video camera was on (but not recording) during the experiment, to create the illusion that their videos were going to be shown to opposite-sex participants. The video targets were presented electronically to participants using E-Prime 2.0 software (Psychology Software Tools, Inc., 2012; www.pstnet.com), and the order of the six opposite-sex target videos was fully randomised for each participant/session. Immediately following each video, participants were asked to rate the attractiveness of each target (on a 7-point scale), and monaural audio responses of the participants were digitally recorded using E-Prime (SoundIn object) on a laptop PC, using a ClearChat Stereo™ Headset (Logitech®, 2007), positioning the microphone about 2 cm from the participant's mouth.

Each participant did the experiment during the experimental and control conditions, recording 12 responses altogether (6 during the control, and 6 during the experimental condition). In total 957 recordings were obtained (3 recordings were discarded because of technical problems or background noise that affected audio quality and subsequent acoustic analysis). Similarly to the methods described in Chapter 2 (Section 2.2.1), each participant responded to three targets of each attractiveness category (attractive, unattractive) during both the control and experimental sessions. The values used in the analysis were, therefore, the mean of each participant's three responses on each condition/attractiveness combination: control/attractive, control/unattractive, experimental/attractive, and experimental/unattractive. By doing this, four mean values were entered for each participant (for each analysed acoustic variable) instead of 12 (the number of recordings per participant), avoiding potential pseudo-replication issues.

In addition, in the first session and before the experiment, participants were asked to read and sign the consent form, as well as take the short olfactory sensitivity test. In the second session, and after the experimental procedure, participants were to read a debriefing sheet. Their data were only kept and analysed if they still agreed after being fully debriefed. The study described in this chapter was approved by the Ethics Committee of the Department of Psychology, University of Stirling.

3.2.5 Data Analysis

Acoustical analyses of the recordings were done following the method described in Chapter 2. Using Praat Praat©, Version 5.2 (P. Boersma and D. Weenink, 2011; www.praat.org) values on intensity (dB) and F_0 (Hz) were obtained every 10 ms. A noise-resistant autocorrelation method (75 - 300 Hz for male voices, 100 - 500 Hz for female voices) was used. For intensity, only time points for which the Praat algorithm produced a value of F_0 were used.

Means and standard deviations were then obtained for intensity and F_0, attractiveness ratings, as well as minimum F_0 for male voices, for each of the 957 recordings. For these values, mean scores were calculated for each participant according to the session and attractiveness of the target; because each participant responded to three targets of each attractiveness category in each session, values used in the analysis were the mean of their three responses to attractive and unattractive targets during the control and experimental sessions. These were analysed using repeated-measures generalised linear models (GLM) for each parameter, with Bonferroni-adjusted $\alpha = 0.0083$ (because I performed 6 analyses), using sex of the participant, AND (added, not added) and body odour (HQ, LQ) as between-subjects factors, and condition (control, experimental), and target attractiveness (attractive, unattractive) as within-subjects factors. Post-hoc pairwise comparisons (*t*-tests) were conducted for significant effects of condition. All tests are two-tailed.

3.3 Results

To avoid the possibility that apparent differences between groups might be an artefact of between-subject differences, I tested each participant in two conditions: control (no odour stimuli), and experimental (odour stimuli). The within-subjects effects involving Condition are reported in Table 3.1, reflecting the experimental design (full models are provided in Table B1 in Appendix B). Analysis revealed that both mean F_0 and variability in F_0 (F_0 SD) were particularly sensitive to the inclusion of odour stimuli (Table 3.1). In both cases, a main effect of Condition was significant: in the experimental sessions, participants' mean F_0 was significantly higher, while F_0 SD was significantly lower; these results were driven by the differences in acoustic parameters of female participants between conditions (Fig. 3.1c,d). Unsurprisingly, the interaction between condition and participant sex was again significant for F_0 and F_0 SD: women's F_0 was significantly higher in the experimental condition, while men's F_0 was significantly lower; in contrast, F_0 SD of male participants did not change significantly between conditions, while females' F_0 SD was significantly lower in the experimental condition.

Table 3.1. Context-dependent variation in vocal parameters and attractiveness ratings

Effect	**Vocal parameter**										**Attractiveness Ratings**	
	Mean Intensity		**Intensity SD**		**Mean F_0**		**F_0 SD**		**Minimum F_0 (male only)**			
Within-subject Effect	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**
C	0.60	0.44	2.38	0.13	**279.02**	**<0.0001**	**56.26**	**<0.0001**	2.84	0.10	0.04	0.85
C * A	0.98	0.33	0.36	0.55	0.04	0.85	0.14	0.71	0.34	0.56	0.10	0.75
C * BO	0.31	0.58	1.34	0.25	0.73	0.40	0.00	0.98	0.37	0.55	0.69	0.41
C * BO * A	1.95	0.17	0.18	0.68	0.00	0.95	0.27	0.60	0.21	0.65	0.49	0.48
C * PS	0.04	0.83	2.23	0.14	**354.29**	**<0.0001**	**46.31**	**<0.0001**	.	.	1.80	0.18
C * PS * A	0.15	0.70	0.28	0.60	1.63	0.21	2.21	0.14	.	.	1.80	0.18
C * PS * BO	0.04	0.85	0.64	0.43	1.19	0.28	0.91	0.34	.	.	0.92	0.34
C * PS * BO * A	0.01	0.90	0.00	0.97	0.15	0.70	0.02	0.89	.	.	2.98	0.09
C * TA	1.20	0.28	0.85	0.36	**524.72**	**<0.0001**	**49.41**	**<0.0001**	0.47	0.50	0.01	0.93
C * TA * A	0.24	0.63	0.37	0.55	0.05	0.82	0.18	0.68	**15.37**	**<0.001**	3.13	0.08
C * TA * BO	0.70	0.40	0.98	0.33	1.08	0.30	1.13	0.29	3.56	0.07	0.22	0.64
C * TA * BO * A	0.12	0.73	1.54	0.22	0.06	0.81	0.02	0.90	0.07	0.79	0.01	0.93
C * TA * PS	0.88	0.35	0.00	0.96	**537.52**	**<0.0001**	**65.53**	**<0.0001**	.	.	0.70	0.40
C * TA * PS * A	0.03	0.86	0.41	0.53	0.29	0.59	0.00	0.98	.	.	0.22	0.64
C * TA * PS * BO	3.00	0.09	0.05	0.83	0.82	0.37	0.30	0.58	.	.	0.43	0.52
C * TA * PS * BO * A	3.98	0.05	0.47	0.50	1.12	0.29	0.89	0.35	.	.	0.70	0.40

C = Condition (control, experimental), A= Androstadienone (yes, no), BO = Body Odour (HQ, LQ), PS = Participant Sex (male, female), TA = Target Attractiveness (attractive, unattractive). Results are from repeatedmeasures generalized linear models (d.f. = 1, 72 in each case) for each vocal parameter, with Bonferroni adjustment or multiple tests ($\alpha = 0.0083$). Significant effects are in bold.

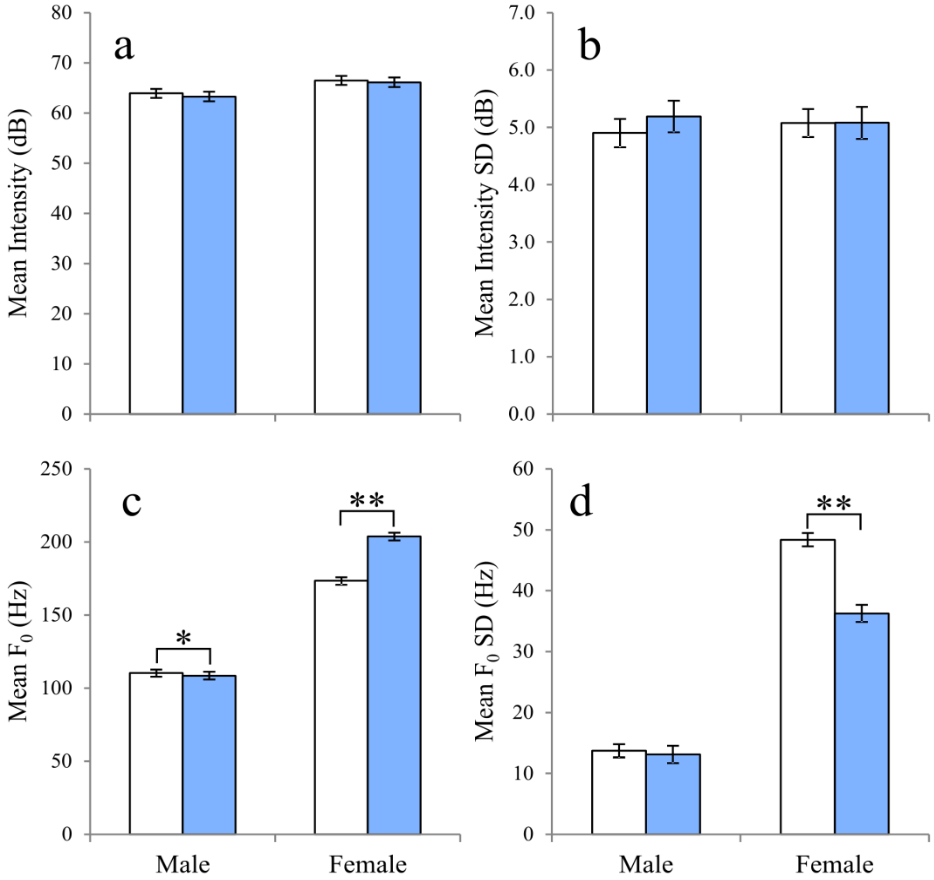


Figure 3.1. Modulation of acoustic parameters in speech towards opposite-sex targets, split by condition (control: white bars; experimental: blue bars). (a) Mean Intensity; (b) Intensity SD; (c) Mean F_0; (d) F_0 SD. Standard deviation (SD) for intensity and F_0 were used as a measure of variability. Bars represent mean ± 1 s.e.m. Post-hoc tests, *$p < 0.01$, **$p < 0.0001$.

Similarly, F_0 and F_0 SD were sensitive to the interaction between condition and target attractiveness (Table 3.1), but these differences were driven by female participants (Fig. 3.2), so that the interaction between condition, target attractiveness and participant sex was significant: while in the control condition female F_0 was lower (Fig. 3.2a) and F_0 SD was higher (Fig. 3.2b) when responding to unattractive targets, these differences were non-significant in the experimental condition (i.e. in the presence of odour stimuli), and no such pattern was found for male participants.

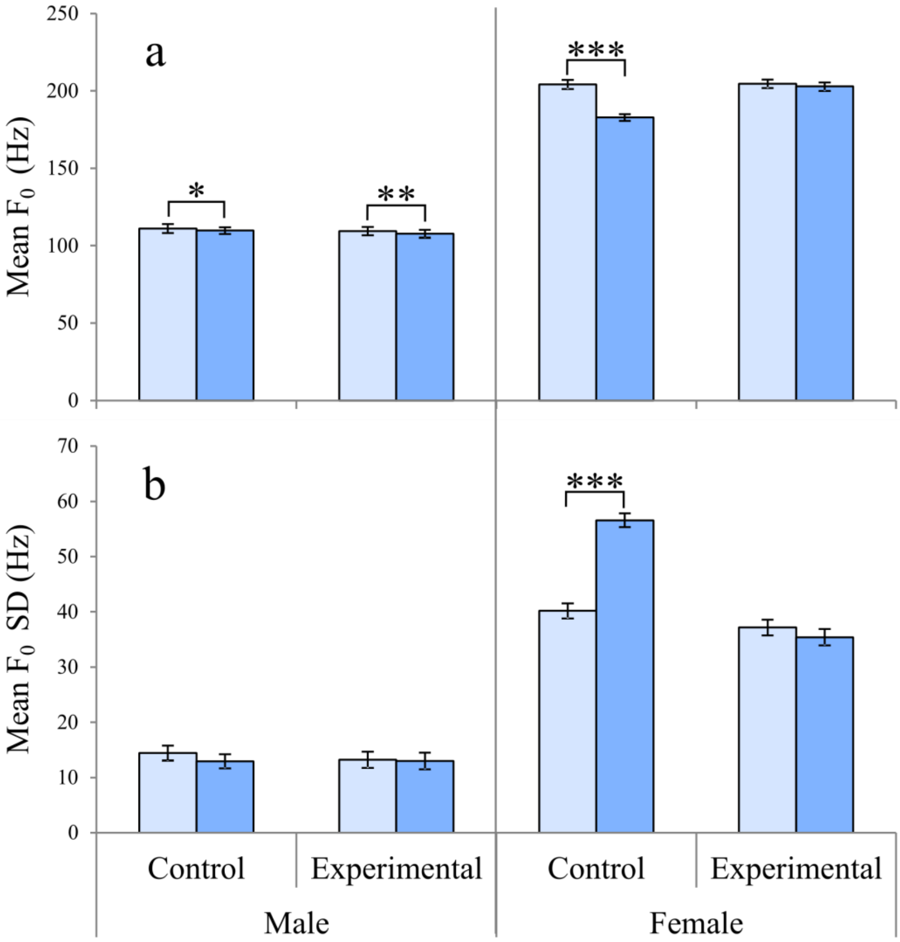


Figure 3.2. Modulation of F0 and F0 SD parameters in speech towards opposite-sex targets, split by participant sex (male, female), condition (control, experimental), and target attractiveness (attractive: light blue bars; unattractive: dark blue bars). (a) Mean F_0; (b) F_0 SD. Standard deviation (SD) for F_0 was used as a measure of variability. Bars represent mean ± 1 s.e.m. Within subjects post-hoc tests, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

No significant interactions involving body odour or AND were found for F_0 or F_0 SD, suggesting that the inclusion of *any* odour stimulus was sufficient to elicit changes in these acoustic parameters (particularly in female participants), but that the specific qualities of the odour samples (HQ or LQ, with or without AND) were not.

In contrast to the effects and interactions involving condition on F_0 and F_0 SD, no significant results were found for intensity and intensity SD, or the attractiveness ratings given to the target stimuli by the participants. However, the interaction between condition, target attractiveness and AND had a significant effect on the minimum F_0 of male participants (Table 3.1): while in the control condition male participants had a significantly lower minimum F_0 when responding to attractive female targets, this difference was no longer significant when participants were exposed to odour stimuli containing AND, and was reversed (i.e. participants had a significantly lower minimum F_0 when responding to unattractive targets) when they were exposed to body odour only, without AND (Fig. 3.3).

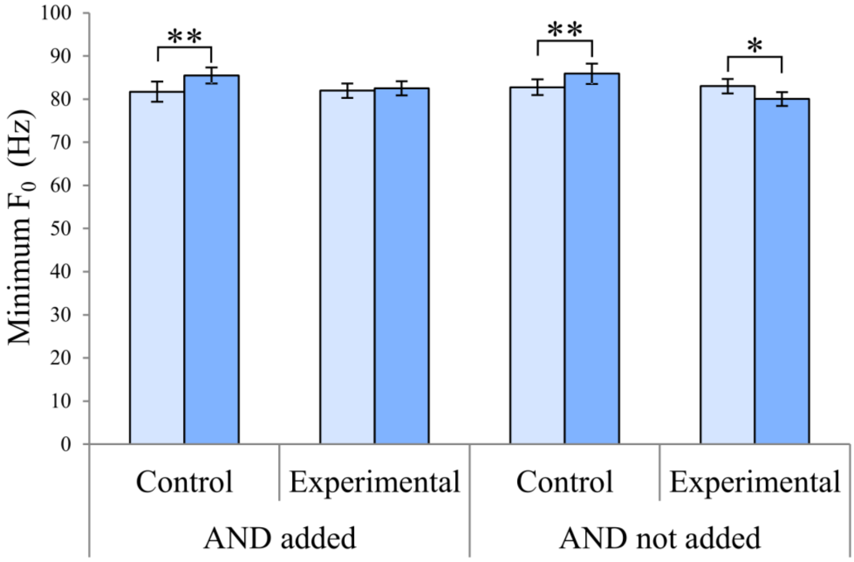


Figure 3.3. Modulation of minimum F_0 of male participants in speech towards opposite-sex targets, split by AND (added, not added), condition (control, experimental), and target attractiveness (attractive: light blue bars; unattractive: dark blue bars). Bars represent mean ± 1 s.e.m. Post-hoc tests, $*p < 0.05$, $**p < 0.01$.

3.4 Discussion

Results from this chapter corroborate the main effects described in Chapter 2 (Study 1), for vocal modulation during courtship in responses to opposite sex individuals. Specifically, these results show that when responding to attractive targets individuals tend to speak with higher variability in $F_{0,}$ higher mean F_0, and men tend to speak reaching a lower minimum F_0. The implications of these results are discussed in section 2.4.

Furthermore, as expected, the addition of male odour (any odour) produced some changes in vocal parameters, in particular decreasing male F_0 and increasing female F_0. These sex-specific changes suggest that the presence of the male odour leads to vocal changes that increase perceived femininity in female voices, and perceived masculinity/dominance in male voices, which is in line with previous research.

Men's perceived attractiveness is increased in women by the presence of male axillary secretions (Thorne et al., 2002), as well as by exposure to androstadienone (Saxton et al., 2008). Because of this, men portrayed in the target videos may have been regularly perceived as more attractive during the experimental condition in comparison to the control condition, leading women to speak with increased voice $F_{0,}$ which is attractive to men (e.g. Feinberg, Jones, DeBruine, et al., 2005; Jones, Feinberg, DeBruine, Little, & Vukovic, 2008). Men, on the other hand, showed a tendency to speak with a lower mean F_0 during the experimental condition, when they were exposed to male odour. This could be because of increased self-perceived confidence (see Roberts et al., 2009), and manifested in voices that emphasise masculinity and dominance (see e.g. Puts et al., 2007; Wolff & Puts, 2010).

The effects of high quality odour, or added androstadienone, however, did not have additional effects. This was somewhat unexpected, but it may be that the presence of the odour alone is sufficient to generate change in modulation, or that the variability in odour quality was not sufficient to elicit quality-related changes in modulation.

Perhaps if I used a larger sample of odour donors, accentuating differences between high and low quality, the effect of quality might have been measurable. With respect to added androstadienone, other constituents of the axillary odour could have a more prominent role in odour evaluation, or these other constituents may be more perceivable in the odour mixture.

What the results demonstrate is that increasing the ecological validity of the environment, providing not just a context and the images of the putative target listeners, but also an associated odour, can produce measurable changes in vocal parameters. This lends support to the idea that our use of voice is extremely context-sensitive, and so can play an important part in shaping how we are perceived by others. In the following chapter, I go on to explore how perception of such changes might reap adaptive benefits.

CHAPTER 4:

PITCH DISCRIMINATION AND LIFE HISTORY

4.1 Introduction

Results from Chapters 2 and 3 show that both men and women modulate their voices in courtship scenarios, and that these modulations, although subtle, can be detected. In particular, F_0 variability seems to play an important role in signalling courtship intent. As discussed in Chapter 2, variability in paralinguistic prosody in courtship contexts has implications for ideas about the evolution of musicality in humans, as an important part of musicality is the ability to process the pitch variations that produce the contours and, ultimately, a melody (Peretz & Coltheart, 2003; Peretz & Hyde, 2003). Musicality is a complex phenomenon that has received relatively little attention in comparison to language, and whose origins remain largely unexplained, but results presented so far offer support (a possible strong evolutionary pressure) for a potential mechanism that could have shaped, at least partially, human musicality.

This raises a further series of important questions. Why do people modulate their voices during courtship? Why do individuals tend to find speech with higher pitch variability more attractive? And perhaps more important: if vocal modulation (and particularly increased pitch variability) plays a role in courtship, does the ability to perceive these modulations allow individuals to choose more suitable partners? If the answer to the latter question is affirmative, there are several possibilities: for example, people with greater detection of vocal modulations might be better equipped to detect cues of underlying quality and relationship satisfaction potential, and/or might be more sensitive to cues that denote better parenting skills. This study aims to investigate these possibilities.

Vocal cues such as low pitch are important in signalling underlying quality. For example, testosterone, which negatively predicts voice pitch, suppresses the immune system function (Chen & Parker, 2004; Folstad & Karter, 1992) and is associated with more risky social and sexual behaviours (Archer, 2006). In hunter-gatherers, however, both men and women associate low-pitched opposite-sex voices with better skills to acquire resources (Apicella & Feinberg, 2009), suggesting that low-pitched voices can be associated with both positive and negative attributions. Increased F_0 variability has been associated in male speakers with traits like dynamism, femininity and aesthetic inclinations (Addington, 1968), as well as friendliness and playfulness (Trainor et al., 2000), but decreased F_0 variability is associated with higher aggressiveness in both foraging and industrial societies (Puts et al., 2012) and occurs in competitive contexts (Hodges-Simeon et al., 2010, 2011), suggesting that F_0 variability can be more generally interpreted as a mechanism to signal threat in social interactions (see Hodges-Simeon et al., 2010, 2011; Puts et al., 2012). Thus, acute perception of these vocal cues could have repercussions on relationship satisfaction, because the selection of complementary traits (but not necessarily similar traits; see Luo, 2009; Roberts & Little, 2008) might be associated with relationship satisfaction (Blum & Mehrabian, 1999; Robins, Caspi, & Moffitt, 2000). For example, as oral contraceptive (OC) use affects cyclic changes in partner preference as well as sexual functioning (see Roberts, Cobey, Klapilová, & Havlíček, 2013), alterations in OC use during the course of a relationship have been shown to have a negative effect on relationship satisfaction (Roberts et al., 2014).

Then again, if the ability to perceive vocal modulations during courtship allows individuals to choose *better* partners, a second option is the intriguing possibility that pitch variability could be an indicator of parenting behaviour and parental investment. For example, the ability to produce *good* infant-directed speed (IDS), which has important effects on infants' arousal, focus on the speaker, strengthening of parent-infant bonds and, potentially, language acquisition (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 2009), could be predicted by such modulations. If this is true, it could help to elucidate the evolutionary pressures that shaped human musicality. Furthermore, and because IDS has been argued to be an important component in the development of musicality (Trehub, 2003), it could have implications for our understanding of the complicated relation between musicality, language and IDS, as well as their evolutionary origins, and could explain why music and language are indistinguishable during the early developmental stages (Chen-Hafteck, 1997; see also Trehub & Trainor, 1993).

Based on the findings of chapter 2, I tested whether pitch discrimination (i.e. the ability to detect small pitch changes), which would allow individuals to detect subtle vocal modulations, is associated with the *quality* of the partners that people choose. To do this, I developed a pitch discrimination test that can be completed online, and collected information about relationship satisfaction, partner attractiveness, and partner's parental investment of the participants; these data were then used to test whether individuals with better pitch discrimination skills would tend to be more satisfied with their partners, which could be reflected as a positive association between pitch discrimination and relationship satisfaction, pitch discrimination and partner attractiveness, and/or pitch discrimination and their partners' parental investment (for those who have children). Additionally, I predicted that individuals with better pitch discrimination skills would have a stronger preference for vocal responses to attractive individuals, over responses to unattractive individuals (based on results and methods presented in study 2, Chapter 2).

4.2 Materials and Methods

4.2.1 Participants

333 participants (268 women, mean age ± SD = 23.7 ± 9.78; 65 men, 24.6 ± 10.41) recruited using social media and the University of Stirling's PsychWeb participant portal took part in the study. I excluded 84 participants because they did not complete the pitch discrimination test (see section 4.2.2 below), and a further 38 because they self-identified as either homosexual or bisexual. The final sample included 211 participants (165 women, mean age ± SD = 22.5 ± 7.94; 46 men 23.0 ± 8.85). All participants accepted an online consent form.

Of the 165 women in the final sample 74 were users of oral contraceptives (OCs) (mean age ± SD = 20.6 ± 4.99), and 91 were nonusers (24.0 ± 9.46), of which 5 had onset of their last menstruation more than 40 days before completing the questionnaire, and were therefore not considered as *regularly cycling* and excluded from analyses regarding menstrual cycle. Of the total final sample, 16 women (mean age ± SD = 41.3 ± 8.85) and 6 men (39.7 ± 10.60) had children.

4.2.2 Pitch Discrimination Test

To assess pitch discrimination skills, I developed a test in which participants listened to sequences of three beeps, one of which (first, second, or third) was slightly lower than the other two. Participants were then asked to identify the beep that was lower than the other two.

All unmanipulated beeps had a fundamental frequency of 333.34 Hz which is well within the normal hearing range for all age groups (Brant, 1990), at a constant intensity of 88.48 dB (μE), and a duration of 0.250 ms. Using Praat© 5.2.44 (P. Boersma and D. Weenink , 2011; www.praat.org), 10 altered versions of this beep were created, by lowering the original beep by an amount (in semitones) that ranged from easily perceptible, to very subtle: 0.8, 0.6, 0.5, 0.4, 0.3, 0.2, 0.15, 0.1, 0.075, and 0.05 semitones (for frequency values, see Fig. 4.1). Manipulations were performed using semitones to control for perceptual changes (see chapter 6 for information about psychoacoustic scales). Three sequences of three beeps (two originals, one lowered) were created using each one of the 10 lowered beeps (placing it first, second, or third), each one lasting 2.250 ms including a 500 ms silence after each beep. In total, 30 sequences were created.

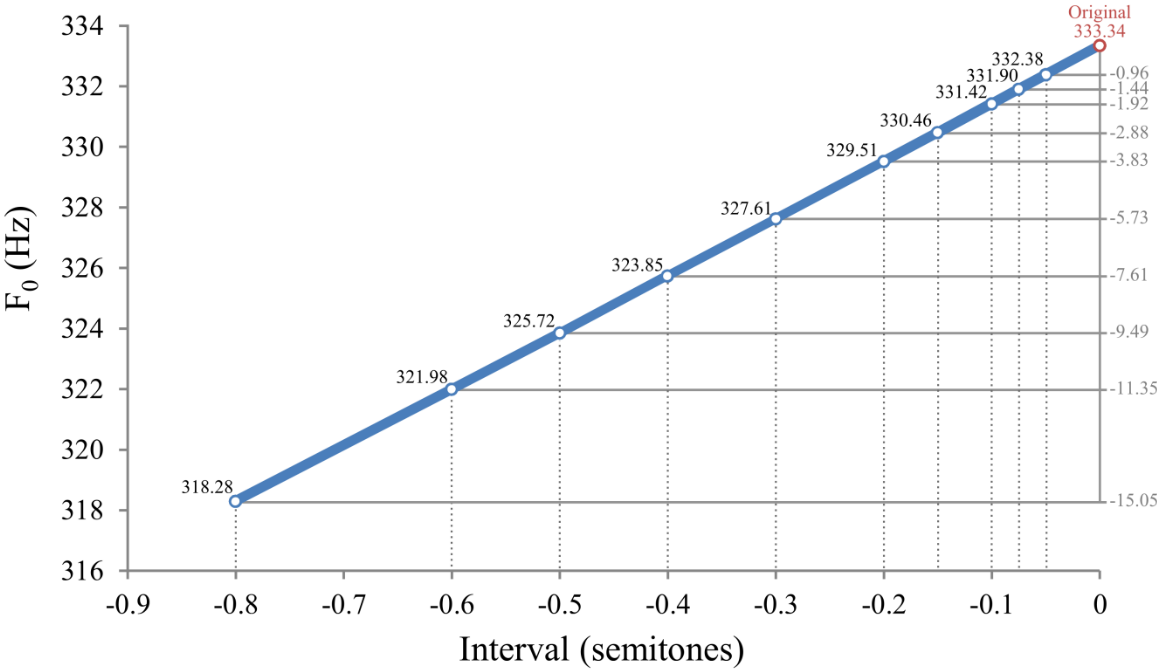


Figure 4.1. Fundamental frequency and frequency difference for each lowered beep in the pitch discrimination test. White dots represent each lowered beep. Values in black represent F_0 for each beep, and values in grey represent the frequency difference to the original beep (in red).

For the pitch discrimination test, these 30 sequences were presented in fully randomised order, in a different order for each participant, and the number of correct responses was counted. The number of correct responses ranged from 9 to 22 (mean ± SD = 17.49 ± 2.54) out of a possible maximum of 30.

4.2.3 Experimental Procedure

The study was distributed online to participants, using Qualtrics software (Qualtrics, Provo, UT, 2013; www.qualtrics.com). The main study consisted of four central sections: (1) demographic and personal information, (2) pitch discrimination test, (3) voice preference test, and (4) relationship and satisfaction and partner’s investment in children.

Demographic information included age, sex, sexual orientation, relationship status and length of the relationship (if currently in a relationship), number of children (if any), and self-rated attractiveness (in a 1-100 scale). Women were additionally asked to state whether they were currently using hormonal contraception (OC), and nonusers were asked to indicate the number of days since the onset of their last menstruation.

For the pitch discrimination test (see section 4.2.2 above), participants were presented with 30 sequences of three beeps, one of which was slightly lower than the other two, and were asked to identify which of the three beeps was the lowered one for each sequence.

For the voice preference test, following the procedure used in Chapter 2 (Study 2, Section 2.3), participants were presented with 10 pairs of short audio recordings (in total) of the same person speaking in a foreign language (Czech), and were asked to select the one they thought sounded more attractive from each pair, in a forced choice test. The audio recordings were not manipulated, and contained responses of native Czech individuals. Each pair had recordings of one individual speaking in Czech to one attractive and one unattractive target. Participants were presented with opposite-sex recordings only.

To measure relationship satisfaction, participants were asked to complete the 7-item Relationship Assessment Scale (RAS) (Hendrick, 1988), and to rate the physical attractiveness of their partners on a 1-100 scale (Appendix C). This was done taking into consideration the relationship status of the participants: those who were currently in a relationship were asked to respond in relation to their current partner (women: n =

86, men: n = 16), and those who were not currently in a relationship were asked to respond in relation to their last partner (women: n = 79, men: n = 30). Participants who had children were additionally asked to complete a 11-item questionnaire about the investment in children of the other parent of their only/youngest child (Appendix C); this questionnaire contained adapted versions of the 7-item *Delight* construct of the Parental Investment in Children (PIC) questionnaire (Bradley & Whiteside-Mansell, 1997), as well as 4 additional items designed to further measure parental investment: “I believe my partner gives our child a lot of attention”, “my partner spends a lot of time with our child”, “my partner is/was/will be involved with our child’s schoolwork”, and “my partner is the best possible parent”. The first three of the last four items were adapted from Apicella & Marlowe (2004). Participants were also asked to specify whether or not they were still in a romantic relationship with the parent of their only/youngest child.

4.2.4 Data Analysis

Scores were calculated for each participant on the pitch discrimination test, as well as for voice preference, partner satisfaction and partner’s parental investment tests. *Pitch discrimination* was not normal, with skewness of -0.72 (SE = 0.17), and the same was true for the *days since the onset of the last menstruation* (skewness ± SE = 7.36 ± 0.25), the *duration of the current relationship* (2.90 ± 0.24), *partner satisfaction* (-0.57 ± 0.17), and *partner attractiveness* (-1.24 ± 0.17). *Self-rated attractiveness* (-0.40 ± 0.17), *voice preference* (-0.09 ± 0.17), and *partner’s parental investment* (-0.24 ± 0.49) were not significantly skewed.

To correct the skewness, square root, logarithmic and inverse transformations were attempted. However, only two variables could be corrected, in both cases using logarithmic transformations: *days since the onset of the last menstruation* (skewness ± SE = 0.57 ± 0.25) and *relationship duration* (0.16 ± 0.24). The remaining variables that were significantly skewed (i.e. pitch discrimination, *partner satisfaction*, and *partner attractiveness*), were converted into Z scores for any subsequent analyses.

4.3 Results

First, I tested relations between pitch discrimination and other variables (days since the onset of last menstruation, relationship duration, self-rated attractiveness, preference for responses to attractive targets, partner satisfaction, partner attractiveness, partner's parental investment, age; Table 4.1).

Table 4.1. Correlations between all collected variables.

	RD		**SFA**		**PD**		**VP**		**PS**		**PA**		**PPI**		**Age**	
	r	n	*r*	n	*r*	n	*r*	n	*r*	n	*r*	n	*r*	n	*r*	n
DMC	**0.27***	**38**	-0.17	91	**-0.32*****	**91**	-0.04	91	**-0.18***	**91**	**-0.31*****	**91**	**-0.66*****	**15**	**0.23****	**91**
RD	-	-	-0.14	101	0.02	101	-0.11	101	**-0.20***	**101**	**-0.20****	**101**	0.25	20	**0.67******	**101**
SRA	-	-	-	-	**0.14***	**211**	-0.08	211	-0.02	211	0.04	211	-0.07	22	0.01	211
PD	-	-	-	-	-	-	-0.05	211	-0.08	211	-0.1	211	**0.36***	**22**	0.02	211
VP	-	-	-	-	-	-	-	-	0.03	211	-0.03	211	-0.23	22	-0.05	211
PS	-	-	-	-	-	-	-	-	-	-	**0.58******	**211**	**0.56*****	**22**	-0.1	211
PA	-	-	-	-	-	-	-	-	-	-	-	-	**0.66*****	**22**	**-0.20*****	**211**
PPI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**-0.54*****	**22**

DMC = Days of menstrual cycle, RD = Relationship duration, SRA = Self-rated attractiveness, PD = Pitch discrimination, VP = Voice preference, PS = Partner satisfaction, PA = Partner attractiveness, PPI = partner's parental investment. Significant effects and non-significant trends are in bold.
*p < 0.1; **p < 0.05, ***p < 0.01; ****p < 0.001.

I expected pitch discrimination (PD) to predict relationship satisfaction (RS), partner's investment in children (PPI), and partner attractiveness (PA), but these correlations were not significant. However, because these relations could be obscured by third variables (e.g. relationship status, sex), I further explored these relations using partial correlations, as well as GLMs in the cases where it was appropriate to control for other variables.

Pitch discrimination was, however, significantly correlated to days since the onset of last menstruation (DMC) and self-rated attractiveness (SRA). For this reason, I also explored these relations.

4.3.1 Pitch discrimination, relationship satisfaction and partner attractiveness

Pitch discrimination was not significantly correlated with relationship satisfaction for men (r = -0.45, n = 46, p = 0.76) nor women (r = -0.82, n = 165, p = 0.29) (all participants: r = -0.83, n = 211, p = 0.23). This result combines data from participants who were currently in a relationship, and hence were asked to respond in relation to their current partner, and those who were not, and were asked to respond in relation to their previous partner.

Because relationship satisfaction was significantly higher for those participants who were currently in a relationship than to those who were not ($F_{1,207}$ = 68.39, p < 0.001), but it did not significantly differ between men and women ($F_{1,207}$ = 1.09, p = 0.30), nor by the interaction between sex and relationship status ($F_{1,207}$ = 0.001, p = 0.98), I performed a partial correlation for all participants between pitch discrimination and relationship satisfaction, controlling for relationship status. This further showed that pitch discrimination and relationship satisfaction were not significantly associated (r = -0.39, df = 208, p = 0.58).

The correlation between pitch discrimination and partner attractiveness was not significant (r = -0.10, n = 211, p = 0.13). Perhaps surprisingly, self-rated attractiveness and partner attractiveness were not significantly correlated (r = 0.04, n = 211, p = 0.517), even when controlling for relationship status (and therefore whether they were responding in relation to their current or previous partner) (r = 0.05, df = 208, p = 0.21).

4.3.2 Pitch discrimination and partner's parental investment

The correlation between partner's investment in children and pitch discrimination showed a non-significant trend for participants with better pitch discrimination, to rate their partners' parental investment in their children higher (r = 0.36, n = 22, p = 0.098).

Similarly to the case of relationship satisfaction, partners' investment in their children was significantly higher for those participants who were currently in a relationship with the parent of their youngest/only child than for those who were not ($F_{1,19}$ = 33.44, $p < 0.001$), but neither sex ($F_{1,207}$ = 1.72, p = 0.26), nor the interaction between sex and whether participants were currently in a relationship with the parent of their youngest/only child ($F_{1,207}$ = 0.16, p = 0.70), were significant.

When controlling for whether participants were currently in a relationship with the parent of their youngest/only child, the correlation between partner's investment in children and pitch discrimination was significant (r = 0.47, df = 19, p = 0.03). To further examine this relationship, I ran independent correlations using the data of those participants who were currently in a relationship with the parent of their youngest/only child, and those who were not. This showed that the significant positive association between pitch discrimination and partner's investment in children, was marginally significant for those who were still in such a relationship ($r = 0.48$, n = 17, $p = 0.049$), but did not reach significance for those who were not ($r = 0.40$, n = 5, $p = 0.50$).

4.3.3 Pitch discrimination and self-rated attractiveness

Self-rated attractiveness positively correlated with pitch discrimination (Table 4.1), indicating that individuals who rated themselves as more attractive tended to also score higher in the pitch discrimination test. This correlation, was driven by women ($r = 0.16$, n = 165, $p = 0.036$), and was not significant for men ($r = 0.03$, n = 46, $p = 0.83$).

4.3.4 Pitch discrimination and menstrual cycle

Pitch discrimination did not reliably predict relationship satisfaction or partner attractiveness, and although it seems to predict partner's parental investment, this result is only based on a small sample size. Unexpectedly, however, I found a significant negative correlation between pitch discrimination and days since the onset of the last menstruation (Table 4.1). To further investigate this relationship and avoid potential biases, I looked at the data of regularly cycling women (i.e. OC nonusers, whose last menstruation onset was no more than 40 days before they took part in the study). The analysis of this relationship showed that there was a negative correlation between days since the onset of the last menstruation and pitch discrimination scores ($r = 0.32$, n =

86, $p = 0.003$), but also that this relation is better expressed as a quadratic relation ($F_{2,83} = 6.07$, $p = 0.003$), in which pitch discrimination slightly increases towards the middle of the menstrual cycle, but decreases sharply in the last days of the menstrual cycle ($\Delta R^2 = 0.026$) (Fig. 4-2).

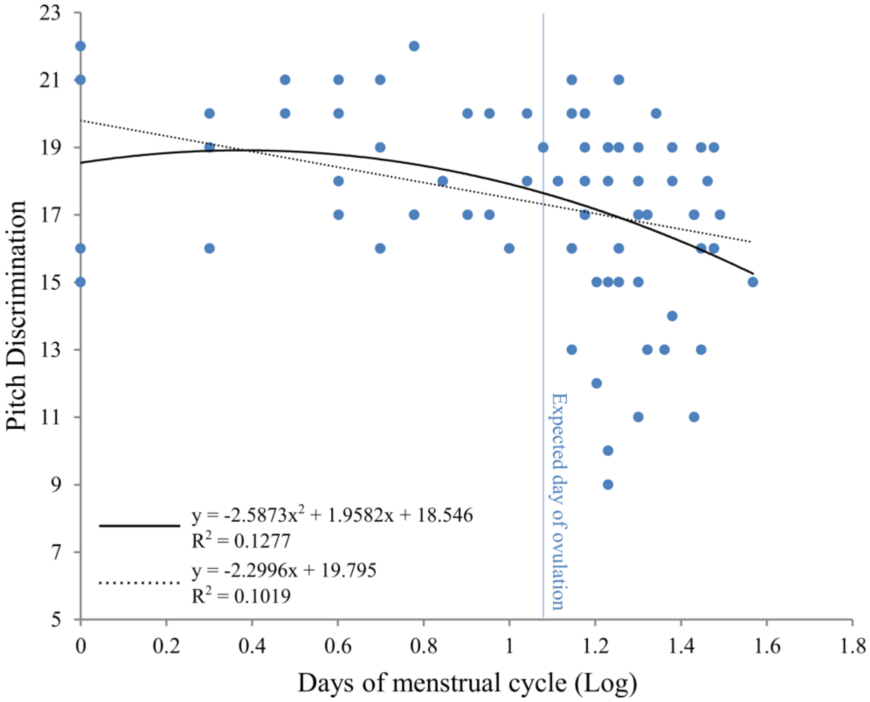


Figure 4.2. Correlation between days of the menstrual cycle (days since the onset of the last menstruation) and pitch discrimination. Black lines represent the relation (solid = quadratic; dotted = linear) between these parameters. The blue line represents day 14 of the menstrual cycle (expected day of ovulation).

To test whether there was a tendency for pitch discrimination to increase according to conception risk, I explored differences in mean pitch discrimination scores of women who were in high conception risk (days 9 – 14), to those in low conception risk (all other days of the menstrual cycle), and OC users. To do this, I conducted a one-way ANOVA, with planned contrasts (simple); this analysis showed that pitch discrimination was significantly different between women in high conception risk, women in low conception risk, and OC users ($F_{2,155}$ = 3.685, *p* = 0.027; Fig. 4.3). Furthermore, planned contrasts revealed that women in the high conception risk group tended to have better pitch discrimination than those on the low conception risk group (F = -1.658, *p* = 0.008) and those who were using oral contraceptives (F = -1.038, *p* = 0.088), although this last difference did not reach significance.

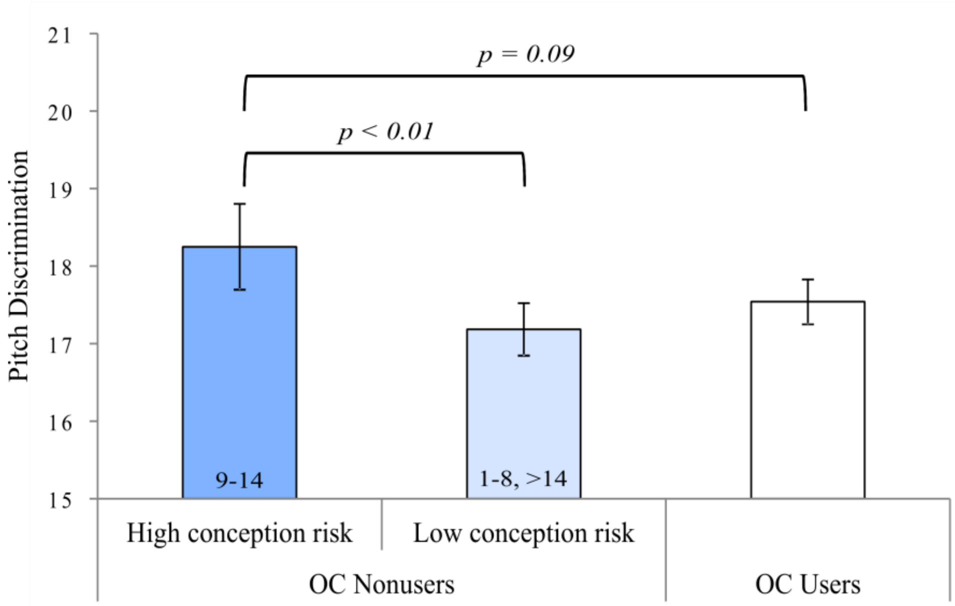


Figure 4.3. Participants' mean pitch discrimination for OC nonusers (high conception risk: dark blue bar; low conception risk: light blue bar) and OC users (white bar). For OC nonusers, the days of the menstrual cycle included in each group are written at the bottom of the bars. Bars represent means ± 1 s.e.m.

To investigate a potential proximate cause for this tendency of women during high conception risk days of their menstrual cycle to have better pitch discrimination, I looked at the relation between steroid hormonal concentrations and pitch discrimination. To do this, we calculated expected hormonal levels based on women's cycle day, following the approach of Puts (2006). We used weighted averages of daily hormone concentrations from Puts (2006), who calculated hormonal levels using data from several previous studies (Abraham, 1974; Cooke, Lenton, Adams, & Sobowale, 1978; Franchimont et al., 1976; Judd & Yen, 1972; Midgley & Jaffe, 1968). Figure 4.4 shows data for estradiol and progesterone from Puts (2006) (Fig. 4.4a), as well as the average ratio between these two hormones (Fig. 4.4b).

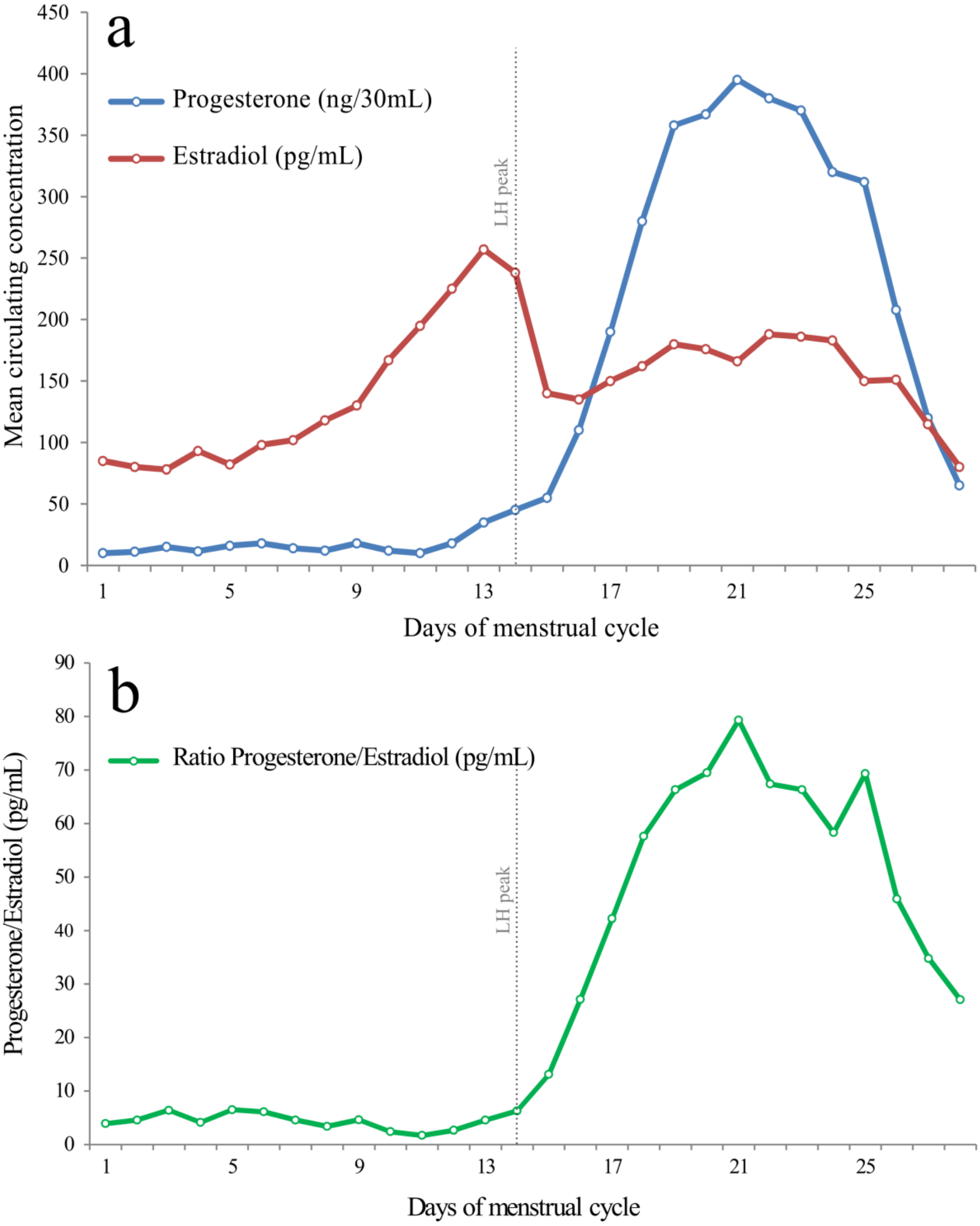


Figure 4.4. Average daily steroid hormone concentrations (estradiol and progesterone) during the menstrual cycle. (a) mean concentration of progesterone (blue) and estradiol (red), obtained from Puts (2006). (b) progesterone to estradiol ratio. To calculate the ratio, progesterone reference values (originally in ng/30mL) were converted into pg/mL.

As in the case of the relation between pitch discrimination and days since the onset of the last menstruation, there was a negative correlation between expected progesterone/estradiol ratio and pitch discrimination ($r = -0.33$, n = 86, $p = 0.002$), but this relation was also better expressed as a quadratic relation ($F_{2,83} = 7.28$, $p = 0.001$), in which, as the progesterone to estradiol ratio increases to around 40, pitch discrimination decreases; however, when the progesterone/estradiol ratio is larger than around 40, pitch discrimination increases ($\Delta R^2 = 0.043$) (Fig. 4.5).

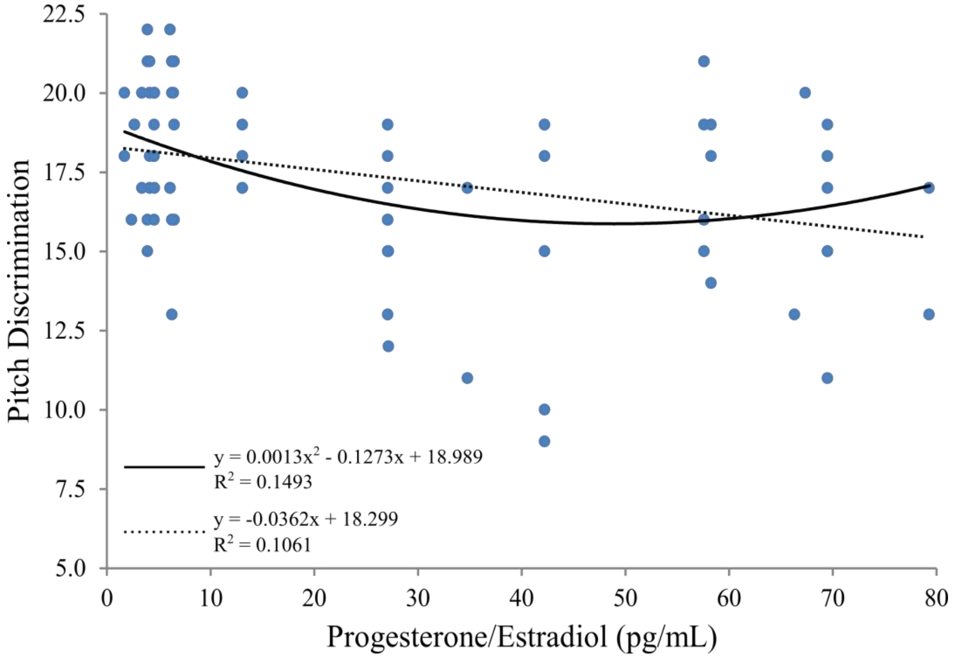


Figure 4.5. Correlation between average progesterone/estradiol ratio and pitch discrimination. Black lines represent the relation (solid = quadratic; dotted = linear) between these parameters.

Given that the average ratio between progesterone and estradiol during the menstrual cycle significantly predicted pitch discrimination, I analysed the relation between the progesterone to estradiol ratio and voice preferences (again in OC nonusers). In addition to preference for responses to attractive individuals (voice preference), I tested preference for voices with lower mean F_0 (because the literature predicts women to prefer lower mean F_0 in men's voices), as well as for higher F_0 SD and lower minimum F_0 (because results from Chapter 2 indicated a potential preference for increased F_0 variability and lower minimum F_0). Progesterone to estradiol ratio was significantly correlated with pitch discrimination (Table 4.2), but not with voice preference, or preference for lower mean $F_{0,}$ higher F_0 SD, or lower minimum F_0 (Table 4.2). Interestingly, however, pitch discrimination was negatively correlated with preference for responses to attractive targets (voice preference), but not with preference for any F_0-related acoustic parameter (low mean F_0, high F_0 SD, or lower minimum F_0) (Table 4.2). Voice preference, on the other hand, correlated positively with preferences for lower mean F_0, and particularly with higher F_0 SD, and showed a non-significant trend to be positively associated with lower minimum F_0 (Table 4.2). Preference for lower mean F_0 was negatively associated with preference for higher F0 variability, and positively associated with preference for lower minimum F0 (Table 4.2).

Table 4.2. Correlations for progesterone/estradiol ratio, pitch discrimination and voice preference.

	Pitch discrimination	**Voice preference**	**Preference mean F_0**	**Preference F_0 SD**	**Preference minimum F_0**
Progesterone/estradiol	**-0.33*** **	0.06	-0.05	-0.06	-0.03
Pitch discrimination	-	**-0.26** **	0.02	0.11	-0.04
Voice preference	-	-	**0.27** **	**0.69**** **	**0.19* **
Preference mean F_0	-	-	-	**-0.30*** **	**0.59**** **
Preference F_0 SD	-	-	-	-	-0.11

Voice preference was calculated as the number of responses to attractive targets that were selected; preference mean F_0 as the number of responses with lower mean F_0; preference F_0 SD as the number of responses with higher F_0 SD; preference minimum F_0 as the number of responses with lower minimum F_0. Significant effects and non-significant trends are in bold. n = 86 in all cases. *$p < 0.1$; **$p < 0.05$,

4.4 Discussion

This study was designed to answer the question of whether the ability to perceive vocal modulations allow individuals to choose *better* partners. Although the results are somewhat inconclusive because the main findings were unexpected to some extent and the study was not designed to test them, they are interesting and offer important questions that should be addressed in future studies.

Pitch discrimination was not significantly correlated to relationship satisfaction, or partner attractiveness, which suggests that there is no relation between these variables; however, the association between pitch discrimination and/or partner attractiveness could be weak, and obscured by other variables. Nevertheless, I found a positive association between pitch discrimination and partner's investment in children, as predicted, which is consistent with the hypothesis that pitch variability could be an indicator of parenting behaviour and parental investment. The number of participants who had children, however, was small to obtain convincing results, particularly in men. To make any conclusions, further research should be performed with an adequate number of participants (both men and women), of different ages, and relationship lengths.

Unexpectedly, however, I found two results that open new lines of research that should be addressed in studies specifically designed for that purpose: first, pitch discrimination is positively associated to self-rated attractiveness in women, and second, pitch discrimination skills appear to change across the menstrual cycle (Fig. 4.2), being seemingly higher for women during the follicular phase of their menstrual cycle (high conception risk) than for those in menses and luteal phases (low conception risk) (Fig. 4.3).

Women tend to invest more in their children than men, which means that women with more sexual partners experience less benefits in reproductive fitness than men. Because of this, choosiness is more beneficial to women than men (Buss & Schmitt, 1993; Trivers, 1972). According to this, it would be possible to expect that more attractive women might be able to *afford* to be especially selective in regards to potential partners, in comparison to less attractive women. This could theoretically explain a higher sensitivity (or attention) to cues (including vocal cues) of underlying quality, relationship satisfaction potential, and/or cues that denote better parenting skills. This hypothesis, however, must be directly addressed, controlling not only for self-rated attractiveness, but also for attractiveness ratings given by an independent panel of raters. In addition, testing sensitivity not only to vocal, but also to visual and odour, cues, could provide a broader picture of the potential effect that individual quality has on mate choice.

Similarly, the relationship between the menstrual cycle phase and pitch discrimination must be addressed in a more controlled experiment: a within-subjects experiment testing changes in pitch discrimination across the menstrual cycle, for both OC users and nonusers, taking into account individual cycle lengths and, ideally, individual hormonal levels. The results here presented are, nonetheless, thought-provoking, particularly in the light of the progesterone to estradiol ratio, which could indicate a proximate mechanism for such changes.

Changes in preference for traits that denote masculinity across the menstrual cycle have been well studied, particularly in terms of visual cues such as facial structure (Johnston, Hagel, Franklin, Fink, & Grammer, 2001; Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999; Roney & Simmons, 2008) and face skin colour (Frost, 1994) (for a review about changes in face preferences during the menstrual cycle, see Jones et al., 2008). Preference changes, however, have also been found for body odour (Gangestad & Thornhill, 1998; Grammer, 1993; Havlíček et al., 2005; Thornhill & Gangestad, 1999) and vocal cues (Feinberg et al., 2006; Puts, 2005, 2006), as well as behavioural displays (Gangestad et al., 2004).

Moreover, not only preferences, but also sensitivity, has been shown to change across the menstrual cycle: women have been shown to be more sensitive to certain social cues like facial expressions and have increased attention to social stimuli (in comparison to non-social stimuli), during the luteal phase of the menstrual cycle (Maner & Miller, 2014). Maner & Miller (2014) suggest that changes in progesterone during the menstrual cycle could be associated with perceptual attunement for social cues; during the luteal phase women's bodies prepare for potential pregnancy, and having psychological mechanisms that would allow them to successfully recruit allies and avoid social threats could directly impact on their reproductive fitness.

My results suggest a similar mechanism: pitch discrimination skills change across the menstrual cycle according to hormonal concentrations (particularly progesterone and estradiol). If progesterone has been linked with preference for social cues (Maner & Miller, 2014), estradiol concentration has been linked to an increased preference for testosterone cues in men (Roney & Simmons, 2008). This could lead to an interaction between the effects of these two hormones, emphasising cues of prosociality or underlying quality. In fact, pitch discrimination scores in my participants were better predicted by a quadratic model based on average progesterone to estradiol ratio during the menstrual cycle ($R^2 = 0.1493$; Fig. 4.5), than by a model based on the days of the menstrual cycle ($R^2 = 0.1277$; Fig. 4.4).

While the progesterone to estradiol ratio did predict pitch discrimination, it did not significantly predict voice preference (preference for responses to attractive targets). Voice preference was associated, in naturally cycling women, with preferences for low mean F_0, low minimum F_0, and particularly with a preference for higher F_0 SD (Table 4.2), which is consistent with the results from Chapter 2.

In conclusion, although more research is needed, pitch discrimination seems to predict partners' parenting behaviour and parental investment, and in naturally cycling women, was associated with the point of the cycle when the risk of conception is high. Because F_0 variability is associated with traits such as dynamism, femininity and aesthetic inclinations (Addington, 1968), as well as friendliness (Trainor et al., 2000), it could indicate parental investment. These results raise the exciting possibility that individuals could obtain cues of potential parental investment from vocal modulations, and that women evolved a mechanism (increased sensitivity to acoustic cues, potentially modulated through hormonal changes) to assess this potential, particularly during periods of high conception risk, when it matters most.

CHAPTER 5:

PERCEIVED DIFFERENCES IN SOCIAL STATUS BETWEEN SPEAKER AND LISTENER AFFECT SPEAKER'S VOCAL CHARACTERISTICS

5.1 Introduction

In authority ranking relationships, individuals who are of high social status normally have privileges that other members of their group lack (Fiske, 1992). Examples of this type of relationship in human societies include the ranking system within the military and company organisation models (e.g. an employer is higher in social status than an employee) (Fiske, 1992). Recent research suggests that individuals can obtain high social status through one of two main ways: by using force and intimidation (dominance), or by being knowledgeable and skilful (prestige) (Cheng, Tracy, & Henrich, 2010; Henrich & Gil-White, 2001). The ways in which humans communicate their social status to others range from behaviours shared with non-human animals such as facial expressions and body postures (Tiedens & Fragale, 2003), to linguistic cues (i.e. the use of formal and informal linguistic tenses, as well as using spatial metaphors that make reference to hierarchies or imply a large personal space (Fiske, 1992; Pinker, 1997)).

In terms of non-verbal behaviour, apart from facial expressions and body postures, voice characteristics are an important means to communicate socially relevant information, including social status (Borkowska & Pawlowski, 2011; Hodges-Simeon et al., 2010; e.g. Jones, Feinberg, DeBruine, Little, & Vukovic, 2010). The acoustic qualities of the human voice, aside from linguistic elements such as syntax and semantic content, can communicate an important array of biological information about the speaker including sex, femininity, attractiveness, fertility and sexual maturity, physical strength, and body size (Bryant & Haselton, 2009; Dabbs & Mallinger, 1999; Feinberg, Jones, Little, et al., 2005; Feinberg, 2008; Feinberg, Jones, DeBruine, et al., 2005; Mulac & Giles, 1996; Pipitone & Gallup, 2008; Sell et al., 2010; Xu et al., 2013). Human voices are sexually dimorphic, with men, for example, having lower pitched voices than women. While the precise evolutionary reasons for this pronounced difference are unclear, it has been suggested that it could be a product of sexual selection (Collins, 2000), including dominance competition (Puts et al., 2006).

While no research to my knowledge has explored vocal parameters with respect to prestige, effects of dominance have been widely studied. Voices low in fundamental frequency (F_0), the parameter most closely related to voice pitch, are perceived as more dominant in both men (Hodges-Simeon et al., 2010; Jones et al., 2010) and women (Borkowska & Pawlowski, 2011) (but see Tusing & Dillard, 2000), where a significant positive correlation between F_0 and dominance judgments was found for male, but not female, speakers). Perceptions of dominance appear to be based on multiple cues: F_0, which is related to androgen levels, as well as formant dispersion (D_f), related to vocal tract length and skeletal size, affect dominance perceptions (Puts et al., 2007). The information obtained from vocal cues can also have real-world consequences. In a recent study, surgeons whose voices were rated as higher in dominance and lower in concern/anxiety, perhaps reflecting an 'arrogant' and 'lack-of-care' approach, were also more likely to have been previously sued for malpractice, even when controlling for speech content (Ambady et al., 2002).

Vocal parameters, however, are not constant, and can be modulated during social interactions. Shouting during aggressive displays is a typical example, and in humans and some non-human animals, intensity (loudness) modulations are associated with dominance (Ohala, 1982; Tusing & Dillard, 2000) and hostility (Collias, 1960; Kudo, 1987). Similar to changes in body posture that increase perceived body size, changes in vocal parameters can affect perception of the speaker. Puts et al. (Puts et al., 2006) reported that men tend to lower their voices during interactions with a competitor when they perceive themselves as physically dominant, and raise it when they believe they are not, exemplifying how elements of self-perceived social status may affect social interactions. Furthermore, taller and more dominant men are less sensitive to visual cues of dominance in other men (Watkins, Jones, & DeBruine, 2010; Watkins, Fraccaro, et al., 2010), indicating that authority relationships appear to be dependent on perception of relative, rather than absolute, social status.

To date, most studies have measured responses to voices with artificially manipulated acoustic parameters (typically F_0 and D_f) to investigate how these affect perceptions of dominance (e.g. Borkowska & Pawlowski, 2011; Puts et al., 2006, 2007; Wolff & Puts, 2010), but little is known regarding vocal modulations during interactions with dominant or prestigious individuals, particularly in free speech as opposed to phonemes or standardised sentences. Furthermore, whether people respond to these two forms of social status in similar ways remains unanswered. In my experiment, I aimed to address these questions by measuring within-subject vocal modulations, in both men and women's voices, in response to dominant, prestigious, or average (control) targets. I did this by using a simulated job interview scenario where participants were required to act as a candidate and answer three standardized interview questions (ranging from introductory to interpersonal).

I predicted that (1) participants' vocal characteristics would change based on whether they were talking to a dominant, prestigious, or average target, and (2) that these changes would also be related to the participant's own self-perceived dominance and prestige. Firstly, I predicted that those participants rating themselves as more dominant would speak more loudly (i.e. with higher intensity) than those who rated themselves as low in dominance (Tusing & Dillard, 2000), especially when speaking to high status individuals. Additionally, I expected these high dominance participants to lower their F_0 when speaking to the dominant target (Puts et al., 2006). I had no a priori predictions about how participant prestige would affect their interaction with the targets, or how men and women would differ in their interactions with the male targets. Finally, as the three interview questions differed semantically (see full description of questions in methods) I hypothesized that there might be a question effect, with the greatest variation of vocal parameters found in the most interpersonal question (question 3).

5.2 Materials and Methods

5.2.1 Participants

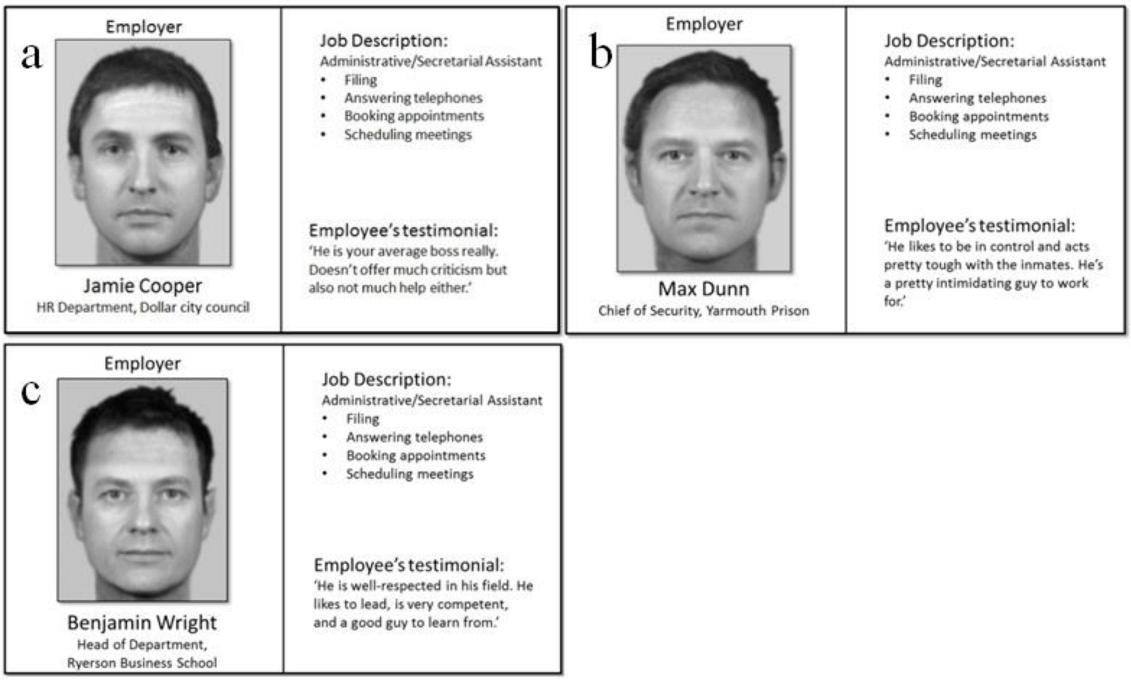
I recruited 48 participants who were students at the University of Stirling (24 men, mean age ± SD = 20.8 ± 6.56; 24 women, 20.2 ± 5.51). All participants provided informed consent and were offered course credit for their participation; the University of Stirling Psychology Ethics Committee approved the study design.

5.2.2 Target Stimuli

I used EvoFit software (Frowd, Hancock, & Carson, 2004) to create the face stimuli used in this experiment. This software allows the user to ‘evolve’ a face from sets of available faces over successive iterations, in a holistic (whole face) process as opposed to featurally (adding single features to the face one-by-one). An independent group of 32 participants (18 women; mean age ± SD = 22.7 ± 5.7) were asked to create same-sex faces using written descriptions of dominant and prestigious individuals based on definitions used in current literature (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Cheng et al., 2010; Henrich & Gil-White, 2001). Dominant individuals were described as 'An approximately 36-45 year old male/female. He/she is an extremely dominant individual. This person likes to be in control and to get their way. They will use force, coercion, and intimidation to achieve their goals if necessary.' Prestigious individuals were described as 'An approximately 36-45 year old male/female. He/she is a highly valued, prestigious and influential individual. He/she has many valued skills and qualities and others follow him/her freely. This ultimately leads to his/her achieving his/her goals.'

These 32 novel faces were rated for dominance and prestige using a 7 point scale (1 = low dominance/prestige; 7 = high dominance/prestige) by 69 undergraduate students (19 men; mean age ± SD = 29.0 ± 9.7). The two faces which received the highest dominance (mean ± SD = 5.1 ± 1.3) and highest prestige (mean ± SD = 3.99 ± 1.3) scores were used as stimuli (i.e., as the dominant and prestigious employers). For the 'average' employer, the face receiving the median rating on dominance (mean ± SD = 3.3 ± 1.3) and prestige (mean ± SD = 3.1 ± 1.3) was used.

I then created three different 'employer profiles', which contained a face image and text description, including a name, a job title, and an employee testimonial. The job title and testimonial were used to further manipulate the impression of targets as either dominant, prestigious or average (Fig. 5.1 shows the three profiles). The three profiles were also scored by an independent group of raters (see Appendix D for information on raters) for prestige and dominance, confirming that in all cases the attributes of the dominant target were rated as more dominant, the attributes of the prestigious target as more prestigious, and the attributes for the average were rated as neither high in dominance or prestige; faces were additionally rated for perceived attractiveness and age (results of these ratings are presented in Table D1 and Fig. D1 in Appendix D). Finally, job descriptions were identical (i.e. administrative/secretarial assistant including filing, answering telephones, booking appointments and scheduling meetings).



'He is your average boss really. Doesn't offer much criticism but also not much help either.'

'He likes to be in control and acts pretty tough with the inmates. He's a pretty intimidating guy to work for.'

'He is well-respected in his field. He likes to lead, is very competent, and a good guy to learn from.'

Figure 5.1. Final targets, as presented to participants, including facial images, names, job titles and employee testimonials. (a) average; (b) dominant; (c) prestigious.

5.2.3 Experimental Procedure

Participants were first told that the 'experiment' they were participating in was in fact a 'pilot' to test the effectiveness of a new interviewing technique which did not require the interviewee and interviewer to be in the same room. After written informed consent was obtained, participants were presented with the experiment using Qualtrics software (Qualtrics, Provo, UT, 2013; www.qualtrics.com), on a desktop computer located in a quiet room. Monaural audio responses of the participants were digitally recorded using Praat© 5.2.44 (P. Boersma and D. Weenink , 2011; www.praat.org), with a sampling frequency of 44.1 kHz, using a head mounted microphone positioned about 2 cm from the participant’s mouth.

To control for any potential order effects, 24 male and 24 female participants were shown the three targets in one of six possible sequences (i.e. (1) Dominant (D)-Prestigious (P)-Average (A); (2) D-A-P; (3) P-D-A; (4) P-A-D; (5) A-D-P; (6) A-P-D; the sequences were counterbalanced across participants). For each of the three targets, participants were asked to record responses to three common interview questions; hence I recorded 9 instances of speech from each participant. The interview questions were: (1) ‘please introduce yourself to this potential employer in a few sentences’, (2) ‘please tell this employer why you are a good candidate for the job’, and (3) ‘if you had a problem with a colleague at work how would you convey it to your boss?’. Aside from the generic nature of the questions, they were also selected to differ in their interpersonal characteristics. That is, while question 1 was purely a request for the subject to introduce themselves, question 2 added a personal component in requiring the participant to think about and articulate what personal attributes they believed would make them qualified for the job. Finally, question 3 had an interpersonal emphasis and required the participant to think about how they might engage with and approach the employer (target) with a problem.

After recording their responses, participants were asked to enter some basic demographic information, fill in a self-report scale of dominance and prestige (Cheng et al., 2010), rate the dominance and prestige of the three targets, and explain what they thought the purpose of the study was (see Appendix D for additional information). The entire experiment was presented using Qualtrics software, and was completed by participants while they were alone in a room. Once they had finished the experiment, participants were debriefed, given the opportunity to ask any remaining questions, and were asked to confirm whether they still consented to the use of their data.

In total, 429 recordings were obtained (3 were discarded due to background noise that affected audio quality), with length ranging from 4 to 107 seconds (mean ± SD = 25.02 ± 16.41s). Length of recording did not differ significantly depending on which target participants were responding to (repeated-measures GLM: $F_{2, 86} = 0.95$, $p = 0.39$).

5.2.4 Manipulation Check

As a final manipulation check, once the participants had completed the experiment, I asked them to rate the full profiles for prestige and dominance. These ratings confirmed that the mean dominance rating of the dominant target (mean ± *SD* = 6.58 ± 0.65) was significantly higher than the ratings of both the prestigious (mean ± *SD* = 4.66 ± 1.46) and average (mean ± *SD* = 3.27 ± 1.32) targets ($F_{2,94} = 87.99$, $p < 0.001$; Fig. 5.2a), and the prestigious target was rated as more prestigious (mean ± *SD* = 6.06 ± 1.04) than the dominant (mean ± *SD* = 4.25 ± 1.49) and average (mean ± *SD* = 3.44 ± 1.22) targets ($F_{2,94} = 57.62$, $p < 0.001$; Fig. 5.2b).

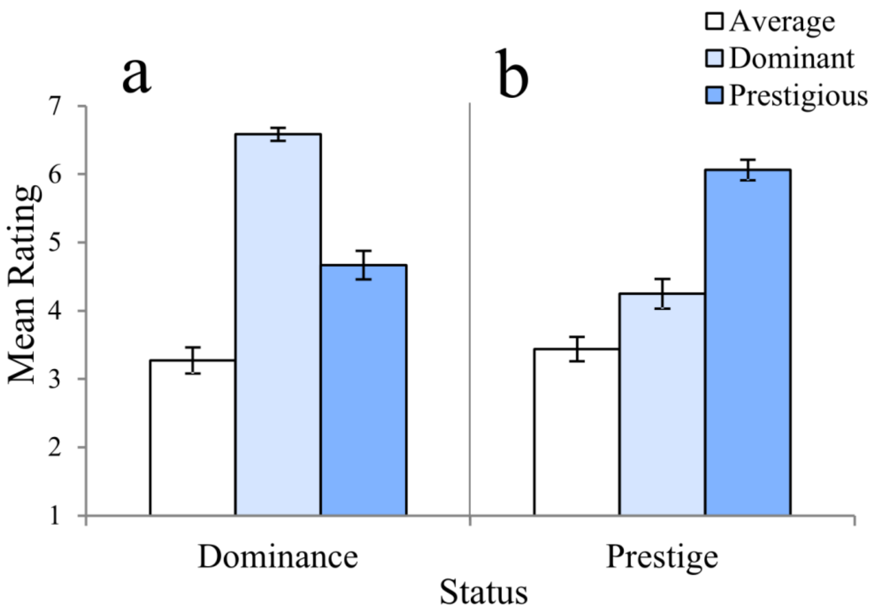


Figure 5.2. Ratings of dominance (a) and prestige (b) of the final target profiles (average: white bars; dominant: light blue bars; prestigious: dark blue bars) as were presented to participants. Bars represent estimated marginal means ± 1 s.e.m.

5.2.5 Data Analysis

I analysed each recording using Praat, obtaining values every 10 ms on intensity (dB) and F_0 (Hz). F_0 was measured using a noise-resistant autocorrelation method, between 75 and 300 Hz for male voices, and 100 and 500 Hz for female voices, as recommended by the software programmers. To ensure that intensity values were not affected by differences in the length or number of silent periods, and to control for background noise during these, I only used values which corresponded to times points in which the Praat algorithm produced a value of pitch.

For the statistical analysis, I calculated five variables from each recording, two of which were related to intensity: mean intensity and intensity variability (intensity SD), and three to F_0: mean F_0, F_0 variability (F_0 SD), and minimum F_0. These final values were analysed using repeated-measures generalised linear models (GLM) for each parameter (with Holm-Bonferroni (Holm, 1979) adjustments for multiple tests, because I performed two analyses of intensity parameters, and three of F_0 parameters), using sex of the participant (PS) as a between-subjects factor, target and question as within-subjects factors, and participant dominance (PD) and participant prestige (PP) as covariates. All tests are two-tailed.

5.3 Results

First, I tested whether individuals' self-rated status (prestige and dominance) predicted their vocal parameters, in response to each target. Then, I tested if individuals altered their vocal parameters in speech directed at dominant or prestigious individuals. I conducted separate analyses testing within-subject differences in parameters related to intensity (mean intensity and intensity SD) and F_0 (mean F_0, F_0 SD, and minimum F_0), with planned contrasts (Helmert) comparing responses to the average versus the high status targets (dominant and prestigious), and between the two high status targets (dominant versus prestigious).

5.3.1 Relationships between vocal parameters and self-rated status

As I predicted participants would adjust their vocal characteristics based on their self-rated status (prestige and dominance), in my analyses I used these self-ratings as covariates, and tested whether there were relationships between each acoustic parameter, in response to each target, and the participants' own ratings of dominance (PD) and prestige (PP; Table 5.1). Mean (± SD) self-rated scores of PD were 3.07 ±

0.56 and 2.71 ± 0.91 for men and women, respectively; scores for PP were 4.66 ± 0.59 and 4.79 ± 0.83. Because there were no significant differences in PD or PP between men and women (*t*-tests: PD: $t_{46} = 0.63$, $p = 0.11$; PP: $t_{46} = 1.67$, $p = 0.53$), I pooled these data in the analyses below.

Table 5.1. Correlations between vocal parameters in responses to each target and participants' status.

	Mean Intensity			Intensity SD			Mean F_0			F_0 SD			Minimum F_0		
	A	D	P	A	D	P	A	D	P	A	D	P	A	D	P
PD	0.113	0.023	0.115	0.073	-0.025	0.013	**-0.254***	**-0.282****	**-0.291****	-0.165	**-0.256***	-0.226	-0.161	-0.183	-0.045
PP	0.023	-0.004	-0.042	**-0272***	**-0.283***	-0.215	0.037	0.035	0.023	0.125	0.075	0.138	0.020	0.058	0.012

PD = Participant Dominance, PP = Participant Prestige. Results are from correlations for the responses to each target (A = Average, D = Dominant, P = Prestigious) with participants' status (PD, PP), for each vocal parameter. *p < 0.1; **p < 0.05

As expected, participants who rated themselves as higher in dominance had lower F_0, although this trend did not reach significance in responses to the average target ($p = 0.08$).There were also non-significant trends for more prestigious participants to vary their intensity less, and for more dominant individuals to speak with lower F_0 SD, particularly when responding to the dominant target.

5.3.2 Intensity Parameters

Previous research showed that voices with higher mean amplitude and amplitude SD (amplitude is directly proportional to intensity) are perceived as more dominant (Tusing & Dillard, 2000). Because of this, I anticipated that participants would adjust the intensity of their voices depending on the perceived status (dominance or prestige) of the targets, and their self-perceived dominance (PD) and prestige (PP). However, the analysis of intensity parameters revealed no significant differences in the mean intensity or intensity SD of the participants' responses depending on the target, even when controlling for PP and PD, nor a significant interaction between participant sex and target (for detailed results, see Table D2 in Appendix D).

5.3.3 Fundamental Frequency (F_0) Parameters

The analysis of F_0 parameters revealed that mean F_0 was particularly sensitive to my manipulation (Table 5.2). Although the main effect of target did not reach significance, it showed a trend in which the mean F_0 of the participants progressively increased in responses to the average, dominant, and prestigious targets (Fig. 5.3). When controlling for PD, this trend did reach significance (p = 0.01), suggesting that participants raise their F_0 when responding to high status targets (Table 5.2).

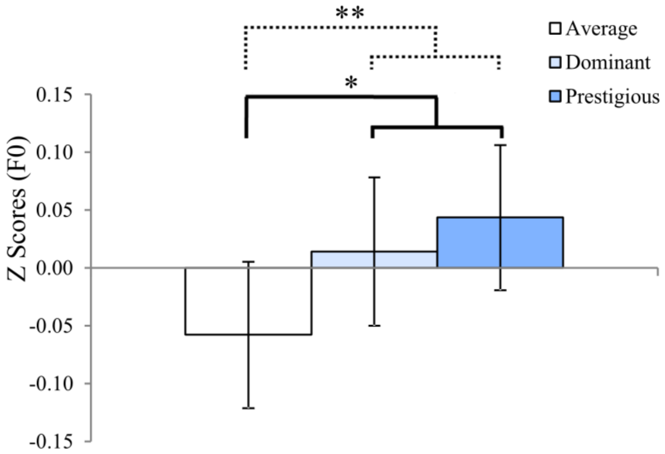


Figure 5.3. Participants' mean F_0 in responses to the three targets (average: white bars; dominant: light blue bars; prestigious: dark blue bars). Results were standardised (to z scores) for each participant to make results equivalent and account for between-subjects' differences. Solid lines represent a significant difference between responses to average versus high status targets (dominant and prestigious); dashed lines represent a significant difference between responses to average versus high status targets when controlling for participant dominance (PD). $*p < 0.05$, $**p < 0.01$. Bars represent estimated marginal means ± 1 s.e.m.

Table 5.2. Context-dependent variation in vocal parameters related to F_0.

Within-subject Effect	Mean F_0			F_0 SD			Min F_0		
	F	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**
T	2.484	2, 82	0.09	*1.195*	*1.65, 67.48*	*0.302*	0.076	2, 82	0.927
T * PD	**4.56**	**2, 82**	**0.013**	*1.136*	*1.65, 67.48*	*0.319*	2.331	2, 82	0.104
T * PP	0.214	2, 82	0.808	*0.631*	*1.65, 67.48*	*0.506*	0.427	2, 82	0.654
T * PS	1.526	2, 82	0.223	*3.078*	*1.65, 67.48*	*0.062*	0.088	2, 82	0.916
Q	***6.775***	***1.50, 61.61***	***0.005***	3.078	2, 82	0.051	0.462	2, 82	0.632
Q * PD	*0.404*	*1.50, 61.61*	*0.611*	1.39	2, 82	0.255	2.644	2, 82	0.077
Q * PP	***6.439***	***1.50, 61.61***	***0.006***	2.606	2, 82	0.08	0.024	2, 82	0.976
Q * PS	***14.394***	***1.50, 61.61***	***<0.001***	**15.31**	**2, 82**	**<0.001**	1.443	2, 82	0.242
T * Q	***3.379***	***2.72, 111.33***	***0.025***	0.841	4, 164	0.501	*2.428*	*2.52, 103.13*	*0.08*
T * Q * PD	***3.984***	***2.72, 111.33***	***0.012***	1.265	4, 164	0.286	*2.416*	*2.52, 103.13*	*0.081*
T * Q * PP	*1.412*	*2.72, 111.33*	*0.245*	0.532	4, 164	0.712	*0.717*	*2.52, 103.13*	*0.521*
T * Q * PS	*2.015*	*2.72, 111.33*	*0.122*	1.358	4, 164	0.251	*1.763*	*2.52, 103.13*	*0.168*

T = Target (average, dominant, prestigious), Q = Question, PD= Participant Dominance, PP = Participant Prestige, PS = Participant Sex (male, female), A = Average Target, HS = High Status Targets (dominant, prestigious), D = Dominant Target, P = Prestigious Target. Results are from repeated-measures generalized linear models for each vocal parameter, with Holm–Bonferroni adjustment for multiple tests. Significant effects are in bold. Where sphericity could not be assumed, Greenhouse–Geisser correction was used (in italics). For all results, including between-subjects effects, see table S2 in the supplementary material.

Planned contrasts revealed that in the cases of the main effect of target and the interaction between target and PD, there was a significant difference in the mean F_0 of the participants between the average versus the high status targets (dominant, prestigious), but not between the two high status targets (dominant versus prestigious; Table 5.3). Simply, this suggests that participants were raising their F_0 when speaking to high status targets but not when speaking to the average target. Similarly, for the interaction between target and participant sex, F_0 SD was significantly different when comparing responses to the average versus the high status targets, but not between the two high status targets (Table 5.3, Fig. 5.4b). Thus, it appears that women varied F_0 more when talking to average targets than dominant and prestigious targets, while the opposite effect was evident in men: they varied their F_0 less when speaking to average targets than dominant and prestigious targets.

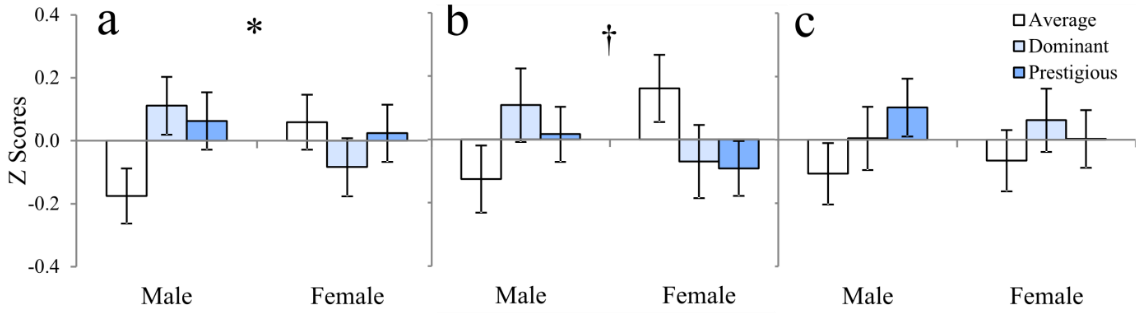


Figure 5.4. Modulation of acoustic vocal parameters related to F_0 in speech towards the three targets (average: white bars; dominant: light blue bars; prestigious: dark blue bars), split by sex of the participants. (a) Mean F_0; (b) F_0 SD; (c) Minimum F_0. Standard deviation (SD) was used as a measure of variability. Results were standardised (to z scores) for each participant to make results equivalent and account for between-subjects' differences. * represents significant main effect of target; † represents a significant interaction between target and participant dominance (PD). Bars represent estimated marginal means ± 1 s.e.m.

In addition, the general analysis and planned contrasts revealed the importance of the effects of question in the vocal parameters of spoken responses: there was a significant main effect of question, as well as significant interactions between question and PP on the mean F_0 of the participants (Table 5.2), and a significant interaction between question and participant sex for both mean F_0 and F_0 SD (Table 5.2); furthermore, the interaction between target and question was significant, as well as the interaction between target, question and PD, suggesting that the specific characteristics of the questions (introductory, personal, interpersonal) had an effect on the vocal parameters of the responses (Table 5.2). Planned contrasts revealed that in the cases of the interactions between target and question (for mean F_0), and the interaction between target, question and participant sex (for F_0 SD), there was a significant difference between the average versus and high status targets, but not between the high status targets (Table 5.3).

Table 5.3. Planned contrasts for variation in vocal parameters related to F_0.

Effect	Planned Contrasts	Mean F_0		F_0 SD		Minimum F_0	
		F	**p**	***F***	**p**	***F***	**p**
T	A vs HS	**4.31**	**0.044**	0.889	0.351	0.18	0.673
	D vs P	0.847	0.363	1.558	0.219	0.009	0.926
T * PD	A vs HS	**9.019**	**0.005**	2.018	0.163	0.447	0.508
	D vs P	0.563	0.458	0.086	0.77	3.531	0.067
T * PP	A vs HS	0.152	0.699	0.047	0.83	0.043	0.836
	D vs P	0.269	0.607	1.325	0.256	0.672	0.417
T * PS	A vs HS	3.078	0.087	**5.318**	**0.026**	0.124	0.726
	D vs P	0.134	0.716	0.413	0.524	0.065	0.8
T * Q	A vs HS	**4.435**	**0.041**	1.191	0.281	4.036	0.051
	D vs P	2.253	0.141	1.93	0.172	0.539	0.467
T * Q * PD	A vs HS	2.538	0.119	3.827	0.057	1.12	0.296
	D vs P	3.339	0.075	1.227	0.275	0.09	0.765
T* Q * PP	A vs HS	2.73	0.106	0.042	0.838	1.966	0.168
	D vs P	0.42	0.52	0.719	0.401	0.236	0.63
T * Q * PS	A vs HS	3.89	0.055	**6.952**	**0.012**	0.057	0.812
	D vs P	1.731	0.196	0.099	0.754	0.303	0.585

T = Target (average, dominant, prestigious), Q = Question, PD= Participant Dominance, PP = Participant Prestige, PS = Participant Sex (male, female), A = Average Target, HS = High Status Targets (dominant, prestigious), D = Dominant Target, P = Prestigious Target. Results are from planned contrasts (Helmert) for each vocal parameter (d.f. = 1, 41). Significant effects are in bold.

5.3.4 Analysis of Fundamental frequency (F_0) parameters by question

Paralinguistic parameters thus vary depending on the target, but participants changed their vocal characteristics of their responses according to the question they were responding to. To further explore this connection, I split the analysis by question in order to test the effect that the specific context of each question had on the responses.

This analysis revealed that in the case of question 1 (Introductory), there were no significant differences in the vocal parameters of the participants depending on the target they were responding to (Table 5.4, Fig. 5.5).

Table 5.4. Context-dependent variation in vocal parameters related to F_0 by question.

Effect	Mean F_0			F_0 SD			Min F_0		
	F	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**
Question 1 (Introductory)									
T	1.761	2, 84	0.178	0.006	2, 84	0.994	*0.006*	*1.60, 67.23*	*0.994*
T x PD	2.355	2, 84	0.101	0.287	2, 84	0.751	*1.626*	*1.60, 67.23*	*0.203*
T x PP	0.801	2, 84	0.452	0.116	2, 84	0.89	*0.696*	*1.60, 67.23*	*0.501*
T x PS	0.378	2, 84	0.686	0.053	2, 84	0.948	*1.013*	*1.60, 67.23*	*0.367*
Question 2 (Personal)									
T	***5.148***	***1.75, 75.16***	***0.011***	*0.248*	*1.52, 65.46*	*0.72*	1.458	2, 86	0.238
T x PD	***8.543***	***1.75, 75.16***	***0.001***	*0.43*	*1.52, 65.46*	*0.598*	1.44	2, 86	0.243
T x PP	*1.598*	*1.75, 75.16*	*0.211*	*0.152*	*1.52, 65.46*	*0.801*	0.337	2, 86	0.715
T x PS	*1.635*	*1.75, 75.16*	*0.204*	*1.045*	*1.52, 65.46*	*0.341*	3.112	2, 86	0.05
Question 3 (Interpersonal)									
T	**4.31**	**2, 88**	**0.016**	*2.053*	*1.74, 76.48*	*0.141*	**3.114**	**2, 88**	**0.049**
T x PD	**6.065**	**2, 88**	**0.003**	*2.929*	*1.74, 76.48*	*0.067*	**3.697**	**2, 88**	**0.029**
T x PP	0.909	2, 88	0.407	*1.309*	*1.74, 76.48*	*0.274*	0.919	2, 88	0.403
T x PS	**5.815**	**2, 88**	**0.004**	***4.84***	***1.74, 76.48***	***0.014***	0.001	2, 88	0.999

T = Target (average, dominant, prestigious), PD= Participant Dominance, PP = Participant Prestige, PS = Participant Sex (male, female). Results are from repeated measures generalized linear models for each vocal parameter. Significant effects are in bold. Where sphericity could not be assumed, Greenhouse–Geisser correction was used (in italics).

Participants responding to questions 2 (Personal) and 3 (Interpersonal), did vary their mean F_0 according to the target they were responding to, even when controlling for PD (Table 5.4, Fig. 5.5a). Planned contrasts revealed that in responses to question 2 (Personal), mean F_0 was significantly lower when responding to high status versus average targets (Table 5.5, Fig. 5.5a). Additionally, responses to question 3 (Interpersonal) were significantly different for male and female participants depending on the target: while the mean F_0 and F_0 SD of male participants were lower in responses to the average target, female participants had lower mean F_0 and F_0 SD in responses to the dominant target (Table 5.4, Fig. 5.5a,b).

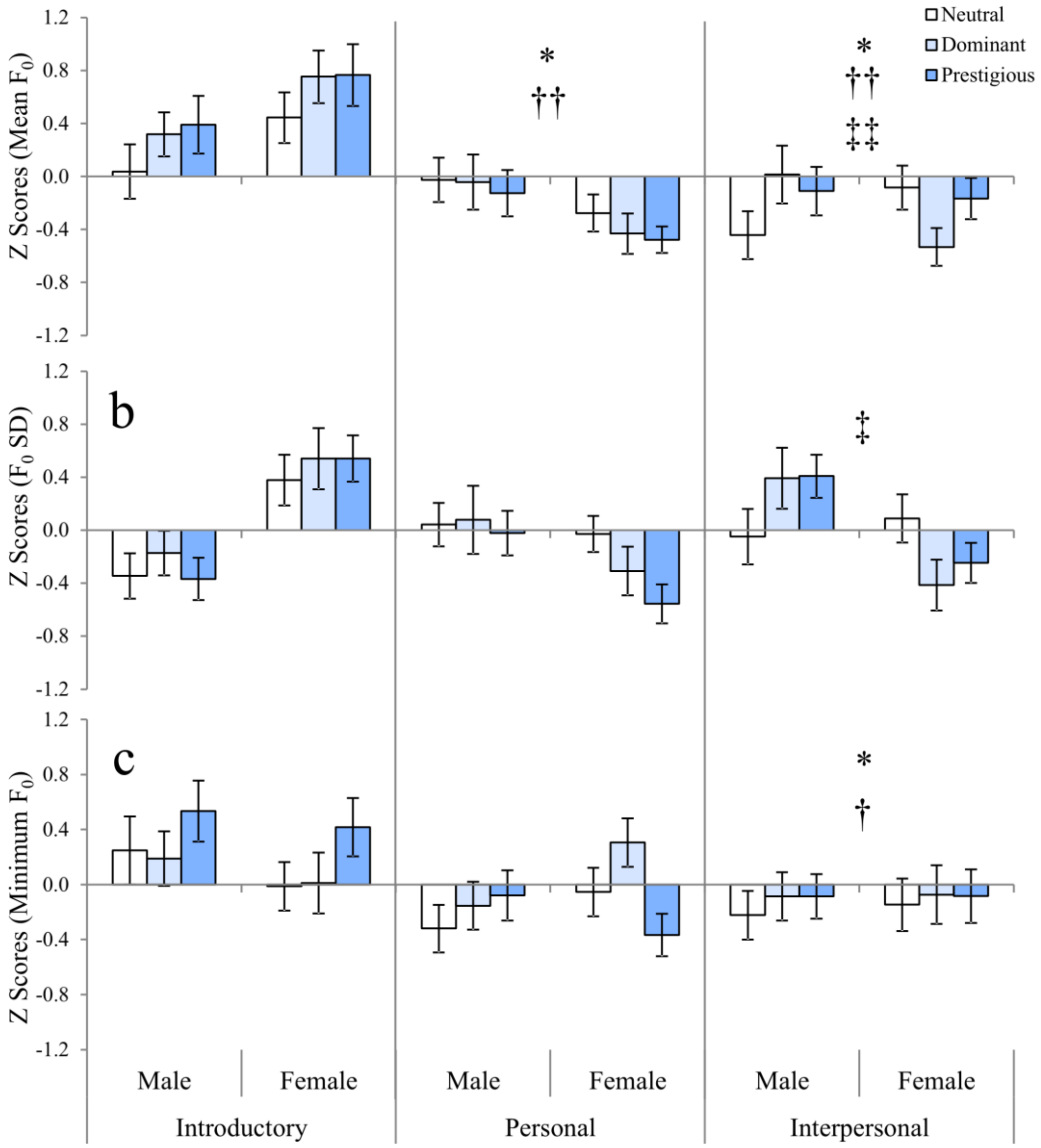


Figure 5.5. . Modulation of acoustic vocal parameters related to F_0 in speech towards the three targets (average: white bars; dominant: light blue bars; prestigious: dark blue bars), split by question (Introductory, Personal, Interpersonal), and sex of the participants. (a) Mean F_0; (b) F_0 SD; (c) Minimum F_0. Standard deviation (SD) was used as a measure of variability. Results were standardised (to z scores) for each participant to make results equivalent and account for between-subjects' differences. Bars represent estimated marginal means ± 1 s.e.m. * represents a significant main effect of target; † represents significant interactions between target and participant dominance (PD); ‡ represents significant interactions between target and participant sex (PS). *,†, or ‡, $P < 0.05$, ††, or ‡‡, $P < 0.01$.

In responses to question 2 (personal) the interaction between target and PD, F_0 was not only different between the high status and average targets, but also between the dominant and prestigious targets (Table 5.5, Fig. 5.5a). The minimum F_0 was significantly different between the dominant and prestigious target in question 2 (Personal), but not between the average and high status targets (Table 5.5, Fig. 5.5c).

Table 5.5. Planned contrasts for variation in vocal parameters related to F_0.

Effect	Planned contrasts	Question 1 (Introductory)						Question 2 (Personal)						Question 3 (Interpersonal)					
		Mean F_0		F_0 SD		Min F_0		Mean F_0		F_0 SD		Min F_0		Mean F_0		F_0 SD		Min F_0	
		F	p	*F*	p	*F*	p	*F*	p	*F*	p	*F*	p	*F*	p	*F*	p	*F*	p
T	A vs HS	0.615	0.437	0.012	0.913	0.234	0.631	**10.977**	**0.002**	0.411	0.525	2.317	0.135	**7.363**	**0.009**	1.092	0.302	**6.358**	**0.015**
	D vs P	2.553	0.118	0.002	0.968	0.052	0.820	0.317	0.576	0.094	0.760	0.444	0.509	0.002	0.966	3.055	0.087	0.579	0.451
T x PD	A vs HS	0.025	0.875	0.000	0.994	0.391	0.535	**13.390**	**0.001**	0.422	0.519	2.640	0.112	**10.218**	**0.003**	**5.378**	**0.025**	**4.417**	**0.041**
	D vs P	3.967	0.053	0.533	0.469	2.393	0.129	**4.528**	**0.039**	0.437	0.512	0.025	0.876	0.203	0.655	0.377	0.542	3.134	0.084
T x PP	A vs HS	1.481	0.230	0.072	0.790	0.530	0.470	2.948	0.093	0.306	0.583	0.307	0.583	1.550	0.220	0.014	0.907	2.076	0.157
	D vs P	0.330	0.569	0.155	0.696	0.799	0.376	0.479	0.493	0.006	0.938	0.372	0.545	0.005	0.942	2.658	0.110	0.016	0.899
T x PS	A vs HS	0.325	0.572	0.001	0.976	0.159	0.692	3.241	0.079	2.129	0.152	0.656	0.422	**7.799**	**0.008**	**9.414**	**0.004**	0.000	0.990
	D vs P	0.415	0.523	0.098	0.756	1.544	0.221	0.303	0.585	0.020	0.888	**6.009**	**0.018**	3.014	0.090	0.073	0.788	0.001	0.971

T = Target (neutral, dominant, prestigious), PD= Participant Dominance, PP = Participant Prestige, PS = Participant Sex (male, female), A = Average Target, HS = High Status Targets (dominant, pretigious), D = Dominant Target, P = Prestigious Target. Results are from planned contrasts (Helmert) for each vocal parameter (d.f. = 1, 42 for question 1, 1, 43 for question 2, and 1, 44 for question 3).Significant effects are in bold.

In responses to question 3 (Interpersonal), responses were significantly different in both mean F_0 and minimum F_0 between the average and high status targets, but not between the two high status targets (Table 5.5, Fig. 5.5a,c). When controlling for PD, this was also true in F_0 SD (Table 5.5, Fig. 5.5). Likewise, the contrast between the average and high status targets revealed a significant interaction between target and participant sex (Table 5.5, Fig. 5.5a,b).

5.4 Discussion

Previous studies have suggested that manipulations of vocal parameters, particularly F_0, affect perceived dominance (Puts et al., 2007), that men adjust their voices during interaction with competitors depending on their perceived relative dominance (Puts et al., 2006) and, more generally, that authority relationships are dependent on relative, rather than absolute, social status perceptions (Watkins, Jones, et al., 2010; Watkins, Fraccaro, et al., 2010). Such studies have, however, focused on dominance, and predominantly on men's voices. My experimental design of a job interview scenario provides new insights into the specific nature of authority relationships and into the vocal differences when addressing dominant and prestigious individuals in both men and women.

Firstly, I found that male and female participants who judged themselves to be more dominant lowered their F_0 when speaking to all targets, in line with previous research on men (Puts et al., 2006). I also found a tendency for more prestigious participants to respond with lower intensity variability, and dominant participants to decrease variability in fundamental frequency (F_0 SD), which would perhaps make them sound calmer and more in control of situations; in fact, decreased F_0 variability is associated with lower aggressiveness in industrial as well as foraging societies (Puts et al., 2012), and it is known to occur in contexts involving competition (Hodges-Simeon et al., 2010, 2011).

Differences in vocal parameters between responses to the different targets were especially noticeable in mean F_0 (Tables 5.2 and 5.3), and when controlling for self-perceived dominance. As predicted by previous research (Puts et al., 2006), participants, and particularly men, responded with a relative higher F_0 when speaking to the high status targets. Additionally, most significant differences arise when comparing responses to the average versus the high status targets, suggesting that the status of the target, whether by means of dominance or prestige, is the key factor. In fact, differences between responses to the dominant and prestigious targets were not significantly different when analysing all questions together.

Contextual vocal modulations, however, were not found to occur in mean intensity or intensity SD. This suggests that while these parameters can be a robust cue of social status, as shown above in the self-perceived prestige and intensity variability, or even context-dependent (e.g. shouting) interactions, speakers do not modulate their voice intensity during free speech depending solely on the relative social status of the listeners. This is likely due to the nature of my interview scenario, as participants were not directly competing, and were not trying to signal aggression in front of a potential employer, but rather to make themselves appear favourable for a position.

Furthermore, the use of a job interview scenario allowed us to include questions with different characteristics: introductory, personal, and interpersonal. The analysis of the vocal characteristics by question revealed significant vocal differences dependent on the perceived social status of the target listener when personal and interpersonal questions are answered, but not during introductory responses. In these cases the effects of target, especially when controlling for PD, were significant (Table 5.4). In general, participants' mean F_0 was raised when responding to the dominant or prestigious targets (Fig. 5.1), and this was especially true in men (Fig. 5.2), supporting previous results (Puts et al., 2006). This may be because participants tended to introduce themselves in a very similar manner to all targets (e.g. "my name is…", "I am currently studying…", "I live in…"), but when confronted with questions that required them to discuss their specific skills to the target (personal), and even more so when asked to imagine a hypothetical interaction with the target (interpersonal), the nature of the questions themselves may have induced participants to *improvise* and respond more naturally.

Differences in vocal parameters between the responses to these questions are apparent in my analysis. Although it could be argued that this is a product of the order in which the questions were presented, I suggest that this is unlikely because of the different characteristics of the questions and, furthermore, because participants participated in three *interviews*, which meant that they responded to question one (introductory) after question 3 (interpersonal) twice during the experiment. The possibility of order effects could be tested in future experiments, to disentangle responses to different types of questions. Additionally, in my experiment I only tested responses to male targets; future studies could address vocal modulations in response to men, but also women, of varying social status.

In conclusion, using a novel job interview scenario, I found that self-perceptions of dominance and prestige affected vocal parameters such that the higher an individual's self-perceived dominance, the lower their mean F_0, and the higher their self-perceived prestige, the higher their intensity variability. Additionally, regardless of self-perceived status, participants changed their vocal characteristics when talking to average versus high status targets, displaying a relatively higher mean F_0 when talking to high status targets. The context of questions (i.e. introductory, personal, or interpersonal) also affected participants' vocal characteristics with the greatest changes in F_0 according to status of the listener observed for the responses to the personal and interpersonal questions. These F_0 effects were even more pronounced when controlling for participant self-perceived dominance. Ultimately my findings suggest that individuals' vocal characteristics are influenced, whether consciously or non-consciously, by the relative difference between their self-perceived social status and the social status of the listeners.

CHAPTER 6:

MECHANICAL VERSUS PERCEPTUAL MANIPULATIONS OF VOCAL FREQUENCY: IMPLICATIONS FOR PSYCHOLOGICAL STUDIES

6.1 Introduction

The way our brain processes acoustic pitch is intriguing. While the perception of the pitch of a sound depends on its frequency (the higher the frequency, the higher the pitch we hear), this relationship is not linear, as is exemplified by musical scales (Fig. 6.1). Small frequency changes, for example, can be easily perceived in low pitch sounds, but become virtually impossible to perceive in increasingly higher sounds. This is a technical problem that has important consequences in several areas, including music acoustics, sound engineering, and sound perception.

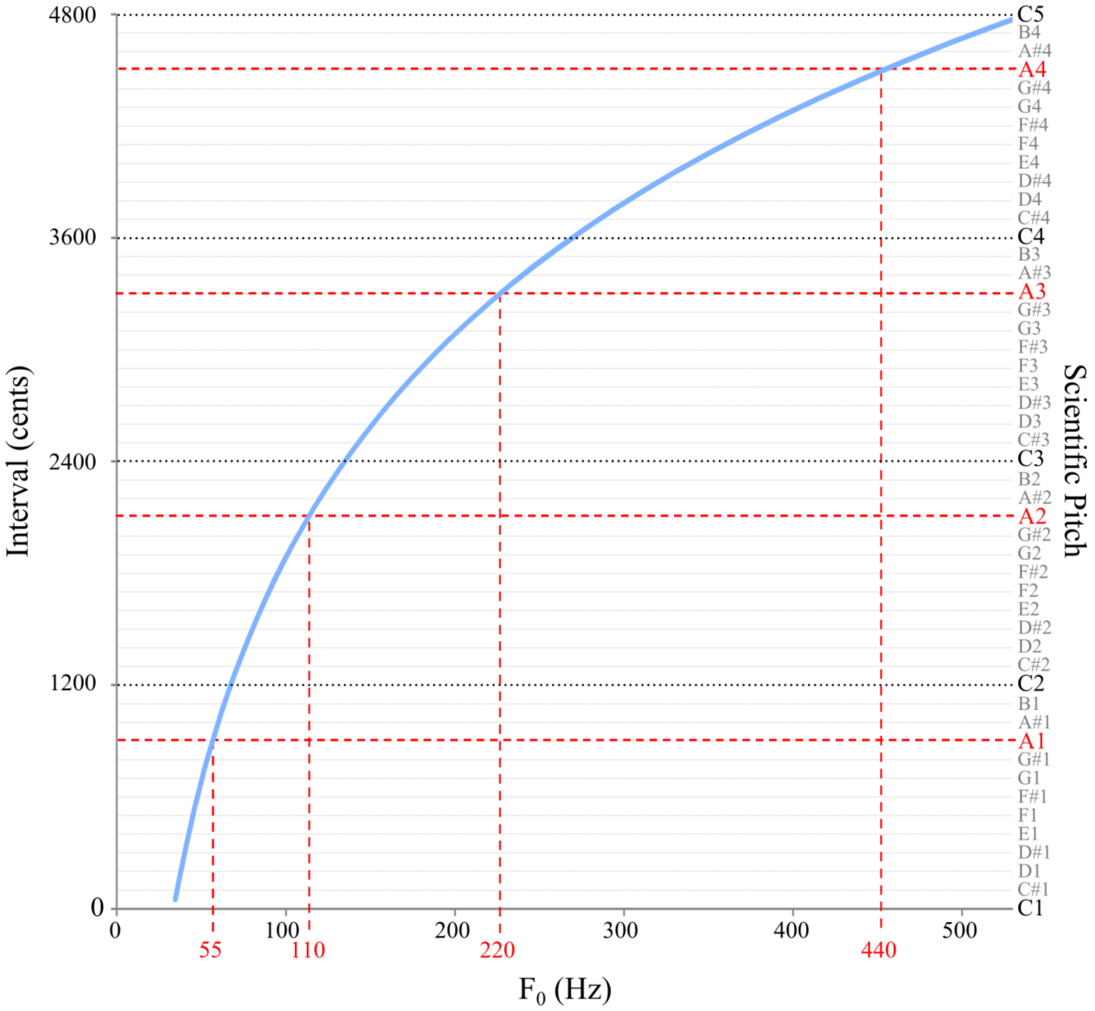


Figure 6.1. Relationship between intervals and scientific pitch with fundamental frequencies (F_0). The X axis represents F_0 in Hz, and the Y axis represents a four-octave scale in cents (1/100 of a semitone)

starting with C1 (left Y axis), and Scientific Pitch (right Y axis) from C1 to C5. Semitones are a measure of relative pitch, or interval (the perceptual *distance* between two pitches), and therefore require a point of reference; here, semitone steps (in cents) are calculated in relation to C1 (32.703 Hz). Horizontal grey lines represent each semitone step; horizontal black dotted lines represent each instance of C. Red dashed lines and numbers represent instances of A on cents and F_0, highlighting the relationship between frequencies and pitch (A1 = 55 Hz, A2 = 110 Hz, A3 = 220 Hz, A4 = 440 Hz).

Musical scales work to the point that a melody can be transposed to different keys, and still be easily identified as the *same*. However, experimental psychoacoustic data suggest that the relation between pitch and frequency is similar, but more complex, than assumed by musical scales (e.g. Dai & Micheyl, 2011; Stevens, Volkmann, & Newman, 1937a, 1937b; Umesh, Cohen, & Nelson, 2002). In fact, several psychoacoustic scales have been proposed, and for most of them numerous equations to transform hertz into an approximation of perceptual pitch have been developed, demonstrating the challenges that developing an accurate psychoacoustic pitch scale poses.

Among the several psychoacoustic scales that attempt to accurately represent perceptual pitch (commonly transforming frequencies into a logarithmic or near-logarithmic scale), the most used ones are the semitone scale, based on an equal temperament musical scale (in which each octave is divided into 12 equal steps), the Bark Scale (Zwicker, 1961), the Equivalent Rectangular Bandwidth (ERB-rate) scale (Moore & Glasberg, 1983), and the mel scale (Stevens et al., 1937a).

For example, two well-known equations have been proposed for the Bark scale. This scale is based on critical bands of hearing (created by the cochlea) within which a sound with a different frequency will interfere with the perception of another sound.

More specifically, the Bark scale is based on the lower 24 critical bands of frequency, with cut-off frequencies ranging roughly from 20 to 15500 Hz. To transform a frequency f (Hz) into barks, Zwicker (1961) proposed the following equation: $\text{Bark} = 13\arctan(0.00076*f) + 3.5\arctan(f/7500)^2$

Traunmüller (1990), however, proposed a different equation (for frequencies between 20 and 6700 Hz): $\text{Bark} = [26.81/(1+1960/f)] - 0.53$.

Similarly, the ERB-rate scale (ERBS) transforms a frequency f into the number of equivalent rectangular bandwidths below the given frequency, for moderate sound levels. The original transformation (Moore & Glasberg, 1983) for a frequency f in kHz (between 0.1 and 6.5 kHz) can be expressed by the following equation: $\text{ERBS} = 11.17 * \log_e \{1 + [(f + 0.312)/(f + 14.675)]\} + 43$.

For a frequency f in Hz (between 100 and 10000 Hz), Glasberg & Moore (1990) later proposed a new approximation that can be expressed as: $\text{ERBS} = 21.4 * \log_{10} (1 + 0.00437 * f)$.

The mel scale, widely used by engineers, was first proposed by Stevens, Volkmann, & Newman (1937a), and attempts to create a purely perceptual scale in which subsequent steps are perceived by listeners as having the same *distance* (in the manner of musical intervals). Its history exemplifies the complications and challenges of creating a psychoacoustic scale that accurately represents pitch perception. This scale has been strongly criticized for having methodological flaws (Greenwood, 1997), including order biases in the presentation of the stimuli, and a relatively small sample size. However, in the words of Russo & Thompson (2005, p. 1069) "it is difficult to dismiss the data that were used to derive the mel scale. Specifically, they imply that there is a dimension of melody perception that is somewhat independent of the explicit labels associated with musical intervals" (see also Makeig, 1982). In fact, as in the case of the Bark ERB-rate scales, several equations have been proposed for the mel scale (Fig. 6.2). The most commonly used (Fant, 1968; Lindsay & Norman, 1977; O'Shaughnessy, 1987) tends to assign a value of 1000 mel to 1000 Hz.

This diversity of transformations is partially a product of the fact that many different equations can fit the data obtained in experiments to calculate the mel scale (Umesh, Cohen, & Nelson, 1999). Moreover, the curves generated by these equations can provide closer approximations to the mel scale in certain frequency regions, at the expense of others (see Ganchev, Fakotakis, & Kokkinakis, 2005). Indeed, Praat (P. Boersma and D. Weenink, 2014; www.praat.org), uses a different equation (Fig. 6.2) to transform an *f* acoustic fundamental frequency (Hz) into perceptual pitch (mel), in which the only value assigned in mel that corresponds to Hz is 0.

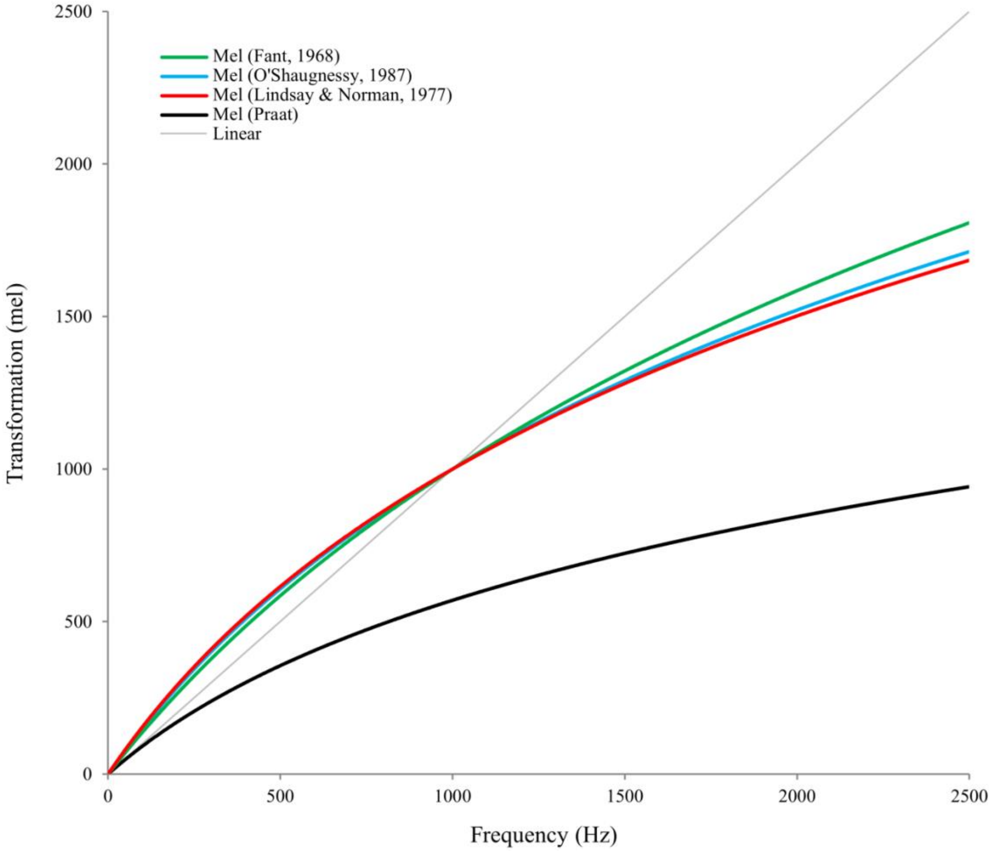


Figure 6.2. Representation of common mel scale approximations, and their relation to frequency between 0 and 2500 Hz. Each curve represents a common equation to transform frequency (Hz) into mel. Green curve (Fant, 1968): $mel = (1000/\log_{10}2)*\log_{10}(1+f/1000)$; blue curve (O'Shaughnessy, 1987): $mel = 2595*\log_{10}[1 + (f/700)]$; red curve (Lindsay & Norman, 1977): $mel = 2410*\log_{10}(1.6 * 10^{-3}f + 1)$; black curve (Praat): $mel = 550 \log_e (1 + f/550)$. For reference, the grey line represents a linear relation.

The challenges of the measurement of perceptual pitch directly impact research in the psychological and behavioural sciences, where there is a growing body of research aiming to understand the effects and relative importance of specific vocal acoustic characteristics in the perception that listeners have of the speaker. Researchers have shown that changes in pitch-related parameters (particularly fundamental frequency, F_0, and formant dispersion, D_f), have important effects in perceptions of relevant physical information such as sex (Puts et al., 2012), body size (Collins, 2000;

Feinberg, Jones, Little, et al., 2005; Xu et al., 2013), physical strength (Sell et al., 2010), femininity (Feinberg, 2008; Feinberg, Jones, DeBruine, et al., 2005), and attractiveness (Collins & Missing, 2003; Collins, 2000; Feinberg, Jones, Little, et al., 2005; Feinberg, Jones, DeBruine, et al., 2005; Xu et al., 2013).

To understand the effects and relative importance of specific vocal acoustic characteristics, studies commonly record voices and manipulate single parameters to test how this changes the perception that listeners have of the speaker (e.g. Feinberg, Jones, Little, et al., 2005). Some studies, however, have manipulated F_0 by equal mechanical amounts, irrespective of the original F_0 of the voice. Feinberg and his colleagues (2005), for example, manipulated the voices of several men, increasing and decreasing their F_0 by 20 Hz. While their general findings are robust, manipulating F_0 of voices of different pitch in this manner means that, in perceptual terms, lower voices tended to be manipulated by a *larger* amount than relatively higher voices (Fig. 6.3), potentially resulting in over- or under-estimation of the results.

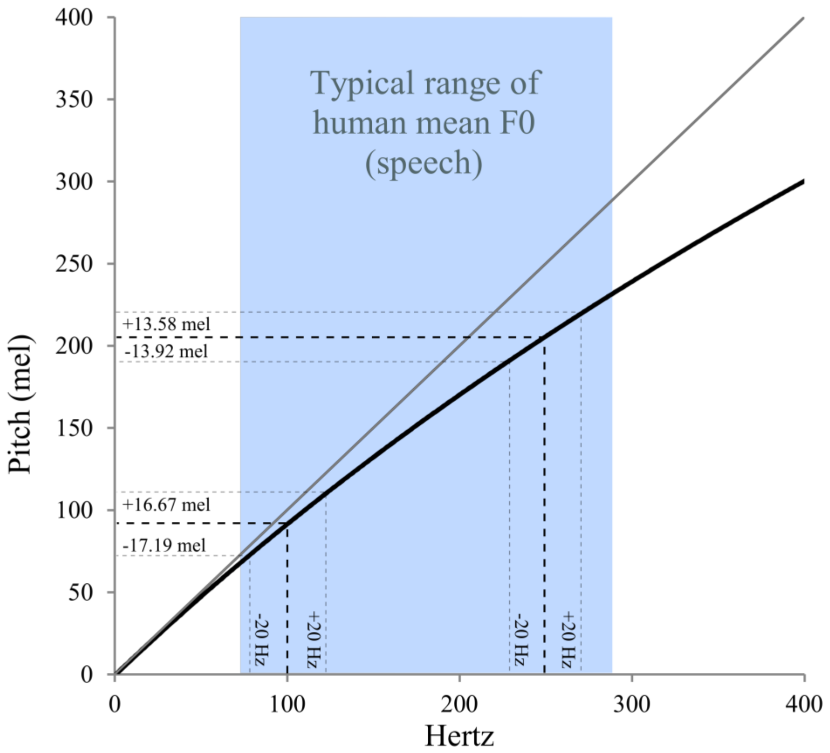


Figure 6.3. Difference in the perceptual magnitude of mechanical manipulations of F_0, using the mel scale (as calculated in Praat) as an approximation of perceptual pitch. The solid black curve represents the relation between frequencies in Hz (X axis) and their perceptual equivalent approximation (Y axis); for reference, the grey line represents a linear relation. As an example of the problems of manipulating voices by the same frequency amount, the resultant perceptual manipulation (mel) of two relatively extreme voices with original F_0 of 100 and 250 Hz (black dashed lines) raised and lowered by 20 Hz is shown (grey dashed lines) is shown.

To test whether the use of a mechanical, physical measure of frequency (Hz) in studies that manipulate F_0 could bias the results, I did a partial replication of the methods described in Feinberg et al. (2005). In that paper, the authors manipulated men’s voices F_0 (±20 Hz), as well as apparent vocal tract. Here, I manipulated only F_0, but using not only a mechanical (Hz) scale, but also a perceptual (mel) scale, and compared differences in ratings of attractiveness, masculinity and femininity, body size, and age. In addition, I manipulated voices of women as well as men, to increase the total F_0 range of the sample towards higher frequencies, where differences in the use of the two scales (Hz and mel) are more likely to become apparent.

Based on Feinberg et al. (2005), as well as the body of knowledge of voice perception, I expected that manipulations of F_0 would tend to affect the perception of men and women speakers in different ways, replicating previous results. First, while perceived attractiveness would be affected by manipulations increasing F_0, positively for women's voices but negatively for men's voices, manipulations lowering F_0 would increase men's, but decrease women's, perceived attractiveness. Second, lowering the voices would increase masculinity of men's voices and decrease femininity of women's voices, but increasing F_0 would decrease the masculinity and increase the femininity. Third, manipulations of voices' F_0 would affect both perception of body size and age in the same direction: lowering F_0 would increase perceived age and body size or both men and women, and the opposite would be true for manipulations increasing F_0.

Regarding the scale used for the manipulations, and based on the current knowledge about the relation between frequency and pitch, I expected that (1) manipulations in hertz would have a larger perceptual effect on relatively low voices in relation to relatively high voices, and (2) that manipulations in mel would tend to have a relatively constant perceptual effect on all voices, regardless of their original F_0.

6.2 Materials and Methods

6.2.1 Participants

78 heterosexual participants who were students at the University of Stirling (49 men, mean age ± SD = 21.2 ± 3.28; 31 women, 20.5 ± 1.39) took part in the experiment

as raters. Voice samples (English vowels A E I O and U with a British accent), were recorded from an independent sample of 60 British men (n = 30; mean age ± SD = 20.9 ± 0.88) and women (n = 30; 20.3 ± 0.95). All participants provided written informed consent.

6.2.2 Voice Stimuli

Monaural audio samples (A E I O U) were digitally recorded in a quiet room using Praat$^{©}$, Version 5.2 (P. Boersma and D. Weenink, 2011; www.praat.org) on a laptop PC, with a sampling frequency of 44.1 kHz, using a sE X1 Large Diaphragm Cardioid Condenser Microphone (http://www.seelectronics.com). Following the procedures of Feinberg et al. (2005), I excluded the vowel U from all recordings. Final recordings had a mean duration ± SD of 3.1 ± 0.4 s, and a fundamental frequency of 112.51 ± 22.10 Hz (ranging from 83.40 to 169.65 Hz) for men's voices, and 202.07 ± 24.54 Hz (ranging from 157.66 to 256.93 Hz) for women's voices.

Using Praat, mean F_0 was obtained with a noise-resistant autocorrelation method (between 75 and 300 Hz for male voices, and 100 and 500 Hz for female voices), and each one of the 60 voice samples was manipulated in Hz (both increasing and decreasing F_0 by 20 Hz, following Feinberg et al.'s (2005) procedure), as well as mel (increasing and decreasing it by 16.65 mel). The magnitude of 16.65 mel was calculated according to the mean F_0 (i.e. 111 Hz) of the sample used in Feinberg et al. (2005, p. 565); a typical voice in that study was then increased by 16.39 mel (111 to 131 Hz, or 101.11 to 117.50 mel), and lowered by 16.90 mel (111 to 91 Hz, or 101.11 to 84.21 mel) or, in other words, was manipulated by an average of 16.65 mel. To transform the F_0 frequency from Hz to mel, I used the equation used in Praat (Fig. 6.2). In the ratings of body size, for manipulations lowered using mel, responses to only 58 (and not 60) recordings were analysed, due to an experimenter's mistake; two manipulations of men's voices lowered in mel were incorrectly coded in the online experiment, which resulted in incorrect recordings being presented.

6.2.3 Experimental Procedure

To avoid distractions, raters were asked to complete the experiment in the lab. Participants were presented with the experiment using Qualtrics software (Qualtrics, Provo, UT, 2013; www.qualtrics.com), on a desktop computer equipped with Philips SBC HP250 headphones, with a frequency response of 20 to 20,000 Hz. Each participant was presented with the recordings of a subsample of 10 opposite-sex original voices, as well as the four manipulations (+20 Hz, -20 Hz, +16.65 mel, -16.65 mel) of each one of those original recordings; in total, each rater listened to 50 recordings. They were asked to rate each recording for attractiveness, masculinity (for men's voices) or femininity (for women's voices), and body size, on a 7-point rating scale (1 = low; 7 = high), as well as age (in years). The 50 recordings were presented in a different, fully randomised order to each rater.

6.2.4 Data Analysis

Within-subject differences in the ratings given to each manipulation were calculated, subtracting the rating given to each original recording by each rater from the rating given to each manipulation of each original recording by the same rater. To test the general tendency that the manipulations produced, mean differences in the ratings given to each manipulation were tested using one-sample *t*-tests (against a value of 0), and then mean differences in ratings for each speaker were correlated to the mean F_0 of their corresponding original recordings. Finally, coefficients of the correlations between rating differences (manipulation minus original) were compared using the Fisher r-to-z transformation for the two manipulations that increased F_0 (+20 Hz and +16.65 mel), and the two that lowered it (-20 Hz and -16.65 mel), for each dimension rated (attractiveness, masculinity/femininity, body size and age). All Fisher r-to-z transformations reported are one-tailed, as I had clear predictions about the direction of effect. In the cases of ratings of attractiveness and masculinity/femininity, independent correlations and comparisons between them were performed for male and female voices, because I predicted opposite effects of increasing or lowering F_0. Conversely, and because I anticipated manipulations of F_0 to affect perceptions in the same way, ratings of body size and age were analysed combining male and female voices, increasing the F_0 range of the original recordings.

6.3 Results

Because I predicted that (1) manipulations in hertz would have a larger perceptual effect on relatively low voices in relation to relatively high voices, and (2) that manipulations in mel would tend to have a relatively constant perceptual effect on all voices, regardless of their original pitch, I expected that correlations between mean rating difference and mean F_0 of the original recordings to be significant for all Hz manipulations (i.e. a tendency for the correlation to progressively approach a value of 0 in the Y axis as the F_0 of the original recordings, plotted in the X axis, increased). Conversely, I expected the same correlations for mel manipulations to be non-significant, as the average change for any perceived characteristic would tend to remain constant for all voices (because all recordings should be manipulated by the *same* perceptual amount), irrespective of their original F_0. Because of this, I also expected correlations for the manipulations in Hz to be statistically significantly different from those in mel.

6.3.1 Attractiveness

Perceived attractiveness was affected by the acoustic manipulations (Fig. 6.4), with increased F_0 showing a tendency to negatively affect attractiveness for men's voices (Fig. 6.4a) when performed both in Hz ($M = -0.40$, $SD = 0.49$; one-sample *t*-test: $t_{29} = -4.43$, $p < 0.001$) and mel ($M = -0.66$, $SD = 0.45$; one-sample *t*-test: $t_{29} = -7.95$, $p < 0.001$), but neither positively nor negatively for women's voices (Fig. 6.4b) when manipulations were performed in Hz ($M = 0.09$, $SD = 0.32$; one-sample *t*-test: $t_{29} = 1.46$, $p = 0.16$) or mel ($M = 0.06$, $SD = 0.45$; one-sample *t*-test: $t_{29} = 0.77$, $p = 0.45$). Manipulations lowering F_0 showed an unexpected tendency to decrease men's attractiveness (Fig. 6.4c) both in Hz ($M = -0.51$, $SD = 0.80$; one-sample *t*-test: $t_{29} = -3.47$, $p = 0.002$) and mel ($M = -0.32$, $SD = 0.49$; one-sample *t*-test: $t_{29} = -3.62$, $p = 0.001$), but as expected tended to decrease women's perceived attractiveness (Fig. 6.4d) both in Hz ($M = -0.22$, $SD = 0.42$; one-sample *t*-test: $t_{29} = -2.85$, $p = 0.008$) and mel ($M = -0.42$, $SD = 0.43$; one-sample *t*-test: $t_{29} = -4.97$, $p < 0.001$).

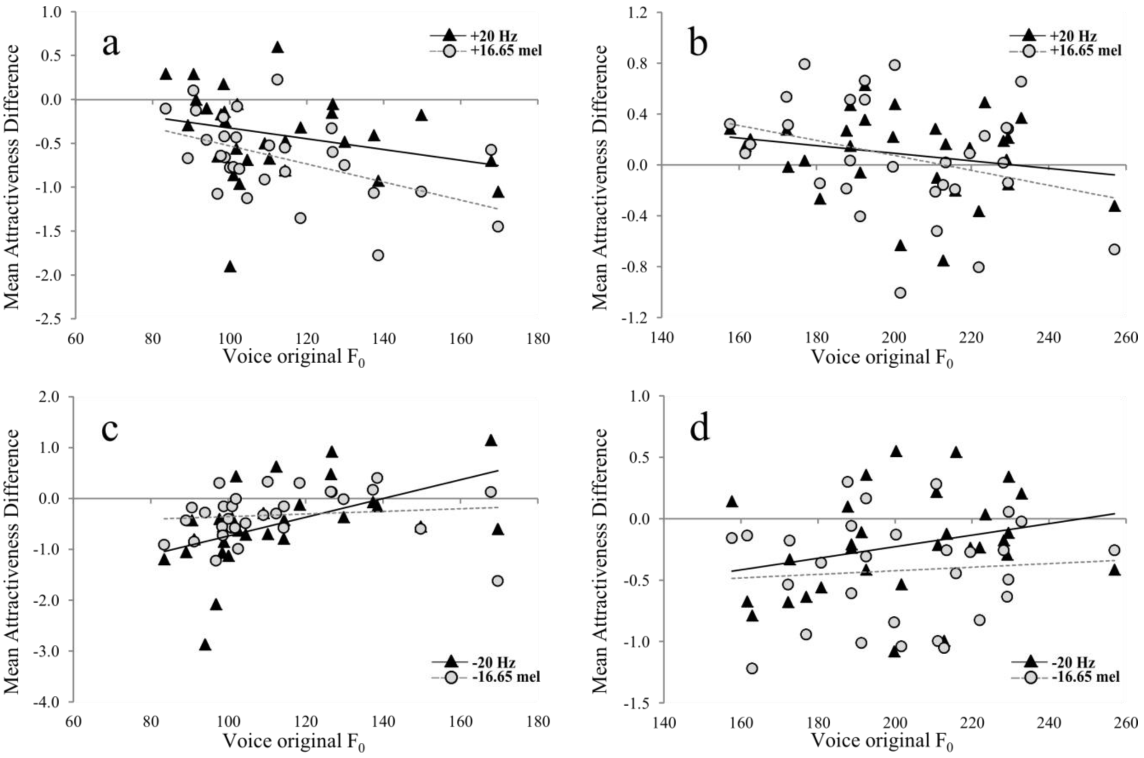


Figure 6.4. Correlations between rating differences of attractiveness (manipulation – original) and original F_0. Markers represent rating differences (black triangles = manipulations in Hz; grey circles = manipulations in mel). Lines represent correlations between rating differences and F_0 of the original recordings (black solid line = manipulations in Hz; grey dashed line = manipulations in mel). (a) Increased F_0 (+20 Hz, +16.65 mel) for men's voices; (b) increased F_0 (+20 Hz, +16.65 mel) for women's voices; (c) lowered F_0 (-20 Hz, -16.65 mel) for men's voices; (d)) lowered F_0 (-20 Hz, -16.65 mel) for women's voices.

The correlation between rating differences and original F_0 for manipulations increasing F_0 for men's voices (Fig. 6.4a) was significant in the case of the mel manipulation (r = -0.50, n = 30, p = 0.002), and showed a non-significant trend in the case of Hz manipulations (r = -0.27, n = 30, p = 0.07), which was negative – as opposed to the predicted positive correlation. Furthermore, the correlations for Hz and mel manipulations were not significantly different (z = 0.99, p = 0.16).

For manipulations of men's voices with lowered F_0 (Fig. 6.4c), the analysis revealed that while no correlation was found for mel manipulations (r = 0.12, n = 30, p

= 0.27), a significant correlation for Hz manipulations (*r* = 0.51, n = 30, *p* = 0.002), and that these two correlations were marginally significantly different (*z* = 1.63, *p* = 0.05).

In voices of women, manipulations increasing F_0 resulted in negative trends (Fig. 6.4b): the correlation for manipulations in Hz was not significant (*r* = -0.23, n = 30, *p* = 0.11), but the correlation for mel manipulations was (*r* = -0.32, n = 30, *p* = 0.04). Furthermore, these correlations did not differ significantly (*z* = 0.35, *p* = 0.36). For manipulations lowering F_0 (Fig. 6.4d), no significant correlations were found, but the manipulation in Hz showed a trend in the predicted direction (Hz: *r* = 0.28, n = 30, *p* = 0.07; mel: *r* = 0.13, n = 30, *p* = 0.25), although manipulations in Hz and mel were not significantly different from one another (*z* = 0.57, *p* = 0.28).

6.3.2 Masculinity and Femininity

Manipulations increasing F_0 tended to lower the perception of masculinity in men's voices (Fig. 6.5a) in both Hz (*M* = -0.36, *SD* = 0.52; one-sample *t*-test: t_{29} = -3.81, *p* = 0.001) and mel (*M* = -0.53, *SD* = 0.56; one-sample *t*-test: t_{29} = -5.19, *p* < 0.001), and increase perceived femininity in women's voices (Fig. 6.5b) in Hz (*M* = 0.29, *SD* = 0.32; one-sample *t*-test: t_{29} = 4.98, *p* < 0.001) and mel (*M* = 0.20, *SD* = 0.24; one-sample *t*-test: t_{29} = 4.73, *p* < 0.001). Conversely, manipulations lowering F_0 tended to increase perceived masculinity in men's voices (Fig. 6.5c) in Hz (*M* = 0.50, *SD* = 0.50; one-sample *t*-test: t_{29} = 5.56, *p* < 0.001) and mel (*M* = 0.45, *SD* = 0.39; one-sample *t*-test: t_{29} = 6.31, *p* < 0.001), but decrease perceived femininity in voices of women (Fig. 6.5d) both in Hz (*M* = -0.32, *SD* = 0.33; one-sample *t*-test: t_{29} = -5.43, *p* <

0.001) and mel (M = -0.39, SD = 0.41; one-sample t-test: t_{29} = -4.85, p < 0.001). These tendencies are consistent with those found in previous research.

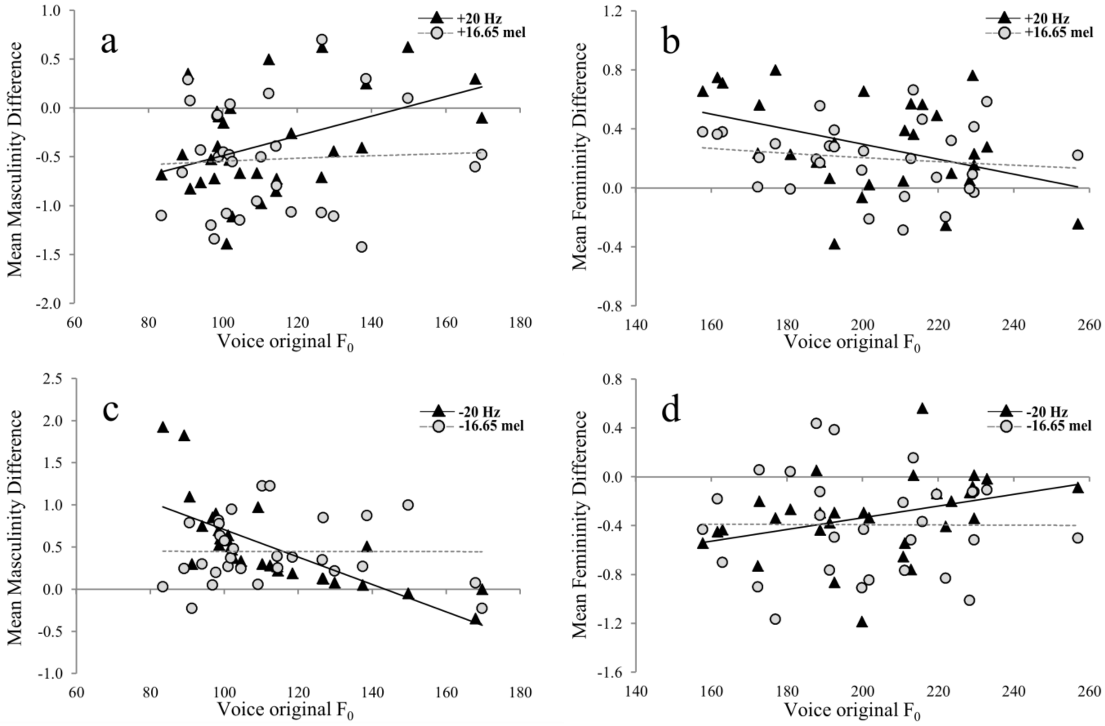


Figure 6.5. Correlations between rating differences of masculinity/femininity (manipulation – original) and original F_0. Markers represent rating differences (black triangles = manipulations in Hz; grey circles = manipulations in mel). Lines represent correlations between rating differences and F_0 of the original recordings (black solid line = manipulations in Hz; grey dashed line = manipulations in mel). (a) Increased F_0 (+20 Hz, +16.65 mel) for men’s voices; (b) increased F_0 (+20 Hz, +16.65 mel) for women’s voices; (c) lowered F_0 (-20 Hz, -16.65 mel) for men’s voices; (d)) lowered F_0 (-20 Hz, -16.65 mel) for women’s voices.

In voices of men, manipulations of F_0 resulted in significant correlations when they were done using Hz, but not when they were performed using the mel scale (Fig. 6.5a, c): the correlation for manipulations increasing F_0 (Fig. 6.5a) was significant when performed in Hz (r = 0.43, n = 30, p = 0.009), but not in mel (r = 0.55, n = 30, p = 0.39), and the comparison between these correlations revealed that their difference was close to significance ($z = 1.49$, $p = 0.07$). Similarly, manipulations lowering F_0 (Fig. 6.5c) resulted in a significant negative correlation when the manipulation was performed using Hz ($r = -0.73$, n = 30, $p < 0.001$), but no correlation was found for manipulations in mel ($r = -0.05$, n = 30, $p = 0.40$). In this last case, however, the two correlations differed significantly ($z = -3.2$, $p < 0.001$).

For women's voices, manipulations affected perceptions of femininity in the predicted direction (Fig. 6.5b,d): increasing F_0 (Fig. 6.5b) resulted in a significant negative correlation when the manipulation was performed using Hz ($r = -0.40$, n = 30, $p = 0.02$), but not when performed using mel ($r = -0.14$, n = 30, $p = 0.24$), although these two correlations did not differ significantly ($z = -1.0$, $p = 0.15$). Likewise, lowering F_0 (Fig. 6.5d) resulted in a significant correlation for manipulations in Hz ($r = 0.36$, n = 30, $p = 0.03$), but not for manipulations in mel ($r = 0.03$, n = 30, $p = 0.43$), but again these correlations did not differ significantly ($z = 1.3$, $p = 0.10$).

6.3.3 Body Size

As expected, increasing F_0 tended to decrease perceptions of body size (Fig. 6.6a), either in Hz ($M = -0.28$, $SD = 0.39$; one-sample *t*-test: $t_{59} = -5.57$, $p < 0.001$) or mel ($M = -0.39$, $SD = 0.40$; one-sample *t*-test: $t_{59} = -7.57$, $p < 0.001$) , while lowering tended to results in perceptions of larger body size (Fig. 6.6b) in both Hz ($M = 0.23$, $SD = 0.37$; one-sample *t*-test: $t_{59} = 4.77$, $p < 0.001$) and mel ($M = 0.30$, $SD = 0.33$; one-sample *t*-test: $t_{57} = 4.99$, $p < 0.001$).

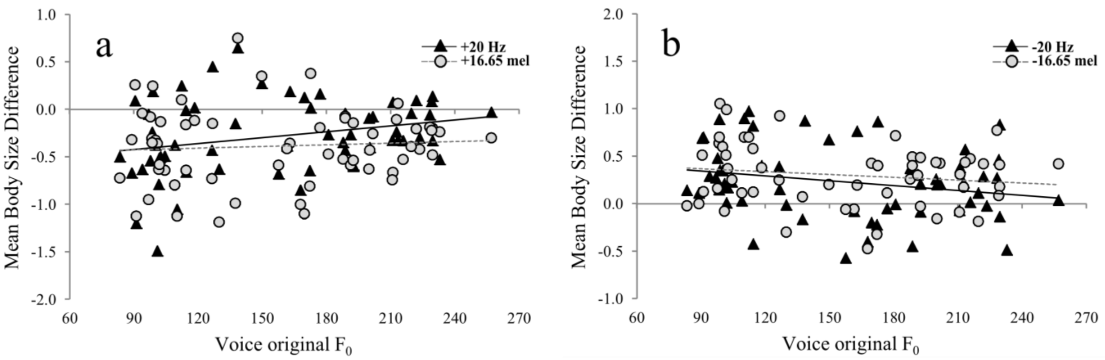


Figure 6.6. Correlations between rating differences of body size (manipulation – original) and original F_0. Markers represent rating differences (black triangles = manipulations in Hz; grey circles = manipulations in mel). Lines represent correlations between rating differences and F_0 of the original recordings (black solid line = manipulations in Hz; grey dashed line = manipulations in mel). (a) Increased F_0 (+20 Hz, +16.65 mel) for all voices; (b) lowered F_0 (-20 Hz, -16.65 mel) for all voices.

Increasing F_0 (Fig. 6.6a) resulted in a significant positive correlation when the manipulation was performed using Hz (r = 0.27, n = 60, p = 0.02), but not when performed using mel (r = 0.07, n = 60, p = 0.30), although these two correlations did not differ significantly (z = 1.1, p = 0.14). Lowering F_0 (Fig. 6.6b) resulted in a significant correlation for manipulations in Hz (r = -0.24, n = 60, p = 0.04), but not for manipulations in mel (r = 0.003, n = 58, p = 0.50), but again these correlations did not differ significantly (z = -1.3, p = 0.10). However, when splitting by sex, correlations were not significant for voices of men or women: for men's voices, increasing F_0 in Hz resulted in a non-significant trend (r = 0.32, n = 30, p = 0.09), but that was not the case when the manipulation was performed using mel, (r = -0.13, n = 30, p = 0.48), nor when F_0 was lowered using either Hz (r = -0.27, n = 30, p = 0.14) or mel (r = -0.24, n = 30, p = 0.21); in women's voices, neither increasing F_0 (Hz: r = 0.25, n = 30, p = 0.18; mel: r = 0.09, n = 30, p = 0.63) nor lowering it (Hz: r = 0.07, n = 30, p = 0.73; mel: r = 0.18, n = 30, p = 0.34), resulted in significant correlations.

6.3.4 Age

Increasing F_0 (Fig. 6.7a), tended to decrease perceived age when the manipulation was performed in both Hz (*M* = -0.49, *SD* = 1.16; one-sample *t*-test: t_{59} = -3.34, *p* = 0.001) and mel (*M* = -0.86, *SD* = 0.95; one-sample *t*-test: t_{59} = -6.93, *p* < 0.001). Conversely, lowering F_0 showed a clear tendency to increase perceived age (Fig. 6.7b) in both Hz (*M* = 1.97, *SD* = 1.79; one-sample *t*-test: t_{59} = 8.28, *p* < 0.001) and mel (*M* = 2.82, *SD* = 1.88; one-sample *t*-test: t_{59} = 10.77, *p* < 0.001).

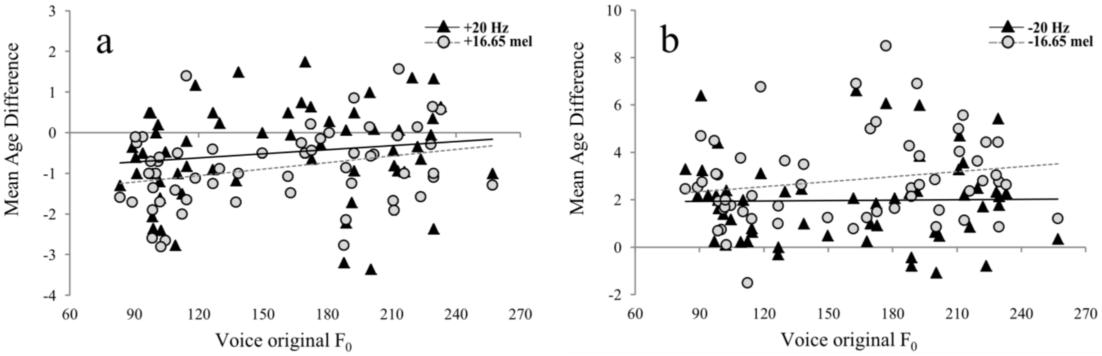


Figure 6.7. Correlations between rating differences of age (manipulation – original) and original F_0. Markers represent rating differences (black triangles = manipulations in Hz; grey circles = manipulations in mel). Lines represent correlations between rating differences and F_0 of the original recordings (black solid line = manipulations in Hz; grey dashed line = manipulations in mel). (a) Increased F_0 (+20 Hz, +16.65 mel) for all voices; (b) lowered F_0 (-20 Hz, -16.65 mel) for all voices.

F_0 manipulations performed using the mel scale produced positive correlations, both when F_0 was increased (*r* = 0.30, n = 60, *p* = 0.01; Fig. 6.7a), or lowered (*r* = 0.22, n = 60, *p* = 0.04; Fig. 6.7b), but manipulations using Hz did not (raised F_0: *r* = 0.15, n = 60, *p* = 0.13; Fig. 7a. Lowered F_0: *r* = 0.02, n = 60, *p* = 0.45; Fig. 6.7b). Correlations between manipulations in Hz and mel did not differ significantly when F_0 was raised (*z* = -0.9, *p* = 0.19), nor lowered (*z* = -1.1, *p* = 0.13).

6.4 Discussion

There is a robust body of knowledge in terms of the effects of manipulating F_0 in perceptions of attractiveness, masculinity/femininity, body size, and age. Except in the case of how F_0 manipulations affected attractiveness, the results obtained here are consistent with previous research. I found that lowering F_0 tended to increase the perceived masculinity of men's voices and decrease the perceived femininity of women's voices, while increasing F_0 tended to decrease masculinity and increase femininity. In both men and women, lowering F_0 tended to increase perceived age and body size, and decrease it when the manipulations increased F_0.

Ratings of attractiveness, however, were not consistent with my predictions: both increasing and lowering F_0 showed a tendency to negatively affect attractiveness for men's voices and, while lowering F_0 in women's voices showed a tendency for participants to perceive the voices as produced by less attractive women (as expected), increasing F_0 did not consistently increased their perceived attractiveness. At least in the case of women, studies measuring vocal responses to attractive targets have found inconsistent results; for example, in one study women were found to increase voice pitch when directing speech towards attractive faces (Fraccaro et al., 2011), while another study found that both men and women lowered their voice pitch when speaking attractive targets of the opposite sex (Hughes et al., 2010). This suggests that while F_0 can have an effect on perceived attractiveness, particularly as it can emphasise sex-specific vocal characteristics (i.e. signalling masculinity or femininity), the relationship between mean F_0 and vocal attractiveness is, at least, partially mediated by other factors such as perceived masculinity/femininity and body size, as well as other vocal parameters such as formant dispersion D_f and breathiness (see Xu et al., 2013).

In addition, increasing F_0 in men's tended to decrease the perceived attractiveness of the voices, which is consistent with previous findings (e.g. Feinberg, Jones, Little, et al., 2005), but surprisingly the same effect was found when the voices were lowered. This could be an effect of increased perceived masculinity as well as body size, because low-pitched masculine voices can be associated with negative attributions (Hodges-Simeon et al., 2010), and both large body size (Xu et al., 2013) and low-pitched voices (Puts et al., 2012), might be associated with aggression.

Regarding the effects of manipulating F_0 using a mechanical scale of frequency (Hz), versus a psychoacoustic scale (mel), I expected that (1) manipulations in hertz would have a larger perceptual effect on relatively low voices in comparison to relatively high voices, and (2) that manipulations in mel would tend to have a relatively constant perceptual effect on all voices, regardless of their original F_0. This would tend to create significant correlations for manipulations in Hz that approached a perceptual difference of 0 (between the ratings given to manipulations their corresponding original recordings) for original voices increasing in F_0. I found these predicted effects only in the case of ratings of femininity/masculinity, and partially in ratings of body size.

In my results, differences in attractiveness (Fig. 6.4), and especially in perceived age (Fig. 6.7), are scattered, suggesting that manipulations of F_0 did not affect voices consistently. This could be because, as discussed earlier, attractiveness perceptions are not directly related to F_0, and the same might be true for perception of age; several vocal parameters (e.g. speaking rate, F_0 instability) have an important effect on age perceptions (Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008; Linville, 1996) and, furthermore, changes in F_0 alone might not be a reliable indication of age (Harnsberger et al., 2008).

As predicted, in the case of perceptions of body size, manipulations in Hz resulted in the F_0 of the original recordings correlating with rating differences, and manipulations in mel did not. However, this was only the case when the ratings for both men's and women's voices were analysed together; because of the sexual dimorphism in body size (Ruff, 1994, 2000) and voices (Titze, 1994), these correlations seem to be a product of between-sex differences.

In the case of femininity and especially masculinity ratings, I found the predicted effects for manipulations in Hz as well as mel. Unlike attractiveness, body size or age, masculinity and femininity perceptions are directly and independently associated with F_0: voice fundamental frequency is related to testosterone levels in men (Dabbs & Mallinger, 1999; S. Evans, Neave, Wakelin, & Hamilton, 2008; see also Ferdenzi, Lemaître, Leongómez, & Roberts, 2011), and most likely related to oestrogen levels in women (Abitbol, Abitbol, & Abitbol, 1999; Feinberg, Jones, DeBruine, et al., 2005).

In conclusion, my results suggest that while the general effects of F_0 changes in perceptions of attractiveness, masculinity/femininity, body size, and age, are well established and robust, the use of particular scales in studies that manipulate frequencies can affect the results, especially when testing changes in directly associated perceptions (i.e. masculinity and femininity). Manipulations using Hz tend to affect relatively lower voices more than relatively higher voices in terms of masculinity and femininity, even within the normal F_0 range during speech, but manipulations using the mel scale seem to help correcting this bias. Because manipulation studies rely on perceptual differences, the use of psycho-acoustic scales is essential, especially for studies that measure more subtle or more intricate perceptual changes. Given the variety and challenging nature of psycho-acoustic scales, the selection of a particular one is, however, complex; future studies should compare the biases that specific psychoacoustic scales (e.g. semitones, Bark, ERB-rate, mel) and their variants have, as well as the differences between them.

CHAPTER 7:

GENERAL DISCUSSION – A MODEL FOR THE EVOLUTION OF MUSICALITY AND ITS ROLE IN HUMAN VOCAL COMMUNICATION

7.1 Introduction

Musicologists have commonly rejected the idea of music as a universal phenomenon (e.g. Bohlman, 1999; Tomlinson, 1984; see also Cross, 2003), and therefore the study of its origins has not been often addressed within this discipline. In contrast, scientists from disciplines such as biology, psychology, and anthropology have addressed this problem more commonly, often focusing on the purpose of music, and its potential evolutionary function. Evolutionary ideas have ranged from purely adaptationist theories such as sexual or group selection (e.g. Brown, 2000a; Charlton, Filippi, & Fitch, 2012; Charlton, 2014; Miller, 2000), to explanations of music as a spandrel (Pinker, 1997). This essential difference, which perhaps exists because ethnomusicologists usually look at cultural differences and focus on the specificity of particular musical manifestations, while other scientists look at universals in music (i.e. looking at music as a universal, human phenomenon), has often prevented communication between these complementary views.

While it seems undeniable that all cultures have some manifestations that can be recognised as music (Blacking, 1995), ethnomusicology highlights an important element that should be considered: variation in the social roles of musical manifestation is enormous. This makes the scientific study of music (as a human universal), and its origins, an extremely difficult task.

Scientists, however, have often focused on *music*, which is a behavioural manifestation –the outcome of any potential adaptations– rather than the adaptations themselves. In other words, it may be less pertinent to examine *music* than *musicality*, our ability to process musical information. Additionally, musicality consists of different separable mechanisms for perception and production that may have evolved independently (Fitch, 2006b). Furthermore, because most theories for the origin of music point to an evolutionary connection between music and language (see section 7.2.2.2 below), the domain of musicality might be not limited to music, but might also play a role infant directed speech (IDS) (Falk, 2004) and perhaps even adult language.

The main body of this thesis has a particular emphasis on contextual vocal modulation, and on its perception, showing how variability in fundamental frequency (F_0 SD) plays an important role in in courtship contexts (Chapters 2 and 3) possibly signalling relationship satisfaction and parental investment potential (Chapter 4), and that the important role of F_0 SD is not apparent in non-courtship scenarios (Chapter 2, study 2) nor authority ranking relationships (Chapter 5). This could be an indication of the role that musicality (or at least an analogous capacity) plays in human communication in specific contexts, and can point to theoretical mechanism that could have partially shaped human musicality and, ultimately, human music. This chapter discusses the problems that the study of the origins of music poses and, based on the results of the empirical chapters of this thesis, as well as the current body of knowledge about this area, proposes a theoretical model for the evolution of musicality and its influence on different modalities of human acoustic communication.

7.1.1 The puzzling study of the origins of human acoustic communication

Forms of acoustic communication with various degrees of complexity are common amongst many animals, but particularly sophisticated (and well-studied) in some bird and mammal species. These studies have been, at least partially, an attempt to inform the staggering complexity of human acoustic communication, and particularly its puzzling evolutionary origins.

There are clear differences in form and complexity between human and non-human forms of vocal communication, particularly in terms of the syntactic and semantic aspects of language, which make the understanding of the evolutionary development of human acoustic communication a huge task. For example, as Falk (2004) asks, “why are we the only animals that talk?”. As pointed out by Brown and Jordania (2011), among more than 4,500 singing species living today, we are the only species with “the ability to follow precise rhythmic patterns so as to permit group singing, drumming, and dancing. What explains the unique place of humans among singing species?”

Amid the many problems that make this matter so complex, is the lack of clear intermediate stages that could allow us to create an accurate picture of the evolutionary process that lead to modern human forms of human acoustic communication. There are no other extant hominin species with varying degrees of acoustic communicative complexity, with which we could compare ourselves. And, as the fossil record does not allow to directly study the acoustic communication of extinct species, we are forced to rely on indirect inferences based on archaeological findings, observations of modern animal species (including humans), or theoretical dissertations. For instance, even in the case of Neanderthals, *Homo neanderthalensis*, probably the most studied extinct human species, and whose cognitive capacities might have been significantly underestimated (see Tanabe, Kochiyama, Ogihara, & Sadato, 2014; Villa & Roebroeks, 2014), there seem to be no clear conclusions about their level of acoustic communication and symbolism (see Mellars, 2010).

If the conditions for the study of human acoustic communication are complex in the case of language, the situation is not better in the case of music. Darwin himself (1871) portrayed music as one of the most mysterious human abilities, because it is a human universal with no obvious function. It is, in fact, a phenomenon that seems to be present in all human cultures (Higgins, 2012), whose roots can be undoubtedly traced for a few tens of thousands of years, back to the earliest known musical instruments. As mentioned in the introduction to this thesis, music does not depend on the construction of instruments to exist, as we can sing, dance, and use our bodies as drums. However, the earliest known flute –a complex and differentiated instrument– was made around 40.000 years ago (e.g. Conard, Malina, & Münzel, 2009; Hahn & Münzel, 1995; see Adler, 2009), and so we can infer that musical ability significantly pre-dates this time.

7.1.2 The question of music universals

The universality of human music is critical to the study of music; if music is in fact a universal phenomenon, the idea of it having a purely cultural origin would be hard to maintain, suggesting instead some form of biological basis.

This is not to say that culture is unimportant. Unquestionably, cultural distinctiveness and variation define individual features, social roles and conceptions of music. As Cross (2006) points out, within the humanities, including main trends in musicology and ethnomusicology, there is a somehow consensual view stating that music is a cultural construction. This view is supported precisely by the enormous cultural variation of musical or music-like phenomena in human societies. Moreover, the notion of *music* itself varies significantly between different cultures. For example, Australian Aboriginal songs combine visual, performing and oral arts (Ellis, 1984), and the Igbo concept of *nkwa* includes, not only actions like singing and playing instruments, but also dancing (Gourlay, 1984). In fact, many academics prefer to use the term *musics* instead of *music*, to account for the uniqueness of these phenomena within each culture (see Cross, 2003).

This idea of *musics*, as different, particular cultural expressions lacking relevant commonalities, which are only valid within the context of a particular human group (Bohlman, 1999; see also Cross, 2003), is essential to understand the limitations of the study of music, and how can we address it. If, as some suggest (and some evidence seems to support), there are no universals in *musics*, no common basic principles that allow measurements and comparisons to be made, then the scientific study of music as one universal, human phenomenon would be irrelevant, and perhaps even implausible. In this paradigm, music –each specific manifestation– can only be described within its cultural context and no generalisations can be made (e.g. Tomlinson, 1984). However, this view has been by no means common to all researchers within the humanities (e.g. Blacking, 1995; Brown & Jordania, 2011; Nettl, 2000); Blacking (1995), for example, stated that every society has some cultural manifestations that can be recognised as music, suggesting that there are, in fact, common features.

In the light of the immense cultural variability it seems difficult to agree about a definition, in a general sense, of what music is, and especially *what is it for*.

However, in spite of these obstacles, we know that our brain, physiology and psychology make us capable of producing and listening to music. In other words, the capacity to process musical information, *musicality*, is universal. This seems to be a major problem for exclusively cultural explanations of music: how can they fully explain the universality of musicality.

For some decades, scientist from disciplines as diverse as biology, psychology, neuroscience or psychiatry have been presenting data that speak to us of a more primal, biological basis of musicality, common to *all* humans. And, furthermore, science has provided an insight on the cognitive demands of the musical capacity. We *all* share the amazing capacity to produce, perceive and enjoy –or dislike– music, probably since, or soon after, we are born (e.g. Papoušek, 1996; Peretz & Hyde, 2003), and music has a significant capacity to affect our emotions (e.g. Husain, Thompson, & Schellenberg, 2002; Juslin & Sloboda, 2001).

In fact, for years many scientists have proposed a variety of music universals. For example, Fritz et al. (2009) found that adult Mafa were successful in identifying three basic emotions (happy, sad, scared/fearful) in Western music, at above chance levels. Western as well as Mafa participants also preferred original versions of both

Western and Mafa music over spectral manipulations of the originals (that affected the sensory dissonance of the music), suggesting that basic emotions of music can be universally recognised, regardless of its cultural origin, and that the perception of pleasantness in music is universally affected by consonance and dissonance. Trehub (2000), analysing the perception of human infants and adults in original and transposed melodies, proposed that the perception of contours (i.e. relational pitch and time features of music) as universal, as well as scales composed of unequal steps, and a preference for small integer frequency ratios (i.e. consonances) like the octave (2:1), perfect fifth (3:2), and perfect fourth (4:3), versus large integer ratios (dissonances) such as the tritone (45:32) across cultures. In addition, Trehub (2000) suggested the universality of a music genre for infants (e.g. lullabies and play songs); in fact, adults are able to recognise a lullaby as such, even when they are unfamiliar with the musical culture, and can identify with almost absolute precision when a song was sang to an infant (see also Trehub, Unyk, & Trainor, 1993).

Furthermore, Brown and Jordania (2011) have recently proposed an extensive list of music universals, categorised in four types: (1) *Conserved universals* which apply to all musical utterances, and include pseudo-syntactic elements such as the fact that music is organised into phrases, relative pitch elements such as the equivalence of octaves (and consequent transposability of melodies) and the use of discreet pitches, as well as factors used for emotive expression, such as register, tempo, and amplitude. (2) *Predominant patterns* which apply to all musical styles, including rhythmic features such as the predominance of isometric rhythms, the use of scales divided into seven or less pitches, the use of motives, and use of texts, among others. (3) *Common patterns*, which apply to many styles, and include, for example, the association of music and dance, and the use of aerophone instruments (wind instruments). (4) *Range universals*, which contains a set of possible options for all musical systems, such as textures (monophony, heterophony, homophony, or polyphony), and type of arrangement (solo or group arrangements).

Musicality, however, might seem like a relatively modest faculty: having the ability to process musical information, regardless of one's playing skills, might seem exclusively dependent on the general sense of hearing. Yet, research from different fields has shown that, even the perception of music, is a far more intricate process, possibly related in complex ways to, or at least analogous to, language. Interestingly, some of the cues to understand the real extent of the cognitive demands of musical perception and processing, and its universality, come precisely from the exceptions: from the singular cases in which a person cannot process musical information, as discussed in the next section.

7.1.3 Music/language relationship

The deep relationship between language and music in terms of shared neural resources, is supported by evidence presented in a variety of studies (e.g. Koelsch, Fritz, et al., 2005; Patel, Peretz, Tramo, & Labreque, 1998; Patel, 2003; Schön et al., 2004), and has become an important area of research and source of debate in recent years. There is an increasing number of studies showing an important overlap of neural resources involved in the processing of specific music and language tasks (e.g.

Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Koelsch, Gunter, et al., 2005; Sammler et al., 2009).

For instance, strong evidence for shared resources in musical and linguistic syntactic processing has been presented in several studies (e.g. Koelsch, Fritz, et al., 2005; Sammler et al., 2009). Moreover, children who suffer from Specific Language Impairment (SLI), which is characterized by deficient processing of linguistic syntax, also show a deficiency of musical syntax processing (Jentschke, Koelsch, Sallat, & Friederici, 2008). There is even evidence suggesting that the human brain does not treat language and music as different kinds of stimuli, at least on early stages of ontogenetic development (see Koelsch & Siebel, 2005) and music therapy (based on singing) has been used in speech rehabilitation (e.g. Racette, Bard, & Peretz, 2006; Skeie, Einbu, & Aarli, 2010; cf. Stahl, Kotz, Henseler, Turner, & Geyer, 2011; see also Hurkmans et al., 2012).

As in language, music processing involves networks of extensively distributed brain regions. In fact, compared to language, music might even comprise a vaster network of regions, from both hemispheres, and with an overall asymmetry towards the right hemisphere for pitch processing (Peretz & Zatorre, 2005; see also Peretz, 2009). Hence, the overlap between the activated neural areas for music and language processing that has been found in several neuroimaging studies –especially clear in production tasks that involve singing with lyrics– is not surprising.

Indeed, Peretz (2009) points out that in this context –in which overlapping of involved neural resources is expected– finding distinct areas of activation for music and language (particularly singing and speaking) can be more enlightening than describing overlaps; several studies (i.e. Brown, Martinez, & Parsons, 2006; D. E. Callan et al., 2006; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Jeffries, J. B. Fritz, & Braun, 2003; Ozdemir, Norton, & Schlaug, 2006; Saito, Ishii, Yagi, Tatsumi, & Mizusawa, 2006) have described, in addition to the expected overlapping, activation of distinct areas for speech and song production.

Furthermore, evidence of domain–specificity of music and language processing becomes apparent from the study of specific cases of brain damage or developmental disorders (Peretz, 2009; see also Sacks, 2007), in which patients might lose musical abilities while maintaining their speaking capacity, like some amusic patients (e.g. Pearce, 2005; Hyde & Peretz, 2004; Peretz, 2009; Peretz & Hyde, 2003; Sacks, 2007), or vice versa: when patients can sing or play music, but can no longer speak, as in the case of some aphasias (e.g. Signoret, van Eeckhout, Poncet, & Castaigne, 1987; Yamadori, Osumi, Masuhara, & Okubo, 1977).

What does this deep relationship tell us about the origins of music and language? Is it possible to think that both channels have common origins? Some evidence seems to suggest that this is precisely the case; for example, Alcock et al. (2000) found that the FOXP2 gene –which plays a crucial role in the neural development necessary for language and speech– seems to affect rhythm perception and production, while not affecting pitch perceptual and production skills (which seem to be affected by independent genetic factors as congenital amusia shows (Hyde & Peretz, 2004)). Furthermore, performance in detecting out–of–key notes in popular melodies showed a stronger correlation between identical (r = 0.79) than fraternal (r =

0.46) twins, suggesting that genetic influence –with a heritability of 70–80%– is more important than shared environments for musical pitch perception (Drayna, Manichaikul, De Lange, Snieder, & Spector, 2001).

The findings regarding similarities and differences found in the processing of music and language have led to an interesting consideration. While Peretz (2006, 2009), based on a variety of data, supports the idea of more complex and specialised cognitive processing requirements than previously thought, and even modularity, pointing out to a biological basis of musicality and some form of natural selection, Patel (2010), argues that universality and processing specialisation can be explained without evolutionary adaptation. Patel gives the example of the ability to make fire, which, although an invention, "extends deep into our species' past and is found in every human culture" and "provides things that are universally valued by humans, including the ability to cook food, keep warm, and see in dark places" (p. 46). He also highlights the example of reading and writing –both cultural inventions– which are each partially associated with functional specializations in specific brain regions (product of neural plasticity) and, as in the case of reading, some disorders are driven by genetic causes (pp. 46-47). Patel, however, seems overlook two important elements: that musicality, unlike making fire, reading and writing, or even music, is not a behaviour *per se* but an *ability* that seems not to be taught and learned, and furthermore, that it appears to be present during early infancy (e.g. Trehub, 2001, 2003; Trehub & Hannon, 2006; see also Bencivelli, 2011). Thus, the question of whether music is an adaptation could in itself be a dead end (see Fitch, 2006a).

7.2 The evolutionary study of musicality

Over more than two decades, researchers have focused their attention towards the evolution of music, producing a great variety of evolutionary theories that range from Pinker's controversial description of music as an "evolutionary cheesecake" (Pinker, 1997), to purely adaptationist views (e.g. Brown, 2000; Miller, 2000). Because these theories have been reviewed and discussed elsewhere (e.g. Cross & Morley, 2008; Fitch, 2006; Perlovsky, 2010), I will not examine them in depth. Instead, this section is divided in three segments addressing: (1) issues in the evolutionary study of musicality, (2) some major ideas in the evolutionary theories of music, and (3) a theoretical model for the evolution of musicality.

7.2.1 *Difficulties of the evolutionary study of musicality*

Besides the fact that music does not seem to play an obvious direct role of biological relevance, the evolutionary study of musicality has to face the problem that it is likely to consist of different, relatively independent components. Strong evidence for this can be found in the cases in which a disorder affects either pitch or rhythm processing, but not both (Alcock, Passingham, et al., 2000; Alcock, Wade, Anslow, & Passingham, 2000; Di Pietro, Laganaro, Leemann, & Schnider, 2004; for a review, see Peretz, 2009), indicating the independence of these modules. This means, as Fitch (2006b) points out, that different components of musicality might have followed independent evolutionary paths; in his words:

> Although all of the mechanisms involved in music perception and production may be grouped together, for convenience, as “the music faculty” or “the capacity for music”, it is important to remember that different components of this capacity may have different evolutionary histories. Thus, discussing “Music” as an undifferentiated whole, or as a unitary cognitive “module”, risks overlooking the fact that music integrates a wide variety of domains (cognitive, emotional, perceptual, motor,...), may serve a variety of functions (mother-infant bonding, mate choice, group cohesion...) and may share key components with other systems like language or speech. Thus, questions like “When did music evolve?” or “What is music for?” seem unlikely to have simple unitary answers (Fitch, 2006b, p. 174).

In addition, Justus & Hutsler (2005) highlight the fact that the evolutionary study of music might have been somewhat biased, favouring explanations based on natural selection over those involving cultural transmission. This is because a majority of the recent abundance of studies of the origins on music have been based on the approach of evolutionary psychology. This approach has required researchers to define criteria to assess whether music emerged as an adaptation (i.e. limited by innate factors, domain-specificity, and conferring survival or reproductive advantages), or as an exaptation (Justus & Hutsler, 2005; McDermott & Hauser, 2005; see also Trainor, 2006). In essence, the central question of adaptationist views is to know if musicality (or, more specifically, its cognitive components) exists because it holds intrinsic biological benefits, or if it has simply acted as a *parasite* and has exploited cognitive abilities that evolved for a different purpose.

The problem, however, is even more complex, as music (and musicality, being a higher-level cognitive domain such as language) probably involves both exaptations and adaptations, making the limits between adaptation and exaptation quite vague (Justus & Hutsler, 2005; Trainor, 2006).

To eventually obtain a complete picture of the evolution of music, both biological (e.g. cognition, mother-infant interactions) and cultural (e.g. learned aesthetic preferences) aspects should be considered. However, in any musical manifestation or its perception, both are so intimately connected that finding the differences is problematic. To overcome this issue, one option is to study infants (assuming them as individuals who have not been, or have only partially been, culturally "contaminated"), comparing them to adults, whose musical abilities are greatly affected by their cultural environment, to see what is innate; this paradigm has provided, and will continue to provide, important answers (e.g. Trehub, 2001, 2003; Trehub & Hannon, 2006; see also Bencivelli, 2011). However, it could intrinsically favour hypotheses related to the evolution of musicality from a parent-infant perspective.

7.2.2 Key ideas in evolutionary theories of music

Evolutionary theories of music are in many cases linked to those of language. There are, at least, two main stages that can be discussed separately: (1) the link between animal precursors and human music and language channels, and (2) the human evolutionary psychology of acoustic communication (including music and language). While both look at the question of origins, the main difference between these two stages is that the first compares modern animal species and their acoustic communication to human language and music, while the second attempts to explain the human evolution of these channels.

7.2.2.1 Animal precursors

Non-human vocal communication has been compared to both language and music. In fact, vocalisations from many species are often called songs, because of their complexity and because they are learned (Fitch, 2006b). However, with the exception of gibbons, these complex song-like vocalisations occur only in birds and non-primate mammals such as cetaceans, suggesting that they do not share a common evolutionary path with music, or any other learned, complex human acoustic signals like language. There are, however, certain similarities and potential instances of convergent evolution that can provide models for the evolution of human acoustic communication (see Fitch, 2005, 2006b). Vocal learning, for example, seems to work in an analogous way in songbirds and humans. In fact, bird brain areas involved in vocal learning have been compared to Broca and Wernicke regions of the human brain, as they activate when a bird hears and sings a song, respectively (Balter, 2010; see also Jarvis, 2004).

Furthermore, there are interesting parallels between human music and language with vocal signals of other animal species, particularly in instances where animal vocalisations have semantic- and syntax-like elements (for a review, see Marler, 2000). Interesting examples of semantic-like elements (i.e. calls that have symbolic functions) come from chickens; Marler and his collaborators, for example, showed that in domestic chickens, *Gallus domesticus,* food calls produced by males are dependent on the quality of the food, and that females respond selectively to these calls (Marler, Dufty, & Pickert, 1986a; see also C. S. Evans & Marler, 1994); furthermore, males are sensitive to the audience, producing significantly less calls when a rival male is present, than in the presence of females, and males are more likely to produce dishonest calls (i.e. in the absence of food) when females were far away than when they were nearby (Marler, Dufty, & Pickert, 1986b; see also Gyger & Marler, 1988). Similar sensitivity to social contexts and audiences has been shown for alarm calls in red junglefowls**,** *Gallus gallus* (Karakashian, Gyger, & Marler, 1988), which have different calls for different types of predators (C. S. Evans, Evans, & Marler, 1993; Gyger, Marler, & Pickert, 1987).

Perhaps some of the most interesting cases of semantic-like elements in non-human vocal communication, because of phylogenetic proximity with humans, are the instances of vocalisations with some degree of symbolic content in primates. For example vervet monkeys, *Cercopithecus aethiops*, like chickens, have different calls for different predators. The presence of leopards, eagles and pythons is communicated through different calls to which individuals respond differently: run into trees, look up, or look down, respectively (Seyfarth, Cheney, & Marler, 1980). These distinct calls are evidence of effective categorization of other species, which individuals progressively develop with age and experience: infants are more likely to produce alarm calls in response to nonraptor birds than juveniles, which in turn are more likely to produce calls in response to incorrect stimuli than adults (Seyfarth et al., 1980). Vervet monkeys, however, are not the only primate species for which calls with symbolic functions have been documented; other species include ring-tailed lemur, *Lemur catta* (Macedonia, 2010), rhesus macaques, *Macaca mulatta* (Hauser & Marler, 1993a, 1993b), toque macaque, *Macaca sinica* (Dittus, 1984), and chimpanzees, *Pan troglodytes* (Hauser, Teixidor, Fields, & Flaherty, 1993).

Syntax-like elements have been widely studied, and are usually present in species that produce vocalisations that are categorised as songs. Marler (2000) divided syntactic elements into two types: the first is phonological syntax (or phonocoding), which is based on the recombination of individual, small phonetic units lacking meaning (e.g. phonemes in human language) to create sequences (e.g. words), and the second is lexical syntax (or lexicoding), in which sequences are recombined to create strings (e.g. sentences) which have meaning both at the sequence (word) and string (sentence) level. While there are strong differences in complexity between human and non-human examples, some animal vocalisations have structures that are similar, from a general point of view, to those of human language and particularly music, because of the absence of symbolic meaning.

In birds, some species have individual song repertoires with a complexity that exceeds that of non-human primates, and that are based on the recombination of elements (for a review, see Fitch, 2006b); swamp sparrows, *Melospiza georgiana*, for example, have songs that consist of short individual, independent units, which are recombined into different sequences (Marler & Pickert, 1984), and in the winter wren, *Troglodytes hiemalis*, each male individual has a large repertoire consisting of around 20 songs that incorporate and transform sequences of other winter wren songs (Kroodsma, 1980; Kroodsma & Momose, 1991; see also Marler, 2000), in a manner that seems to follow a flexible set of *rules* (Van Horne, 1995).

Amongst mammals, however, the most complex vocal behaviour seems to be that of some cetacean species, and particularly humpback whales, *Megaptera novaeangliae*, which have complex songs with syntactic elements analogous to those of songbirds. The songs of the humpback whale are composed of units (analogous to phonemes), which are combined into phrases (relatively fixed sequences of units), and these into themes, which are a collection of phrases (including repetitions and combinations of phrases), which in turn are mixed to create songs with an average duration of 12 to 15 minutes (Payne, 2000).

Furthermore, these humpback whale songs, and the phrases they consist of, constantly *evolve* over time (Payne, Tyack, & Payne, 1983). This creates diversification between populations, similar to that of language and music in human cultures, a phenomenon that has been extensively documented (e.g. Cerchio, Jacobsen, & Norris, 2001; Eriksen, Miller, Tougaard, & Helweg, 2005; Green, Mercado, Pack, & Herman, 2011; Helweg, Herman, Yamamoto, & Forestell, 1990; Maeda et al., 2000). The extent and rate of these changes seems to be motivated by novelty, as exemplified by the documented replacement of the song of the humpback whales from the Pacific Ocean off the Australian east coast, by the song of the Australian west coast population, after the introduction of a small number of members of the latter population (Noad, Cato, Bryden, Jenner, & Jenner, 2000). In songbirds, similar changes in individual and population preferences (as measured by the percentage of individuals singing a song over time) have been documented (e.g. Luther & Baptista, 2010; for a review, see Podos, Huber, & Taft, 2004).

Although seemingly simpler in nature, an interesting phenomenon of recombination of vocal elements has been documented for primate species; greater spot-nosed monkeys, *Cercopithecus nictitans*, combine two alarm calls to create different call series depending on external events (Arnold & Zuberbühler, 2006, 2008). This example is, however, of particular interest because of the semantic-like properties that these call series acquire by the recombination of alarm calls, thus involving both syntactic- and semantic-like properties in a way that seems to be analogous to language, more than music.

A different, yet interesting, example of a potential animal precursor is entrainment, the synchronization to external rhythms, which is a phenomenon central to rhythm processing (and musicality). Entrainment seems to be present in other species, and have direct implications in areas other than music: it has been reported that some patients with Parkinson's disease who are normally unable to walk, can do so when they synchronise to a musical beat (Sacks, 2007; Thaut, 2005; see also Patel, Iversen, Bregman, & Schulz, 2009b). Entrainment has been experimentally confirmed in at least one individual from another species, a sulphur-crested cockatoo, *Cacatua galerita eleonora* (Patel, Iversen, Bregman, & Schulz, 2009a; see also Patel et al., 2009b), and there is evidence of similar behaviour in other species, mainly parrots, *Psittaciformes* (Hasegawa, Okanoya, Hasegawa, & Seki, 2011; Schachner, Brady, Pepperberg, & Hauser, 2009; see also Fitch, 2013). These examples are consistent with the vocal learning and synchronization hypothesis (Patel, 2006), which proposes that the capacity for complex vocal learning is a requisite for any species to show entrainment. However, evidence of certain levels of entrainment in sea lions, *Zalophus californianus* (Cook, Rouse, Wilson, & Reichmuth, 2013) and chimpanzees, *Pan troglodytes* (Hattori, Tomonaga, & Matsuzawa, 2013), seems to contradict this (see Fitch, 2013).

7.2.2.2 The evolution of human acoustic communication

Non-human animal vocal communication provides examples that contain elements analogous to those of music and/or language. These instances, however, are mostly dispersed over phylogenetically distant animal groups (birds, mammals) and families (e.g. cetaceans, primates), suggesting convergent evolution over common ancestry. There are, however, theories that attempt to explain how human communication evolved. Because this chapter attempts to propose a theoretically viable model for the evolution of musicality and its role in human vocal communication, this section briefly reviews the main theories, with an emphasis on music (or musicality).

Probably the best known theory for the evolution of music, and one that proposes an adaptive function, is that music plays a role in mate choice, and more generally sexual selection (e.g. Darwin, 1871; Miller, 2000). This idea seems plausible, in the light of the role that birdsongs play in mate choice, which seems akin to the prevalence of love songs and serenading in human societies. Fitch (2005), however, highlights that there are no studies showing a positive relation between musical skills and reproductive success or offspring survival. Nonetheless, a recent study has provided the first empirical support for a theory of the evolution of music through sexual selection: women have a preference for composers of more complex music around ovulation, but only when selecting partners (based solely on *their* music) for a short-term relationship (Charlton, 2014; see also Charlton, Filippi, et al., 2012). Similarly, changes in pitch discrimination skills, with an increase when the risk of conception is high (as shown in Chapter 4), could be an indication of a mechanism that affects mate choice and that could have played a role in the evolution of human musicality.

Music, however, is in no way limited to courtship contexts, and this may indicate different evolutionary origins. For example, music seems to play an important role in promoting synchronisation and cooperation, as well as group cohesion and identity (e.g. battle music, national anthems, football chants). Because of these social influences, Brown (2000a) suggests that music may have co-evolved with collective rituals, which could explain the universal association between music and rituals, as well as the rewarding properties of music from a psychological perspective. According to him, based on the capacity of music to promote social cohesion, and because music is overwhelmingly a social phenomenon, the survival value of music is not apparent at an individual, but only at group, level.

A somewhat similar hypothesis, based on the potential role of music in promoting group cohesion, is that language evolved as a form of “vocal grooming”, to maintain social bonds in increasingly large groups (Dunbar, 1996, 2003a, 2010). In fact, when phylogenetically controlled, the size of the vocal repertoire strongly predicts group size as well as grooming time in non-human primates (McComb & Semple, 2005). Social bonding is maintained primarily via grooming in primates, but in increasingly large groups this behaviour, which tends to be a one-to-one activity, is less effective. While this theory is presented in relation to the origins of language, it suggests a stage of communal chorusing, lacking propositional meaning, which replaced grooming. Dunbar’s theory of vocal grooming (1996, 2003a, 2010) is consistent with archaeological, as well as social group size and neocortex size data, but lacks direct empirical support.

Dunbar’s hypothesis resembles that of Darwin (1871), who suggested a stage of vocal communication in human evolution more closely related to music (singing) than to spoken language. If this is true, music could be something of a *fossil* of that hypothetical early stage of vocal communication among hominins, often referred to as musical protolanguage (e.g. Fitch, 2011; Kirby, 2011), or music-like protolanguage (e.g. Fitch, 2006b). This, more general, idea of a shared *common ancestor* between music and language, is probably the most recurrent idea in evolutionary musicology. In fact, similar models covering protolanguage stages that relied on musical or music-like elements have been proposed (e.g. Mithen, 2006), including the musilanguage model of Brown (Brown, 2000b; see also Baroni, 2008), which suggests the idea of an expression spectrum, in which purely referential meaning (lacking emotional content) is on one end, and purely emotional meaning is on the other. The main strength of these models, beyond potentially addressing the origins of both music and language, is that they could explain the complex similarities between music and language (see section 7.1.3 in this chapter).

Finally, Trehub (2003) and Dissanayake (2000) have suggested that the primary role of music, and songs in particular, is to aid infant-parent communication. This hypothesis is supported by the apparent universality of lullabies (Trehub, 2000), and their calming effects on infants, which seems to be used in particular to aid them sleep. This idea is also compatible with the existence of IDS and its prevalence in parent-infant interactions; IDS has characteristic vocal modulation patterns which are detected by infants (Fernald & Kuhl, 1987), and has important effects on strengthening of mother-infant bonds, which could indicate that IDS is an important component in the development of musicality (Trehub, 2003). Furthermore, this hypothesis is compatible with Dunbar’s hypothesis of vocal grooming (1996, 2003a, 2010), and Falk (2004, 2005) has suggested that IDS could be a precursor of the social grooming stage that may have originated language.

In comparison to a sexual selection hypothesis for the origin of musicality, a theory based on parent-infant interactions appears to have important advantages: it can explain the early development of musical perception abilities, as pointed out by Fitch (2005), as well as the universal existence and effects of lullabies and IDS, therefore providing hints for a model that could explain, not only music and language, but also IDS.

To summarise, all evolutionary theories about musical capacities share an important component of emotional cohesion or social bonding. In other words, in this picture, music seems to be relevant at an individual level only in terms of benefits obtained during social interactions (parent-infant, mate choice, social bonding), or in group selection scenarios. Some evolutionary theories have tried to explain musicality as a product of one unique selection pressure (e.g. sexual or group selection), but it is important to contemplate the possibility that musicality might have played (and still play) different roles during its evolution, and thus may have been shaped by different selection pressures during this process.

7.3 Creating a simple model for the evolution of musicality

So far in this chapter I have discussed the problems of the scientific study of music, and its complex relation with language, as well as the most important similarities between complex vocal communication in humans and non-human animals, to finally highlight the most important theories for the origins of the music capacity. Musicality, the ability to process musical information, seems to integrate several processing modules, of which at least two, are recognised: pitch and rhythm processing (see Peretz, 2009). This is based on the cases in which a disorder or brain injury affects either pitch or rhythm processing, but not both (e.g. Alcock, Passingham, et al., 2000; Alcock, Wade, et al., 2000; Di Pietro et al., 2004). A model for the evolution of the musical capacity must take into account this apparent relative independency of processing components of musicality, which might have had separate evolutionary origins, and have been shaped under independent evolutionary pressures (see Fitch, 2006b), as discussed on section 7.2.1 in this chapter.

As most current theories for the origins of music suggest, music and language could be *descendants* of an earlier, vanished form of vocal communication among ancestral hominin species (section 7.2.2.2). This could help explain the relationship between music and language (section 7.1.3) and, potentially, infant-directed speech (IDS). Among these theories, a model based on the role of musicality in infant-parent communication has particular strengths, as it could further explain the universal features of IDS and lullabies, as well as the musicality of babies, which seems to be mostly innate (Justus & Hutsler, 2005; McDermott & Hauser, 2005; see also Trainor, 2006). It could also integrate hypotheses based on music playing a role in promoting group cohesion, as in Dunbar’s theory of vocal grooming (1996, 2003a, 2010).

As stated before, I believe evolutionary theories related to this subject should not focus on music, but on the capacity for processing musical information (musicality), understanding the cognitive components and potential modules musicality consists of, and studying their evolutionary history by tracking their role in several domains. It is important to consider that modularity does not equal domain specificity (see Peretz, 2009) and, if music, language, and perhaps IDS have a common evolutionary history, musicality (or components of it) might not be limited to music processing, and could in fact play a role in other domains.

The most important issue for any model, however, is to explain how human populations drifted from a state where musicality was practically inexistent or very modest, towards human groups with more musicality, and especially what force could have driven such a tendency. In the case of pitch processing –an important component of musicality (see Peretz, 2009)– a model based on infant-parent communication provides a potential, but plausible, evolutionary pressure for such drift (Fig. 7.1).

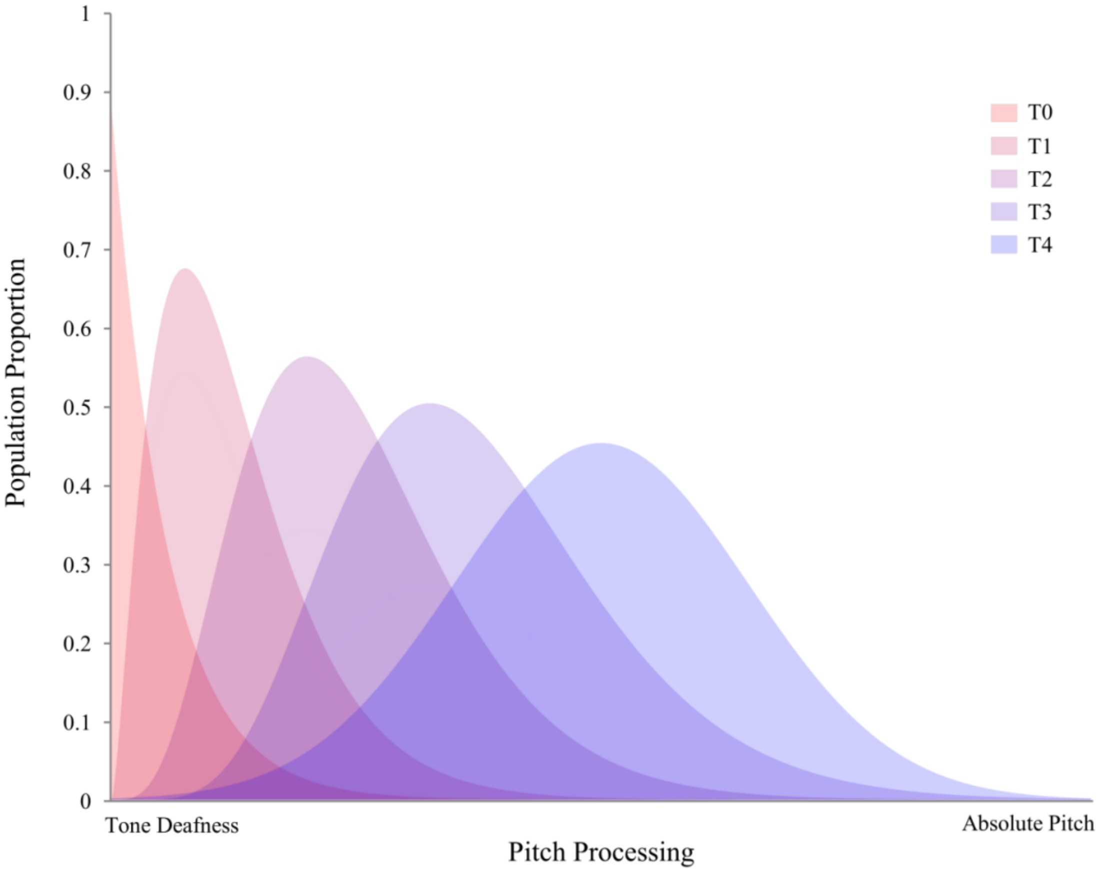


Figure 7.1. Hypothetical progression of pitch processing distribution during evolution. An evolutionary pressure for pitch discrimination (e.g. IDS and its role in parent-infant communication and bonding, and eventually language acquisition) gradually drifts population from a state in which most individuals are tone deaf (T0), to a state in which most individuals have a good level of pitch processing (T4).

Similarly, rhythm processing seems to be of particular importance in terms of synchronisation and entrainment, even from infancy (e.g. Malloch, 2000), highlighting the importance of rhythm, as well as pitch, processing in IDS, which is consistent with a model based on the role of musicality in infant-parent communication for the evolution of musicality. However, the influence of music in synchronising behaviours and promoting bonding, is especially manifest in group activities; today, for example, armies all around the world employ music and/or synchronised behaviours (such as marches), and common analogous examples are rhymes and chants from football fans and protesters. This, however, does not seem to be a modern phenomenon: among many notable traditional examples, are Zulu War Chants as well as the Haka from the Māori people of New Zealand. Music seems to reduce physical exertion (Fritz et al., 2013), which could also partially explain why music is common when human groups perform repetitive tasks.

In general, the evidence seems to be consistent with a model for the evolution of musicality based on its role in infant-parent communication. This theoretical model suggests that communication between infants and parents is at least beneficial to survival, to the point that it became a selective pressure. This seems to make sense since human children are born relatively underdeveloped (in comparison to other primate species), parental care is exceptionally long, and children require strong parent-infant bonds to guarantee parental care and avoid potentially fatal neglect. In fact, IDS is associated with variation in oxytocin levels and other neuropeptides involved in attachment mechanisms (e.g. Feldman, Weller, Zagoory-Sharon, & Levine, 2007; Gordon, Zagoory-Sharon, Leckman, & Feldman, 2010; Weisman et al., 2013). Better parent-infant communication, and particularly mother-infant bonding, could facilitate social learning in infants, allowing them to acquire the necessary skills to survive (see Broad, Curley, & Keverne, 2006).

If musicality evolved primarily as a means to facilitate parent-infant communication and strengthen bonding, to the point of becoming a selective pressure, it is likely that adults with a good level of musicality would have tended to be better parents than those with less musicality. Moreover, because musicality seems to be at least partially hereditary (see Drayna et al., 2001), adults with a good level of musicality could produce offspring better equipped to process this information, and who in turn would have the potential of being better parents, adding a new level to the selective pressure for musicality.

Furthermore, this model could integrate the evidence in support for a sexual selection hypothesis, including the preference for composers of more complex music around ovulation (Charlton, 2014), as well as, if confirmed, the potential increase in pitch discrimination when the risk of conception is high (as shown in Chapter 4). If musicality could affect infant survival, cues of musicality could likely start to be sexually selected, increasing the survival chances of offspring. Mating with someone with musical abilities could be appealing and relevant if that implies a capacity to bond and empathise with infants and other members of the community, and also produce offspring more likely to do so. This could explain the role of vocal modulation – analogous to that of IDS– during courtship described in Chapter 2 (Study 1) and its detection and preference by listeners (Study 2), as well as its apparent exceptional characteristics seen in courtship, but not in other social contexts (see Study 2, Chapter 2, and Chapter 5), which needs to be demonstrated.

This would, however, require some display of musicality, which could have been manifested in a music-like protolanguage, and would exploit the capacity of music to coordinate behaviour and promote social bonding. In a society where basic forms of group chorusing (proto-songs?) start to appear in the context of social rhythmic and coordinated behaviours, the interaction between the voice of male adults and women or children would tend to create octaves and fifths, provided the perceptual preference for these small-integer intervals or consonances (see Trehub, 2000). These group activities would start to promote, not only social bonding, but also group identity.

The theoretical splitting between music and language from a common ancestor (music-like protolanguage) might have been a product of the increased relevance that syntax and semantics played in human communication, building an increasing specialisation towards language. Language itself, however, is of little use in the context of interactions with pre-linguistic infants, leaving a domain in which musicality remained essential, specially being able to communicate and influence emotional states. For this to occur, musicality, however, needs to be present in both infants and adults, allowing the cognitive musical abilities to be employed for other purposes in which it remained influential (e.g. group cohesion and social identity).

In short, I am suggesting that musicality predates music, and that its primary and original purpose was not music (music being, in this view, an epiphenomenon). Musicality, however, as a name, seems relevant as it is today primarily noticeable in music contexts. This model for the evolution of musicality is summarised in Fig. 7.2.

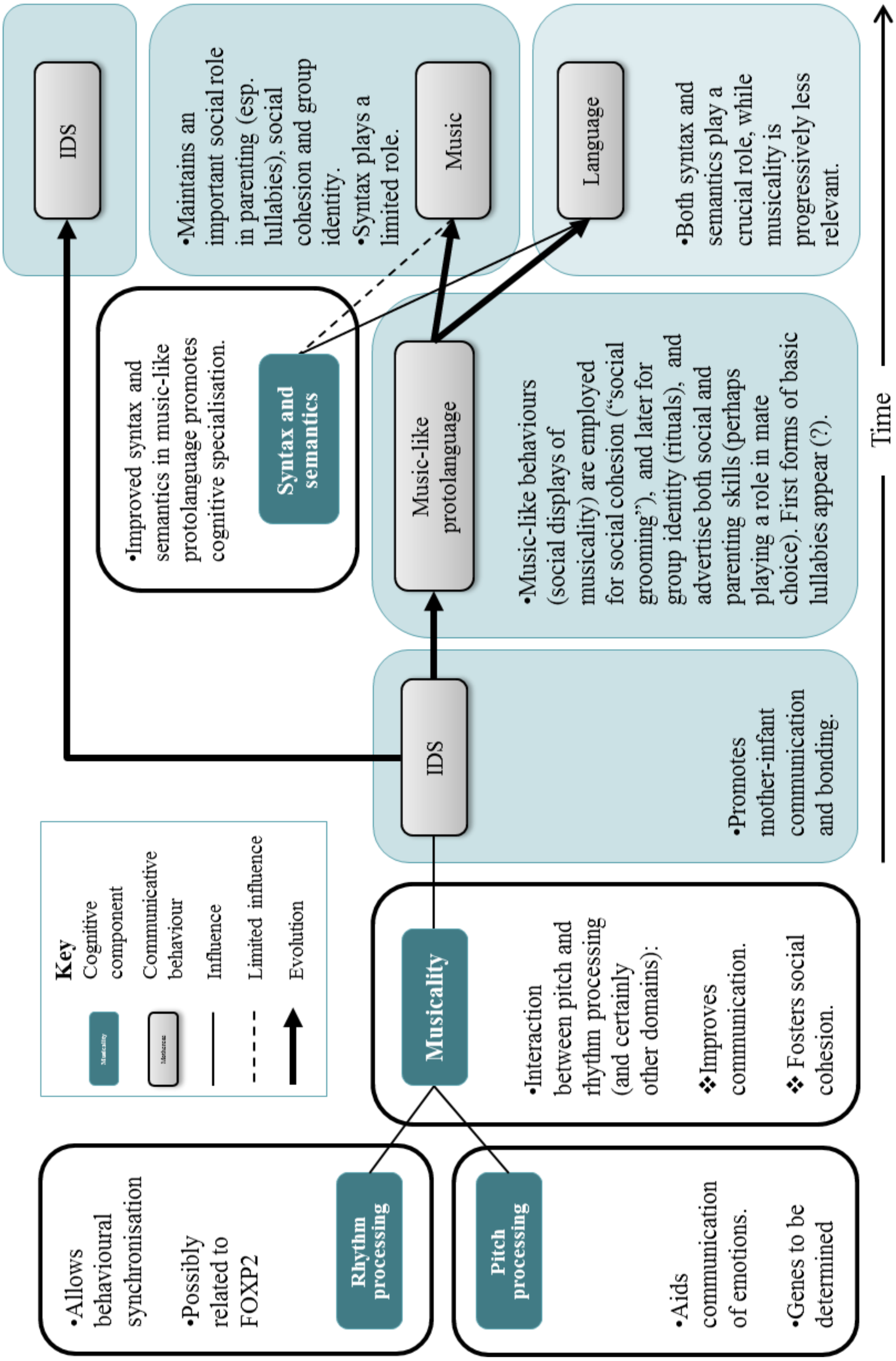


Figure 7.2. Model for the evolution of musicality and its role in human vocal communication. Musicality is presented as a simplified convergence of pitch and rhythm processing, which promotes infant-parent communication and bonding primarily through infant-directed speech (IDS).

If this model accurately portraits the evolution of musicality and its role in complex human vocal communication, music could be partially a *fossil* of our musical brain, whose original communicative purpose was the communication between parents and infants (IDS), and later the social communication of intentions and emotions to promote social bonding and coordination (music-like protolanguage). It is important to emphasise the word *partially*, because its original purpose of aiding communication and bonding between parents and infants is still biologically relevant today in IDS and lullabies, as well as its power to promote group cohesion and social identity, evident today, for example, in ritual music.

While this model is relatively simple, as it based only on two potential cognitive modules (i.e. pitch and rhythm processing), it provides a general view that corresponds with the most current evidence, and could explain the whole range of human complex vocal communication. Although this theory need to be tested, infant-parent bonding might be more than a link between music and language; it could be the very purpose of musicality, and might explain its existence, and its role in particular, non-specifically musical contexts such as IDS and courtship, in which it appears to be a form of contextual musicality.

REFERENCE LIST

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

Abraham, G. E. (1974). Ovarian and adrenal contribution to peripheral androgens during the menstrual cycle. *The Journal of Clinical Endocrinology & Metabolism*, *39*(2), 340–346. doi:10.1210/jcem-39-2-340

Adams, T. G., Stewart, P. A., & Blanchar, J. C. (2014). Disgust and the politics of sex: exposure to a disgusting odorant increases politically conservative views on sex and decreases support for gay marriage. *PLOS ONE*, *9*(5), e95572. doi:10.1371/journal.pone.0095572

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs*, *35*(4), 492–503. doi:10.1080/03637756809375599

Adler, D. S. (2009). Archaeology: The earliest musical tradition. *Nature*, *460*(7256), 695–696. doi:10.1038/460695a

Albrecht, T., Kleven, O., Kreisinger, J., Laskemoen, T., Omotoriogun, T. C., Ottosson, U., … Lifjeld, J. T. (2013). Sperm competition in tropical versus temperate zone birds. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1752), 20122434. doi:10.1098/rspb.2012.2434

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, *75*(1), 34–46. doi:10.1006/brln.2000.2323

Alcock, K. J., Wade, D., Anslow, P., & Passingham, R. E. (2000). Pitch and timing abilities in adult left-hemisphere-dysphasic and right-hemisphere-damaged subjects. *Brain and Language*, *75*(1), 47–65. doi:10.1006/brln.2000.2324

Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: a clue to malpractice history. *Surgery*, *132*(1), 5–9. doi:10.1067/msy.2002.124733

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter-gatherers. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1659), 1077–1082. doi:10.1098/rspb.2008.1542

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.

Apicella, C. L., & Marlowe, F. W. (2004). Perceived mate fidelity and paternal resemblance predict men's investment in children. *Evolution and Human Behavior*, *25*(6), 371–378. doi:10.1016/j.evolhumbehav.2004.06.003

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews*, *30*(3), 319–345. doi:10.1016/j.neubiorev.2004.12.007

Arnold, K., & Zuberbühler, K. (2006). Language evolution: semantic combinations in primate calls. *Nature*, *441*(7091), 303. doi:10.1038/441303a

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), R202–203. doi:10.1016/j.cub.2008.01.040

Atkinson, J., Pipitone, R. N., Sorokowska, A., Sorokowski, P., Mberira, M., Bartels, A., & Gallup, G. G. (2012). Voice and handgrip strength predict reproductive success in a group of indigenous African females. *PLOS ONE*, *7*(8), e41811. doi:10.1371/journal.pone.0041811

Baker, K. K. (2001). Control of vocal loudness in young and old adults. *Journal of Speech, Language, and Hearing Research*, *44*(2), 297–305. doi:10.1044/1092-4388(2001/024)

Balter, M. (2010). Evolution of language. Animal communication helps reveal roots of language. *Science*, *328*(5981), 969–971. doi:10.1126/science.328.5981.969

Baroni, M. (2008). Music, musicality, "musilanguage." *Musicae Scientiae*, *12*(1 Suppl), 197–218. doi:10.1177/1029864908012001091

Bencivelli, S. (2011). *Why We Like Music: Ear, Emotion, Evolution*. Hudson, NY: Music Word Media.

Blacking, J. (1995). Music, culture and experience. In R. Byron (Ed.), *Music, culture and experience: selected papers of John Blacking* (pp. 223–242). London: University of Chicago Press.

Blum, J. S., & Mehrabian, A. (1999). Personality and Temperament Correlates of Marital Satisfaction. *Journal of Personality*, *67*(1), 93–125. doi:10.1111/1467-6494.00049

Bohlman, S. (1999). Ontologies of music. In N. Cook & M. Everist (Eds.), *Rethinking music* (pp. 17–34). Oxford: OUP.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, *82*(1), 55–59. doi:10.1016/j.anbehav.2011.03.024

Bradley, R., & Whiteside-Mansell, L. (1997). Parents' socioemotional investment in children. *Journal of Marriage and Family, 59*(1), 77–90.

Brant, L. J. (1990). Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. *The Journal of the Acoustical Society of America, 88*(2), 813–820. doi:10.1121/1.399731

Broad, K. D., Curley, J. P., & Keverne, E. B. (2006). Mother-infant bonding and the evolution of mammalian social relationships. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2199–2214. doi:10.1098/rstb.2006.1940

Brown, S. (2000a). Evolutionary models of music: From sexual selection to group selection. In F. Tonneau & N. S. Thompson (Eds.), *Perspectives in Ethology* (Vol. 13, pp. 231–281). Kluwer Academic/Plenum.

Brown, S. (2000b). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 271–300). Cambridge, MA: MIT Press.

Brown, S., & Jordania, J. (2011). Universals in the world's musics. *Psychology of Music, 41*(2), 229–248. doi:10.1177/0305735611425896

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23*(10), 2791–2803. doi:10.1111/j.1460-9568.2006.04785.x

Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters, 5*(1), 12–15. doi:10.1098/rsbl.2008.0507

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435. doi:10.1126/science.1069587

Buss, D. M., & Dedden, L. A. (1990). Derogation of competitors. *Journal of Social and Personal Relationships, 7*(3), 395–422. doi:10.1177/0265407590073006

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*(2), 204–232.

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: brain regions involved with perception and covert production. *NeuroImage, 31*(3), 1327–1342. doi:10.1016/j.neuroimage.2006.01.036

Cerchio, S., Jacobsen, J. K., & Norris, T. F. (2001). Temporal and geographical variation in songs of humpback whales, *Megaptera novaeangliae*: synchronous change in Hawaiian and Mexican breeding assemblages. *Animal Behaviour*, *62*(2), 313–329. doi:10.1006/anbe.2001.1747

Charlton, B. D. (2014). Menstrual cycle phase alters women’s sexual preferences for composers of more complex music. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1784), 20140403. doi:10.1098/rspb.2014.0403

Charlton, B. D., Ellis, W. A. H., Brumm, J., Nilsson, K., & Fitch, W. T. (2012). Female koalas prefer bellows in which lower formants indicate larger males. *Animal Behaviour*, *84*(6), 1565–1571. doi:10.1016/j.anbehav.2012.09.034

Charlton, B. D., Filippi, P., & Fitch, W. T. (2012). Do women prefer more complex music around ovulation? *PLOS ONE*, *7*(4), e35626. doi:10.1371/journal.pone.0035626

Charlton, B. D., & Reby, D. (2011). Context-related acoustic variation in male fallow deer (*Dama dama*) groans. *PLOS ONE*, *6*(6), e21066. doi:10.1371/journal.pone.0021066

Chen, C. C. G., & Parker, C. R. (2004). Adrenal androgens and the immune system. *Seminars in Reproductive Medicine*, *22*(4), 369–377. doi:10.1055/s-2004-861553

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, *104*(1), 103–125. doi:10.1037/a0030398

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, *31*(5), 334–347. doi:10.1016/j.evolhumbehav.2010.02.004

Chen-Hafteck, L. (1997). Music and Language Development in Early Childhood: Integrating Past Research in the Two Domains. *Early Child Development and Care*, *130*(1), 85–97. doi:10.1080/0300443971300109

Collias, N. E. (1960). An ecological and functional classification of animal sounds. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal sounds and communication* (pp. 368–391). Washington, D.C.: American Institute of Biological Sciences.

Collins, S. A. (2000). Men’s voices and women's choices. *Animal Behaviour*, *60*(6), 773–780. doi:10.1006/anbe.2000.1523

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

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*(7256), 737–740. doi:10.1038/nature08169

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*(4), 412–427. doi:10.1037/a0032345

Cooke, I. D., Lenton, E. A., Adams, M., & Sobowale, O. (1978). Hormonal therapy and ovulatory disorders. In D. B. Crighton, N. B. Haynes, G. R. Foxcrofl, & G. E. Lamming (Eds.), *Control of Ovulation* (pp. 383–395). London: Butterworths.

Cross, I. (2003). Music and biocultural evolution. In T. Herbert & R. Middleton (Eds.), *The Cultural Study of Music: A Critical Introduction* (pp. 19–30). New York: Routledge.

Cross, I. (2006). Music, Cognition, Culture, and Evolution. *Annals of the New York Academy of Sciences*, *930*(1), 28–42. doi:10.1111/j.1749-6632.2001.tb05723.x

Cross, I., & Morley, I. (2008). The evolution of music: Theories, definitions and the nature of the evidence. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality* (pp. 61–82). Oxford: Oxford university Press.

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences*, *27*(4), 801–804. doi:10.1016/S0191-8869(98)00272-4

Dai, H., & Micheyl, C. (2011). Psychometric functions for pure-tone frequency discrimination. *The Journal of the Acoustical Society of America*, *130*(1), 263–272. doi:10.1121/1.3598448

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

Di Pietro, M., Laganaro, M., Leemann, B., & Schnider, A. (2004). Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia*, *42*(7), 868–877. doi:10.1016/j.neuropsychologia.2003.12.004

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother–infant interaction. In N. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 389–410). Cambridge, MA.: MIT Press.

Dittus, W. P. J. (1984). Toque macaque food calls: Semantic communication concerning food distribution in the environment. *Animal Behaviour*, *32*(2), 470–477. doi:10.1016/S0003-3472(84)80283-3

Drayna, D., Manichaikul, A., De Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, *291*(5510), 1969–1972. doi:10.1126/science.291.5510.1969

Dunbar, R. I. M. (1996). *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Dunbar, R. I. M. (2003a). The origin and subsequent evolution of language. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution. Studies in the evolution of language*. (pp. 219–234). New York, NY: Oxford university Press. doi:10.1093/acprof:oso/9780199244843.003.0012

Dunbar, R. I. M. (2003b). The Social Brain: Mind, Language, and Society in Evolutionary Perspective. *Annual Review of Anthropology*, *32*(1), 163–181.

Dunbar, R. I. M. (2010). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–694. doi:10.1017/S0140525X00032325

Ellis, C. J. (1984). The Nature of Australian Aboriginal Music. *International Journal of Music Education*, *4*(1), 47–50. doi:10.1177/025576148400400110

Eriksen, N., Miller, L. A., Tougaard, J., & Helweg, D. A. (2005). Cultural change in the songs of humpback whales (*Megaptera novaeangliae*) from Tonga. *Behaviour*, *142*, 305–328.

Evans, C. S., Evans, L., & Marler, P. (1993). On the meaning of alarm calls: functional reference in an avian vocal system. *Animal Behaviour*, *46*(1), 23–38. doi:10.1006/anbe.1993.1158

Evans, C. S., & Marler, P. (1994). Food calling and audience effects in male chickens, *Gallus gallus*: their relationships to food availability, courtship and social facilitation. *Animal Behaviour*, *47*(5), 1159–1170. doi:10.1006/anbe.1994.1154

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, *93*(4-5), 783–788. doi:10.1016/j.physbeh.2007.11.033

Falk, D. (2004). Prelinguistic evolution in hominin mothers and babies: For cryin’out loud! *Behavioral and Brain Sciences*, *27*(4), 461–462.

Falk, D. (2005). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(4), 491–503. doi:10.1017/S0140525X04000111

Fant, G. (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), *Manual of phonetics* (pp. 173–177). Amsterdam: North-Holland.

Farley, S. D., Hughes, S. M., & LaFayette, J. N. (2013). People Will Know We Are in Love: Evidence of Differences Between Vocal Samples Directed Toward Lovers and Friends. *Journal of Nonverbal Behavior*, *37*(3), 123–138. doi:10.1007/s10919-013-0151-3

Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Memory & Cognition*, *37*(1), 1–9. doi:10.3758/MC.37.1.1

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology*, *17*(2), 112–118. doi:10.1002/evan.20166

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. doi:10.1016/j.evolhumbehav.2007.12.008

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., … Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *26*(5), 398–408. doi:10.1016/j.evolhumbehav.2005.04.001

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., … Perrett, D. I. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior*, *49*(2), 215–222.

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. doi:10.1016/j.anbehav.2004.06.012

Feldman, R., Weller, A., Zagoory-Sharon, O., & Levine, A. (2007). Evidence for a neuroendocrinological foundation of human affiliation: plasma oxytocin levels across pregnancy and the postpartum period predict mother-infant bonding. *Psychological Science*, *18*(11), 965–970. doi:10.1111/j.1467-9280.2007.02010.x

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*, *278*(1724), 3551–3557. doi:10.1098/rspb.2011.0544

Ferguson, C. A. (1977). Baby talk as a simplified register. In C. Snow & C. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition* (pp. 219–235). Cambridge: Cambridge University Press.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. doi:10.1016/0163-6383(87)90017-8

Fisher, M. L. (2004). Female intrasexual competition decreases female facial attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, *271*(Suppl 5), S283–S285. doi:10.1098/rsbl.2004.0160

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological Review*, *99*(4), 689–723.

Fitch, W. T. (2005). The evolution of music in comparative perspective. *Annals Of The New York Academy Of Sciences*, *1060*(The Neurosciences and Music II: From Perception to Performance), 29–49. doi:10.1196/annals.1360.004

Fitch, W. T. (2006a). On the biology and evolution of music. *Music Perception*, *24*(1), 85–88. doi:10.1525/mp.2006.24.1.85

Fitch, W. T. (2006b). The biology and evolution of music: a comparative perspective. *Cognition*, *100*(1), 173–215. doi:10.1016/j.cognition.2005.11.009

Fitch, W. T. (2011). The biology and evolution of rhythm: unravelling a paradox. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 73–95). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199553426.001.0009

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, *7*, 68. doi:10.3389/fnsys.2013.00068

Fitch, W. T., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on “honest” advertisement. *American Journal of Primatology*, *37*(3), 191–219. doi:10.1002/ajp.1350370303

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., … 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. doi:10.1556/JEP.9.2011.33.1

Fraccaro, P. J., O’Connor, J. J. M., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour*, *85*(1), 127–136. doi:10.1016/j.anbehav.2012.10.016

Franchimont, P., Dourcy, C., Legros, J. J., Reuter, A., Vrindts-Gevaert, Y., Cauwenberge, J. R. Van, & Gaspard, U. (1976). Prolactin levels during the menstrual cycle. *Clinical Endocrinology, 5*(6), 643–650. doi:10.1111/j.1365-2265.1976.tb03867.x

Fritz, T., Hardikar, S., Demoucron, M., Niessen, M., Demey, M., Giot, O., … Leman, M. (2013). Musical agency reduces perceived exertion during strenuous physical performance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(44), 17784–17789. doi:10.1073/pnas.1217252110

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., … Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19*(7), 573–576. doi:10.1016/j.cub.2009.02.058

Frost, P. (1994). Preference for darker faces in photographs at different phases of the menstrual cycle: preliminary assessment of evidence for a hormonal relationship. *Perceptual and Motor Skills, 79*(1), 507–514. doi:10.2466/pms.1994.79.1.507

Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception, 1*(1), 19–39. doi:10.1145/1008722.1008725

Ganchev, T., Fakotakis, N., & Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task. In *Proceedings of the 10th International Conference Speech and Computer (SPECOM 2005)* (pp. 191–194). Patras, Greece: University of Patras. doi:10.1.1.75.8303

Gangestad, S. W. (2003). Facial masculinity and fluctuating asymmetry. *Evolution and Human Behavior*, *24*(4), 231–241. doi:10.1016/S1090-5138(03)00017-5

Gangestad, S. W., Simpson, J. A., Cousins, A. J., Garver-Apgar, C. E., & Christensen, P. N. (2004). Women's preferences for male behavioral displays change across the menstrual cycle. *Psychological Science, 15*(3), 203–207. doi:10.1111/j.0956-7976.2004.01503010.x

Gangestad, S. W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proceedings of the Royal Society B: Biological Sciences, 265*(1399), 927–933. doi:10.1098/rspb.1998.0380

Glasberg, B. R., & Moore, B. C. . (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*(1-2), 103–138. doi:10.1016/0378-5955(90)90170-T

Gordon, I., Zagoory-Sharon, O., Leckman, J. F., & Feldman, R. (2010). Oxytocin and the development of parenting in humans. *Biological Psychiatry*, *68*(4), 377–382. doi:10.1016/j.biopsych.2010.02.005

Gourlay, K. A. (1984). The Non-Universality of Music and the Universality of Non-Music. *The World of Music*, *26*(2), 25–39.

Grammer, K. (1993). 5-αlpha-androst-16en-3αlpha-on: A male pheromone? A brief report. *Ethology and Sociobiology*, *14*(3), 201–207. doi:10.1016/0162-3095(93)90006-4

Gray, R. D., Atkinson, Q. D., & Greenhill, S. J. (2011). Language evolution and human history: what a difference a date makes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1090–1100.

Green, S. R., Mercado, E., Pack, A. A., & Herman, L. M. (2011). Recurring patterns in the songs of humpback whales (*Megaptera novaeangliae*). *Behavioural Processes*, *86*(2), 284–294.

Greenwood, D. D. (1997). The Mel Scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios. *Hearing Research*, *103*, 199–224. doi:10.1016/S0378-5955(96)00175-X

Gregory, S. W., Green, B. E., Carrothers, R. M., Dagan, K. A., & Webster, S. W. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. *Language and Communication*, *21*(1), 37–60.

Gyger, M., & Marler, P. (1988). Food calling in the domestic fowl, *Gallus gallus*: the role of external referents and deception. *Animal Behaviour*, *36*(2), 358–365. doi:10.1016/S0003-3472(88)80006-X

Gyger, M., Marler, P., & Pickert, R. (1987). Semantics of an Avian Alarm Call System: the Male Domestic Fowl, *Gallus Domesticus*. *Behaviour*, *102*(1), 15–40. doi:10.1163/156853986X00027

Hahn, J., & Münzel, S. C. (1995). Knochenflöten aus dem Aurignacien des Geissenklösterle bei Blaubeuren, Alb-Donau-Kreis [Bone flutes from the Aurignacian of Geissenklösterle, near Blaubeuren in the Alb-Donau-Kreis]. *Fundberichte Aus Baden-Württemberg*, *20*, 1–12.

Harnsberger, J. D., Shrivastav, R., Brown, W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice*, *22*(1), 58–69. doi:10.1016/j.jvoice.2006.07.004

Hasegawa, A., Okanoya, K., Hasegawa, T., & Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports*, *1*, 120. doi:10.1038/srep00120

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports, 3*, 1566. doi:10.1038/srep01566

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*(5598), 1569–79. doi:10.1126/science.298.5598.1569

Hauser, M. D., & Marler, P. (1993a). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology*, *4*(3), 194–205. doi:10.1093/beheco/4.3.194

Hauser, M. D., & Marler, P. (1993b). Food-associated calls in rhesus macaques (*Macaca mulatta*): II. Costs and benefits of call production and suppression. *Behavioral Ecology*, *4*(3), 206–212. doi:10.1093/beheco/4.3.206

Hauser, M. D., Teixidor, P., Fields, L., & Flaherty, R. (1993). Food-elicited calls in chimpanzees: effects of food quantity and divisibility. *Animal Behaviour*, *45*(4), 817–819. doi:10.1006/anbe.1993.1096

Havlíček, J., Roberts, S. C., & Flegr, J. (2005). Women's preference for dominant male odour: effects of menstrual cycle and relationship status. *Biology Letters*, *1*(3), 256–259. doi:10.1098/rsbl.2005.0332

Helweg, D. A., Herman, L. M., Yamamoto, S., & Forestell, P. H. (1990). Comparison of Songs of Humpback Whales (*Megaptera Novaeangliae*) recorded in Japan, Hawaii, and Mexico During the Winter of 1989. *Scientific Reports of the Cetacean Research Institute, 1*, 1–20.

Hendrick, S. S. (1988). A generic measure of Relationship Satisfaction. *Journal of Marriage and Family*, *50*(1), 93–98. doi:10.2307/352430

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*(3), 165–196.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. doi:10.1162/089892903322307393

Higgins, K. M. (2012). *The Music Between Us: Is Music a Universal Language?* Chicago, IL: The University of Chicago Press.

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, *21*(4), 406–427. doi:10.1007/s12110-010-9101-5

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2011). Voice correlates of mating success in men: examining “contests” versus “mate choice” modes of sexual selection. *Archives of Sexual Behavior*, *40*(3), 551–557. doi:10.1007/s10508-010-9625-0

Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells like clean spirit. Nonconscious effects of scent on cognition and behavior. *Psychological Science*, *16*(9), 689–693. doi:10.1111/j.1467-9280.2005.01597.x

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker’s F0 range. *The Journal of the Acoustical Society of America*, *117*(4), 2193. doi:10.1121/1.1841751

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 Behaviour*, *34*(3), 155–167. doi:10.1007/s10919-010-0087-9

Hughes, S. M., Mogilski, J. K., & Harrison, M. a. (2013). The Perception and Parameters of Intentional Voice Manipulation. *Journal of Nonverbal Behavior*, *38*(1), 107–127. doi:10.1007/s10919-013-0163-z

Hurkmans, J., de Bruijn, M., Boonstra, A. M., Jonkers, R., Bastiaanse, R., Arendzen, H., & Reinders-Messelink, H. A. (2012). Music in the treatment of neurological language and speech disorders: A systematic review. *Aphasiology*, *26*(1), 1–19. doi:10.1080/02687038.2011.602514

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, *20*(2), 151–171. doi:10.1525/mp.2002.20.2.151

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*(5), 356–360. doi:10.1111/j.0956-7976.2004.00683.x

Jacob, S., & McClintock, M. K. (2000). Psychological state and mood effects of steroidal chemosignals in women and men. *Hormones and Behavior*, *37*(1), 57–78. doi:10.1006/hbeh.1999.1559

Jarvis, E. D. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*, 749.

Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Words in melody: an H215O PET study of brain activation during singing and speaking. *NeuroReport*, *14*(5), 749–754. doi:10.1097/00001756-200304150-00018

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience*, *20*(11), 1940–1951. doi:10.1162/jocn.2008.20135

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. doi:10.1016/S1090-5138(01)00066-6

Jones, B. C., DeBruine, L. M., Perrett, D. I., Little, A. C., Feinberg, D. R., & Law Smith, M. J. (2008). Effects of menstrual cycle phase on face preferences. *Archives of Sexual Behavior*, *37*(1), 78–84. doi:10.1007/s10508-007-9268-y

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2008). Integrating cues of social interest and voice pitch in men's preferences for women's voices. *Biology Letters, 4*(2), 192.

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

Judd, H. L., & Yen, S. S. C. (1972). Serum androstenedione and testosterone levels during the menstrual cycle. *The Journal of Clinical Endocrinology & Metabolism*, *36*(3), 475–481. doi:http://dx.doi.org/10.1210/jcem-36-3-475

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814.

Juslin, P. N., & Sloboda, J. A. (Eds.). (2001). *Music and emotion: Theory and research*. Oxford: Oxford University Press.

Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception*, *23*(1), 1–27. doi:10.1525/mp.2005.23.1.1

Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of Neurophysiology*, *96*(3), 1441–1455. doi:10.1152/jn.01138.2005

Karakashian, S. J., Gyger, M., & Marler, P. (1988). Audience effects on alarm calling in chickens (*Gallus gallus*). *Journal of Comparative Psychology, 102*(2), 129–135. doi:10.1037/0735-7036.102.2.129

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, *132*(2), 1050–1060. doi:10.1121/1.4730893

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (2009). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*(1), 55–68. doi:10.1017/S030500090001343X

Kirby, S. (2011). Darwin's musical protolanguage: an increasingly compelling picture. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 96–102). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199553426.001.0010

Knasko, S. C. (1992). Ambient odor's effect on creativity, mood, and perceived health. *Chemical Senses*, *17*(1), 27–35. doi:10.1093/chemse/17.1.27

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: an fMRI study. *NeuroImage*, *25*(4), 1068–1076. doi:10.1016/j.neuroimage.2004.12.050

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schroger, E., & Friederici, A. D. (2003). Children processing music: Electric brain responses reveal musical competence and gender differences. *Journal of Cognitive Neuroscience*, *15*(5), 683–693.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–84. doi:10.1016/j.tics.2005.10.001

Kroodsma, D. E. (1980). Winter wren singing behavior: a pinnacle of song complexity. *The Condor*, *82*(4), 357–365. doi:10.2307/1367556

Kroodsma, D. E., & Momose, H. (1991). Songs of the Japanese population of the winter wren (*Troglodytes troglodytes*). *The Condor*, *93*(2), 424–432. doi:10.2307/1368959

Kudo, H. (1987). The study of vocal communication of wild mandrills in Cameroon in relation to their social structure. *Primates*, *28*(3), 289–308. doi:10.1007/BF02381013

Lenochova, P., Roberts, S. C., & Havlíček, J. (2009). Methods of human body odor sampling: the effect of freezing. *Chemical Senses*, *34*(2), 127–138. doi:10.1093/chemse/bjn067

Lieberman, P., & Blumstein, S. E. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge, UK: Cambridge University Press.

Lifjeld, J. T., Laskemoen, T., Kleven, O., Albrecht, T., & Robertson, R. J. (2010). Sperm length variation as a predictor of extrapair paternity in passerine birds. *PLOS ONE*, *5*(10), e13456. doi:10.1371/journal.pone.0013456

Lindsay, P. H., & Norman, D. A. (1977). *Human information processing: An introduction to psychology* (2nd ed.). New York: Academic Press.

Linville, S. E. (1996). The sound of senescence. *Journal of Voice*, *10*(2), 190–200.

Luo, S. (2009). Partner selection and relationship satisfaction in early dating couples: The role of couple similarity. *Personality and Individual Differences*, *47*(2), 133–138. doi:10.1016/j.paid.2009.02.012

Luther, D., & Baptista, L. (2010). Urban noise and the cultural evolution of bird songs. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1680), 469–473. doi:10.1098/rspb.2009.1571

Macedonia, J. M. (2010). What is communicated in the antipredator calls of lemurs: Evidence from playback experiments with ringtailed and ruffed lemurs. *Ethology*, *86*(3), 177–190. doi:10.1111/j.1439-0310.1990.tb00428.x

Maeda, H., Higashi, N., Uchida, S., Sato, F., Yamaguchi, M., Koido, T., & Takemura, A. (2000). Songs of humpback whales *Megaptera novaeangliae* in the Ryukyu and Bonin regions. *Mammal Study*, *25*(1), 59–73. doi:10.3106/mammalstudy.25.59

Makeig, S. (1982). Affective versus analytic perception of musical intervals. In M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 227–250). New York: Plenum.

Malloch, S. N. (2000). Mothers and infants and communicative musicality. *Musicae Scientiae*, *2*(2; SPI), 29–57. doi:10.1177/10298649000030S104

Maner, J. K., & Miller, S. L. (2014). Hormones and social monitoring: Menstrual cycle shifts in progesterone underlie women's sensitivity to social information. *Evolution and Human Behavior*, *35*(1), 9–16. doi:10.1016/j.evolhumbehav.2013.09.001

Marler, P. (2000). Origins of music and speech: Insights from animals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 31–48). Cambridge, MA: MIT Press.

Marler, P., Dufty, A., & Pickert, R. (1986a). Vocal communication in the domestic chicken: I. Does a sender communicate information about the quality of a food referent to a receiver? *Animal Behaviour*, *34*, 188–193. doi:10.1016/0003-3472(86)90022-9

Marler, P., Dufty, A., & Pickert, R. (1986b). Vocal communication in the domestic chicken: II. Is a sender sensitive to the presence and nature of a receiver? *Animal Behaviour*, *34*, 194–198. doi:10.1016/0003-3472(86)90023-0

Marler, P., & Pickert, R. (1984). Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Animal Behaviour*, *32*(3), 673–689. doi:10.1016/S0003-3472(84)80143-8

Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews*, *6*(1), 11–22.

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, *1*(4), 381–385. doi:10.1098/rsbl.2005.0366

McDermott, J. H., & Hauser, M. D. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, *23*(1), 29–60.

Mellars, P. (2010). Neanderthal symbolism and ornament manufacture: the bursting of a bubble? *Proceedings of the National Academy of Sciences of the United States of America*, *107*(47), 20147–20148. doi:10.1073/pnas.1014588107

Merker, B. (2000a). Synchronous chorusing and human origins. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 315–327). Cambridge, MA: MIT Press.

Merker, B. (2000b). Synchronous chorusing and the origins of music. *Musicae Scientiae*, *2*(2; SPI), 59–74.

Merker, B. (2001). Music: the missing Humboldt system. *Musicae Scientiae*, *6*(1), 3–22.

Midgley, A. R., & Jaffe, R. B. (1968). Regulation of human gonadotropins. IV. Correlation of serum concentrations of follicle stimulating and luteinizing hormones during the menstrual cycle. *The Journal of Clinical Endocrinology and Metabolism*, *28*(12), 1699–1703. doi:10.1210/jcem-28-12-1699

Miller, G. F. (2000). Evolution of human music through sexual selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 329–360). MIT Press.

Mithen, S. J. (2006). *The singing Neanderthals : the origin of music, language, mind and body*. London: Phoenix.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, *74*(3), 750–753. doi:10.1121/1.389861

Mulac, A., & Giles, H. (1996). “Your’re only as old as you sound”: Perceived vocal age and social meanings. *Health Communication*, *8*(3), 199–215. doi:10.1207/s15327027hc0803_2

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 463–472). Cambridge, MA.: MIT Press.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature*, *408*(6812), 537. doi:10.1038/35046199

O’Shaughnessy, D. (1987). *Speech communication: human and machine*. Reading, MA.: Addison-Wesley.

Ohala, J. J. (1982). The voice of dominance. *The Journal of the Acoustical Society of America*, *72*(S1), S66. doi:10.1121/1.2020007

Ozdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, *33*(2), 628–635. doi:10.1016/j.neuroimage.2006.07.013

Papoušek, M. (1996). Intuitive parenting: a hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical Beginnings* (pp. 88–112). Oxford: OUP.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674–681.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*(1), 99–104. doi:10.1525/mp.2006.24.1.99

Patel, A. D. (2010). Music, biological evolution, and the brain. In M. Bailar (Ed.), *Emerging Disciplines: Shaping New Fields of Scholarly Inquiry in and beyond the Humanities* (pp. 41–64). Houston, TX.: Rice University Press.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009a). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*(10), 827–830. doi:10.1016/j.cub.2009.03.038

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009b). Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, *1169*, 459–469. doi:10.1111/j.1749-6632.2009.04581.x

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain and Language*, *61*(1), 123–144. doi:10.1006/brln.1997.1862

Payne, K. (2000). The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In N. L. Wallin, B. Merker, & S. Brown (Eds.), (pp. 135–150). Cambridge, MA: MIT Press.

Payne, K., Tyack, P., & Payne, R. (1983). Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in hawaii. In R. Payne (Ed.), *Communication and Behavior of Whales* (pp. 9–57). Boulder, Colo.: Westview Press.

Pearce, J. M. S. (2005). Selected observations on amusia. *European Neurology, 54*(3), 145–148. doi:10.1159/000089606

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. doi:10.1016/S1090-5138(99)00033-1

Penton-Voak, I. S., Perrett, D. I., Castles, D. L., Kobayashi, T., Burt, D. M. M., Murray, L. K., & Minamisawa, R. (1999). Menstrual cycle alters face preference. *Nature, 399*(6738), 741–742. doi:10.1038/21557

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*(1), 1–32. doi:10.1016/j.cognition.2005.11.004

Peretz, I. (2009). Music, language and modularity framed in action. *Psychologica Belgica*, *49*, *2*(3), 157–175.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. doi:10.1038/nn1083

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367. doi:10.1016/S1364-6613(03)00150-5

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89–114. doi:10.1146/annurev.psych.56.091103.070225

Perlovsky, L. (2010). Musical emotions: functions, origins, evolution. *Physics of Life Reviews*, *7*(1), 2–27. doi:10.1016/j.plrev.2009.11.001

Pinker, S. (1997). *How the mind works*. New York, NY: Norton.

Pipitone, R. N., & Gallup, G. G. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, *29*(4), 268–274. doi:10.1016/j.evolhumbehav.2008.02.001

Podos, J., Huber, S. K., & Taft, B. (2004). Bird song: the interface of evolution and mechanism. *Annual Review of Ecology, Evolution, and Systematics, 35*, 55–87.

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

Puts, D. A. (2006). Cyclic variation in women's preferences for masculine traits. *Human Nature, 17*(1), 114–127. doi:10.1007/s12110-006-1023-x

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 601–609. doi:10.1098/rspb.2011.0829

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*(4), 283–296. doi:10.1016/j.evolhumbehav.2005.11.003

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340–344. doi:10.1016/j.evolhumbehav.2007.05.002

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain, 129*(10), 2571–2584. doi:10.1093/brain/awl250

Reby, D., Charlton, B. D., Locatelli, Y., & McComb, K. (2010). Oestrous red deer hinds prefer male roars with higher fundamental frequencies. *Proceedings of the Royal Society B: Biological Sciences, 277*(1695), 2747–2753. doi:10.1098/rspb.2010.0467

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences, 272*(1566), 941–947. doi:10.1098/rspb.2004.2954

Riding, D., Lonsdale, D., & Brown, B. (2006). The Effects of Average Fundamental Frequency and Variance of Fundamental Frequency on Male Vocal Attractiveness to Women. *Journal of Nonverbal Behaviour, 30*(2), 55–61. doi:10.1007/s10919-006-0005-3

Rikowski, A., & Grammer, K. (1999). Human body odour, symmetry and attractiveness. *Proceedings. Biological Sciences / The Royal Society, 266*(1422), 869–874. doi:10.1098/rspb.1999.0717

Roberts, S. C., Cobey, K. D., Klapilová, K., & Havlíček, J. (2013). An evolutionary approach offers a fresh perspective on the relationship between oral contraception and sexual desire. *Archives of Sexual Behavior, 42*(8), 1369–1375. doi:10.1007/s10508-013-0126-9

Roberts, S. C., Gosling, L. M., Carter, V., & Petrie, M. (2008). MHC-correlated odour preferences in humans and the use of oral contraceptives. *Proceedings of the Royal Society B: Biological Sciences, 275*(1652), 2715–2722. doi:10.1098/rspb.2008.0825

Roberts, S. C., Gosling, L. M., Spector, T. D., Miller, P., Penn, D. J., & Petrie, M. (2005). Body odor similarity in noncohabiting twins. *Chemical Senses, 30*(8), 651–656. doi:10.1093/chemse/bji058

Roberts, S. C., & Little, A. C. (2008). Good genes, complementary genes and human mate preferences. *Genetica, 132*(3), 309–321. doi:10.1007/s10709-007-9174-1

Roberts, S. C., Little, A. C., Burriss, R. P., Cobey, K. D., Klapilová, K., Havlíček, J., … Petrie, M. (2014). Partner choice, relationship satisfaction, and oral contraception: The congruency hypothesis. *Psychological Science*. doi:10.1177/0956797614532295

Roberts, S. C., Little, A. C., Lyndon, A., Roberts, J., Havlíček, J., & Wright, R. L. (2009). Manipulation of body odour alters men's self-confidence and judgements of their visual attractiveness by women. *International Journal of Cosmetic Science, 31*(1), 47–54. doi:10.1111/j.1468-2494.2008.00477.x

Robins, R. W., Caspi, A., & Moffitt, T. E. (2000). Two personalities, one relationship: Both partners' personality traits shape the quality of their relationship. *Journal of Personality and Social Psychology, 79*(2), 251–259. doi:10.1037/0022-3514.79.2.251

Roney, J. R., & Simmons, Z. L. (2008). Women's estradiol predicts preference for facial cues of men's testosterone. *Hormones and Behavior, 53*(1), 14–19. doi:10.1016/j.yhbeh.2007.09.008

Ruff, C. B. (1994). Morphological adaptation to climate in modern and fossil hominids. *American Journal of Physical Anthropology, 37*(S19), 65–107. doi:10.1002/ajpa.1330370605

Ruff, C. B. (2000). Body size, body shape, and long bone strength in modern humans. *Journal of Human Evolution, 38*(2), 269–290. doi:10.1006/jhev.1999.0322

Russo, F. A., & Thompson, W. F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin & Review, 12*(6), 1068–1075. doi:10.3758/BF03206445

Ryan, M. J. (1980). Female mate choice in a neotropical frog. *Science, 209*(4455), 523–525. doi:10.1126/science.209.4455.523

Sacks, O. W. (2007). *Musicophilia: Tales of music and the brain*. London: Picador.

Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *NeuroReport*, *17*(18), 1893–1897.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., … Schulze-Bonhage, A. (2009). Overlap of musical and linguistic syntax processing: intracranial ERP evidence. *Annals of the New York Academy of Sciences*, *1169*, 494–498. doi:10.1111/j.1749-6632.2009.04792.x

Saxton, T. K., Lyndon, A., Little, A. C., & Roberts, S. C. (2008). Evidence that androstadienone, a putative human chemosignal, modulates women’s attributions of men's attractiveness. *Hormones and Behavior, 54*(5), 597–601. doi:10.1016/j.yhbeh.2008.06.001

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, *19*(10), 831–836. doi:10.1016/j.cub.2009.03.061

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341–349.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., … Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1699), 3509–3518. doi:10.1098/rspb.2010.0769

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803. doi:10.1126/science.7433999

Signoret, J. L., van Eeckhout, P., Poncet, M., & Castaigne, P. (1987). Aphasia without amusia in a blind organist. Verbal alexia-agraphia without musical alexia-agraphia in braille. *Revue Neurologique*, *143*(3), 172–181.

Skeie, G. O., Einbu, T., & Aarli, J. (2010). Singing Improves Word Production in Patients with Aphasia. In F. C. Rose (Ed.), *Neurology of Music* (pp. 347–357). London: Imperial College Press.

Spangenberg, E. R., Sprott, D. E., Grohmann, B., & Tracy, D. L. (2006). Gender-congruent ambient scent influences on approach and avoidance behaviors in a retail store. *Journal of Business Research, 59*(12), 1281–1287. doi:10.1016/j.jbusres.2006.08.006

Sprecher, S. (1989). The importance to males and females of physical attractiveness, earning potential, and expressiveness in initial attraction. *Sex Roles*, *21*(9-10), 591–607.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*, *134*(10), 3083–3093. doi:10.1093/brain/awr240

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937a). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, *8*(3), 185–190. doi:10.1121/1.1915893

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937b). The mel scale equates the magnitude of perceived differences in pitch at different frequencies. *Journal of the Acoustical Society of America*, *8*, 185–190.

Tanabe, H. C., Kochiyama, T., Ogihara, N., & Sadato, N. (2014). Integrated analytical scheme for comparing the Neanderthal brain to modern human brain using neuroimaging techniques. In T. Akazawa, N. Ogihara, H. C Tanabe, & H. Terashima (Eds.), *Dynamics of Learning in Neanderthals and Modern Humans Volume 2* (Vol. 2, pp. 203–207). Tokyo: Springer Japan. doi:10.1007/978-4-431-54553-8

Thaut, M. H. (2005). *Rhythm, music, and the brain: scientific foundations and clinical applications*. New York, NY: Routledge.

Thorne, F., Neave, N., Scholey, A., Moss, M., & Fink, B. (2002). Effects of putative male pheromones on female ratings of male attractiveness: influence of oral contraceptives and the menstrual cycle. *Neuroendocrinology Letters*, *23*(4), 291–297.

Thornhill, R., & Gangestad, S. W. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior*, *20*(3), 175–201. doi:10.1016/S1090-5138(99)00005-7

Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology*, *84*(3), 558–568. doi:10.1037/0022-3514.84.3.558

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

Tomlinson, G. (1984). The web of culture: A context for musicology. *19th-Century Music*, *7*(3), 350–362. doi:10.2307/746387

Trainor, L. J. (2006). Innateness, learning, and the difficulty of determining whether music is an evolutionary adaptation. *Music Perception*, *24*(1), 105–110. doi:10.1525/mp.2006.24.1.105

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195. doi:10.1111/1467-9280.00240

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97–100. doi:10.1121/1.399849

Trehub, S. E. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (Vol. The origin, pp. 427–448). Cambridge, MA.: MIT Press.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, *930*(1), 1–16. doi:10.1111/j.1749-6632.2001.tb05721.x

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669–673. doi:10.1038/nn1084

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition*, *100*(1), 73–99. doi:10.1016/j.cognition.2005.11.006

Trehub, S. E., & Trainor, L. J. (1993). Listening strategies in infancy: the roots of music and language development. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 278–327). Oxford: Clarendon Press.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behavior and Development*, *16*(3), 285–295. doi:10.1016/0163-6383(93)80036-8

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual Selection and the Descent of Man: The Darwinian Pivot* (pp. 136–179). Chicago, IL: Aldine.

Tusing, K., & Dillard, J. (2000). The sounds of dominance: Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, *26*(1), 148–171. doi:10.1111/j.1468-2958.2000.tb00754.x

Tybur, J. M., Bryan, A. D., Magnan, R. E., & Hooper, A. E. C. (2011). Smells like safe sex: olfactory pathogen primes increase intentions to use condoms. *Psychological Science*, *22*(4), 478–480. doi:10.1177/0956797611400096

Umesh, S., Cohen, L., & Nelson, D. (1999). Fitting the Mel scale. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings* (pp. 217–220 vol.1). Phoenix, AZ: IEEE. doi:10.1109/ICASSP.1999.758101

Umesh, S., Cohen, L., & Nelson, D. (2002). Frequency warping and the Mel scale. *IEEE Signal Processing Letters*, *9*(3), 104–107. doi:10.1109/97.995829

Van Horne, B. (1995). Assessing vocal variety in the winter wren, a bird with a complex repertoire. *The Condor*, *97*(1), 39–49. doi:10.2307/1368981

Villa, P., & Roebroeks, W. (2014). Neandertal demise: an archaeological analysis of the modern human superiority complex. *PLOS ONE*, *9*(4), e96424. doi:10.1371/journal.pone.0096424

Watkins, C. D., Fraccaro, P. J., Smith, F. G., Vukovic, J., Feinberg, D. R., DeBruine, L. M., & Jones, B. C. (2010). Taller men are less sensitive to cues of dominance in other men. *Behavioral Ecology*, *21*(5), 943–947. doi:10.1093/beheco/arq091

Watkins, C. D., Jones, B. C., & DeBruine, L. M. (2010). Individual differences in dominance perception: dominant men are less sensitive to facial cues of male dominance. *Personality and Individual Differences*, *49*(8), 967–971. doi:10.1016/j.paid.2010.08.006

Wedekind, C., Seebeck, T., Bettens, F., & Paepke, A. J. (1995). MHC-dependent mate preferences in humans. *Proceedings. Biological Sciences / The Royal Society*, *260*(1359), 245–249. doi:10.1098/rspb.1995.0087

Weisman, O., Delaherche, E., Rondeau, M., Chetouani, M., Cohen, D., & Feldman, R. (2013). Oxytocin shapes parental motion during father-infant interaction. *Biology Letters*, *9*(6), 20130828. doi:10.1098/rsbl.2013.0828

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, *64*(10), 1673–1683. doi:10.1007/s00265-010-0981-5

Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLOS ONE*, *8*(4), e62397. doi:10.1371/journal.pone.0062397

Yamadori, A., Osumi, Y., Masuhara, S., & Okubo, M. (1977). Preservation of singing in Broca’s aphasia. *Journal of Neurology, Neurosurgery & Psychiatry*, *40*(3), 221–224. doi:10.1136/jnnp.40.3.221

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, *33*(2), 248. doi:10.1121/1.1908630

APPENDIX A

Study 1

Supplementary Methods

Experimental Procedure

Participants were told that "at this stage" (to maintain the illusion that they might meet the judges) they should base their responses only on visual characteristics of the person in the video (e.g. attractiveness, body language). Participants viewed the same six same-sex and six opposite-sex target videos in a different, randomised order. Immediately following each video, monaural audio responses of the participants were digitally recorded in a quiet room using Praat©, Version 5.2 (P. Boersma and D. Weenink, 2011; www.praat.org) on a laptop PC, with a sampling frequency of 44.1 kHz, using a ClearChat Stereo™ Headset (Logitech®, 2007), positioning the microphone about 2 cm from the participant's mouth.

Data Analysis

I use GLM because I test within-subject changes in mean scores for each acoustic parameter. As a confirmatory analysis and because the relevant unit of analysis was mean standard deviations rather than directly measured mean scores (Albrecht et al., 2013; Lifjeld, Laskemoen, Kleven, Albrecht, & Robertson, 2010), I additionally compared mean F_0 SD scores in response to attractive or unattractive targets (shown in Fig. 2.1d) using non-parametric pairwise comparisons; these are not reported here but they produced an identical pattern of significant and non-significant effects as the GLM.

Supplementary Results

Table A1. Descriptive statistics for the analysed acoustic parameters and length of the recordings, according to the target.

Measure		**Target**							
		Same Sex				**Opposite Sex**			
		Attractive		**Unattractive**		**Attractive**		**Unattractive**	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Male voices									
Mean Intensity (dB)	En.	63.50	6.33	63.31	6.97	63.71	6.71	63.10	6.55
	Cz.	77.33	8.77	77.30	9.17	77.10	9.13	77.55	7.73
Intensity SD (dB)	En.	4.34	0.60	4.42	0.80	4.69	0.82	4.54	0.73
	Cz.	4.27	0.73	4.14	0.67	4.34	0.88	4.18	0.54
Mean F0 (Hz)	En.	110.48	13.66	108.95	13.42	112.40	14.14	112.16	14.40
	Cz.	119.81	16.92	117.20	16.59	120.34	16.29	118.94	17.03
F0 SD (Hz)	En.	12.67	5.24	10.74	2.96	16.23	6.24	12.52	3.74
	Cz.	22.32	8.79	18.27	8.23	22.19	9.31	18.30	9.21
Minimum F0 (Hz)	En.	83.09	7.62	84.62	8.30	81.80	6.05	85.02	8.08
	Cz.	82.67	6.48	86.27	11.51	83.02	7.01	87.62	10.23
Length (s)	En.	16.24	4.79	19.70	8.54	17.18	5.92	17.15	4.86
	Cz.	14.08	6.90	13.01	6.41	12.92	7.22	12.63	6.68
Female voices									
Mean Intensity (dB)	En.	65.79	4.74	66.10	5.20	66.03	4.95	65.93	5.60
	Cz.	77.32	10.30	77.76	9.31	78.13	8.06	77.87	8.38
Intensity SD (dB)	En.	4.42	0.67	4.39	0.67	4.72	0.69	4.45	0.59
	Cz.	4.65	0.61	4.54	0.58	4.64	0.61	4.64	0.60
Mean F0 (Hz)	En.	190.15	19.35	193.62	19.18	191.73	20.11	190.94	19.73
	Cz.	205.46	18.38	203.15	16.56	207.52	19.25	204.75	16.22
F0 SD (Hz)	En.	40.16	8.32	34.63	9.11	39.01	9.72	38.42	8.36
	Cz.	43.48	14.41	35.77	13.26	43.30	18.21	42.16	16.55
Length (s)	En.	10.84	7.53	11.73	6.94	13.38	6.10	13.98	7.33
	Cz.	13.01	9.00	13.49	7.21	15.37	10.00	15.17	9.91

En. = English participants, Cz. = Czech participants.

Table A2. Context-dependent variation in vocal parameters: complete models including between-subject effects and within-subject effects unrelated to attractiveness.

Effect	**Vocal parameter**							
	Mean intensity		**Intensity SD**		**Mean F_0**		**F_0 SD**	
Within-subject Effect	*F*	p	*F*	p	*F*	p	*F*	p
TS	0.39	0.533	**10.39**	**0.002**	4.49	0.036	**11.85**	**<0.001**
TS x PS	0.37	0.542	0.18	0.672	1.07	0.304	0.79	0.377
TS x L	0.29	0.593	3.89	0.051	0.16	0.687	0.21	0.649
TS x PS x L	0.27	0.606	0.12	0.728	2.64	0.107	4.85	0.03
TA	0	0.99	**8.18**	**0.005**	4.88	0.029	**68.15**	**<0.001**
TA x PS	0.51	0.476	0.03	0.864	0.83	0.364	0.16	0.687
TA x TS	1.37	0.244	1.98	0.162	0.34	0.563	**9.85**	**0.002**
TA x PS x TS	1.71	0.194	0.19	0.661	3.49	0.065	**17.45**	**<0.001**
TA x L	1.28	0.261	0.01	0.921	**7.27**	**0.008**	2.11	0.15
TA x PS x L	1.41	0.239	1.98	0.163	2.22	0.139	0.02	0.9
TA x TS x L	0.46	0.5	4.08	0.046	0.92	0.339	1.28	0.26
TA x PS x TS x L	1.72	0.193	0.35	0.558	1.01	0.317	0.01	0.921
Between-subject Effect	*F*	p	*F*	p	*F*	p	*F*	p
PS	1.16	0.284	2.76	0.1	**701.95**	**<0.001**	**168.73**	**<0.001**
L	**84.9**	**<0.001**	0.41	0.524	**11.87**	**0.001**	**8.57**	**0.004**
PS x L	0.57	0.453	2.9	0.091	0.77	0.381	1.35	0.249

TS = Target sex, PS = Participant sex, L = Language, TA = Target Attractiveness. Results are from repeated-measures generalized linear models (d.f. = 1, 36 in each case) for each vocal parameter, with Bonferroni adjustment for multiple tests ($\alpha = 0.0125$). Significant effects are in bold.

Table A3. Context-dependent variation in vocal parameters: post-hoc comparison of responses to attractive versus unattractive targets.

Acoustic Parameter		**Speakers**							
		Male				**Female**			
		Same Sex		**Opposite Sex**		**Same Sex**		**Opposite Sex**	
		t	p	*t*	p	*t*	p	*t*	p
Mean Intensity (dB)	En.	0.71	0.49	**2.25**	**0.03**	-0.91	0.37	0.29	0.77
	Cz.	0.06	0.95	-0.93	0.36	-1.39	0.18	1.36	0.19
Intensity SD (dB)	En.	-0.8	0.43	1.93	0.06	0.33	0.74	**2.65**	**0.01**
	Cz.	1.5	0.15	1.13	0.27	1.48	0.15	0.03	0.97
Mean F_0 (Hz)	En.	**2.74**	**0.01**	0.28	0.78	**-2.6**	**0.01**	0.38	0.7
	Cz.	1.58	0.13	**2.69**	**0.01**	1.56	0.13	1.58	0.13
F_0 SD (Hz)	En.	**2.48**	**0.02**	**3.93**	**<0.001**	**5.26**	**<0.001**	0.50	0.62
	Cz.	**6.97**	**<0.001**	**7.18**	**<0.001**	**5.05**	**<0.001**	0.55	0.59

En. = English participants, Cz. = Czech participants. Results are from paired-samples t-tests (d.f. = 29 for English participants, and d.f. = 24 for Czech participants) for each acoustic parameter, comparing responses to attractive versus unattractive targets. Significant effects are in bold.

Study 2

Supplementary Methods

Participants

For the test using original, unfiltered voice recordings rated for attractiveness, 332 heterosexual participants were recruited in Stirling and Prague (204 judging opposite-sex recordings, 128 judging same-sex recordings) via social networks, university announcements, and the Stirling University Psychology Sign-Up System. I ensured that listeners did not speak the other language in the following ways. First, I anticipated that many Czech participants might understand some English, so during recruitment I advertised specifically for non-English speakers. Second, during the test described below, all participants (Czech and English) were asked to select how much they understood of the speech in the recordings, with options of “Nothing”, “Very little (isolated words)”, “Some (more than a few words)”, “A lot (whole sentences/meaning)” and “Everything”. Only responses from participants who answered “Nothing” were analysed; by doing this, I excluded a number of English participants who recognised some Czech (7 men, 5 women) and a much larger number of participants recruited in the Czech Republic who recognised some English (28 men, 48 women). I also excluded a further 23 UK participants who indicated they were not native English speakers. No repeated IP addresses were found.

Although in the test using original voice recordings rated for attractiveness opposite-sex listeners were older than same-sex listeners, this had no significant effect on their performance: after opposite-sex raters were assigned to an 'older' (age ≥ 25) and a 'younger' (age < 25) group via a median split, analysis revealed no significant difference in the proportion of responses to attractive individuals that 'older' (mean proportion ± SD = 0.81 ± 0.23, n = 67) and 'younger' (0.85 ± 0.21, n = 56) raters selected as more attractive (paired-samples *t*-test: $t_{121} = 0.93$, $p = 0.35$).

For the test using original voice recordings rated for friendliness, 23 participants were excluded because of not being heterosexual (1 man), or because they indicated they understood at least some content of the recordings (9 men, 13 women).

Audio samples

I used the opening portion of each recording, using all speech until the end of the sentence closest to 10 seconds into the recording. This ensured that the length of sampled responses to the attractive and unattractive targets was not significantly different (paired-samples *t*-test: $t_{39} = 0.56$, $p = 0.58$).

Supplementary Results

Table A4. Descriptive statistics for the proportion of recordings towards attractive targets that were selected as more attractive or friendly by naïve listeners.

Recordings		**Raters**							
		Attractiveness				**Friendliness**			
		OS		**SS**		**OS**		**SS**	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Male									
Original	En.	0.86	0.06	0.63	0.15	0.63	0.12	0.53	0.13
	Cz.	0.83	0.07	0.47	0.11	0.57	0.06	0.48	0.11
Filtered	En.	0.53	0.2	0.55	0.21	0.69	0.18	0.62	0.21
	Cz.	0.54	0.12	0.52	0.11	0.55	0.14	0.54	0.21
Female									
Original	En.	0.87	0.08	0.51	0.3	0.75	0.23	0.53	0.11
	Cz.	0.73	0.08	0.54	0.11	0.55	0.03	0.51	0.08
Filtered	En.	0.55	0.17	0.56	0.13	0.57	0.17	0.6	0.16
	Cz.	0.57	0.08	0.55	0.09	0.59	0.22	0.57	0.21

En. = English recordings, Cz. = Czech recordings. OS = opposite-sex raters, SS = same-sex raters.

Table A5. Perception of vocal modulation: complete models including between- and within-subject effects.

Effect	**Recordings**			
	Original	**Filtered**	**Filtered**	
Within-subject Effect	*F*	p	*F*	p
Rater Sex	**63.19**	**<0.001**	**8.15**	**0.01**
Rater Sex x Gender	0.58	0.45	0.34	0.56
Rater Sex x Language	0.16	0.69	0.70	0.41
Rater Sex x Gender x Language	1.79	0.19	1.53	0.22
Context	**10.27**	**<0.01**	1.35	0.25
Context x Gender	0.07	0.8	0.13	0.72
Context x Language	0.09	0.77	4.22	0.05
Context x Gender x Language	0.31	0.58	0.71	0.40
Rater Sex x Context	**50.93**	**<0.001**	**5.81**	**0.02**
Rater Sex x Context x Gender	0.19	0.67	0.15	0.70
Rater Sex x Context x Language	0.01	0.93	2.28	0.14
Rater Sex x Context x Gender x Language	**7.75**	**0.01**	0.79	0.38
Between-subject Effect	*F*	p	*F*	p
Gender	0.47	0.50	1.24	0.27
Language	3.30	0.08	3.06	0.09
Gender x language	1.11	0.30	0.15	0.70

Results are from repeated-measures generalized linear models (d.f. = 1, 36 in each case) for the proportion of responses to attractive targets rated as more attractive or friendly, depending on the type of recordings presented (original, filtered), with Bonferroni adjustment for multiple tests ($\alpha = 0.025$). Significant effects are in bold.

APPENDIX B

Supplementary Results

Table B1. Context-dependent variation in vocal parameters and attractiveness ratings: complete models including between-subject effects and within-subject effects unrelated to condition.

Effect	Vocal parameter										Attractiveness Ratings	
	Mean Intensity		Intensity SD		Mean F_0		F_0 SD		Minimum F_0 (male only)			
Within-subject Effect	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**
C	0.60	0.44	2.38	0.13	**279.02**	**<0.0001**	**56.26**	**<0.0001**	2.84	0.10	0.04	0.85
C * A	0.98	0.33	0.36	0.55	0.04	0.85	0.14	0.71	0.34	0.56	0.10	0.75
C * BO	0.31	0.58	1.34	0.25	0.73	0.40	0.00	0.98	0.37	0.55	0.69	0.41
C * BO * A	1.95	0.17	0.18	0.68	0.00	0.95	0.27	0.60	0.21	0.65	0.49	0.48
C * PS	0.04	0.83	2.23	0.14	**354.29**	**<0.0001**	**46.31**	**<0.0001**	.	.	1.80	0.18
C * PS * A	0.15	0.70	0.28	0.60	1.63	0.21	2.21	0.14	.	.	1.80	0.18
C * PS * BO	0.04	0.85	0.64	0.43	1.19	0.28	0.91	0.34	.	.	0.92	0.34
C * PS * BO * A	0.01	0.90	0.00	0.97	0.15	0.70	0.02	0.89	.	.	2.98	0.09
C * TA	1.20	0.28	0.85	0.36	**524.72**	**<0.0001**	**49.41**	**<0.0001**	0.47	0.50	0.01	0.93
C * TA * A	0.24	0.63	0.37	0.55	0.05	0.82	0.18	0.68	**15.37**	**<0.001**	3.13	0.08
C * TA * BO	0.70	0.40	0.98	0.33	1.08	0.30	1.13	0.29	3.56	0.07	0.22	0.64
C * TA * BO * A	0.12	0.73	1.54	0.22	0.06	0.81	0.02	0.90	0.07	0.79	0.01	0.93
C * TA * PS	0.88	0.35	0.00	0.96	**537.52**	**<0.0001**	**65.53**	**<0.0001**	.	.	0.70	0.40
C * TA * PS * A	0.03	0.86	0.41	0.53	0.29	0.59	0.00	0.98	.	.	0.22	0.64
C * TA * PS * BO	3.00	0.09	0.05	0.83	0.82	0.37	0.30	0.58	.	.	0.43	0.52
C * TA * PS * BO * A	3.98	0.05	0.47	0.50	1.12	0.29	0.89	0.35	.	.	0.70	0.40
TA	1.59	0.21	**7.50**	**<0.01**	**753.56**	**<0.0001**	**25.72**	**<0.0001**	2.22	0.15	**673.60**	**<0.0001**
TA * A	1.76	0.19	5.41	0.02	1.29	0.26	0.60	0.44	2.72	0.11	1.47	0.23
TA * BO	0.06	0.81	0.13	0.72	**14.64**	**<0.001**	0.02	0.88	**8.35**	**<0.01**	5.87	0.02
TA * BO * A	0.26	0.61	1.48	0.23	4.02	0.05	1.31	0.26	0.48	0.49	0.04	0.84
TA * PS	3.58	0.06	0.29	0.59	**621.35**	**<0.0001**	**41.13**	**<0.0001**	.	.	**29.74**	**<0.0001**
TA * PS * A	7.04	0.01	0.11	0.74	0.00	0.98	1.37	0.25	.	.	5.87	0.02
TA * PS * BO	3.76	0.06	1.40	0.24	**10.56**	**0.00**	0.28	0.60	.	.	0.01	0.92
TA * PS * BO * A	1.71	0.20	0.17	0.68	0.39	0.53	0.18	0.68	.	.	1.47	0.23
Between-subject Effect	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**	**F**	**p**	***F***	**p**
A	1.31	0.26	0.13	0.72	1.21	0.27	0.00	0.96	0.00	1.00	2.10	0.15
BO	0.35	0.56	0.23	0.63	2.70	0.10	0.62	0.43	0.01	0.92	0.14	0.71
BO * A	1.11	0.30	2.20	0.14	0.02	0.89	0.77	0.38	0.58	0.45	7.35	0.01
PS	5.58	0.02	0.01	0.93	**524.20**	**<0.0001**	**348.81**	**<0.0001**	.	.	0.78	0.38
PS * A	1.73	0.19	1.96	0.17	0.01	0.92	0.16	0.69	.	.	4.33	0.04
PS * BO	0.08	0.78	2.56	0.11	0.91	0.34	0.21	0.65	.	.	0.04	0.85
PS * BO * A	0.01	0.94	0.58	0.45	0.73	0.40	1.21	0.27	.	.	0.32	0.57

C = Condition (control, experimental), A= Androstadienone (yes, no), BO = Body Odour (HQ, LQ), PS = Participant Sex (male, female), TA = Target Attractiveness (attractive, unattractive). Results are from repeatedmeasures generalized linear models (d.f. = 1, 72 in each case) for each vocal parameter, with Bonferroni adjustment or multiple tests ($\alpha = 0.0083$). Significant effects are in bold.

APPENDIX C

Relationship Assessment Scale (RAS)

To measure relationship satisfaction, participants were asked to complete the 7-item Relationship Assessment Scale (RAS) (Hendrick, 1988). Participants were instructed to rate each phrase, using a 1 - 5 scale, where 1 = low and 5 = high. Individual scores were calculated by adding the responses to each phrase (items 4 and 7 are reverse-scored).

- How well does your partner meet your needs?
- In general, how satisfied are you with your relationship?
- How good is your relationship compared to most?
- How often do you wish you hadn't gotten into this relationship?
- To what extent has your relationship met your original expectations?
- How much do you love your partner?
- How many problems are there in your relationship?

Parental investment in children

Participants who had children were asked to complete a 11-item questionnaire about the investment in children of the other parent of their only/youngest child. Participants were asked to rate each phrase answer using a 1-9 scale, where 1 = completely does not apply and 9 = completely applies. They were asked to answer each question with respect to their only/youngest child. Individual scores were calculated by adding the responses to each phrase.

This questionnaire was based on adapted versions of the 7-item *Delight* construct of the Parental Investment in Children (PIC) questionnaire (Bradley & Whiteside-Mansell, 1997) (items 1-7), as as well as items adapted from Apicella & Marlowe (2004).

- My partner is always bragging about our child to his/her friends and family.
- My partner carries pictures of our child with him/her wherever he/she goes.
- I often find my partner is thinking about our child.
- Holding and cuddling our child is more fun to my partner than most other things he/she does.
- My partner enjoys going to places our child will enjoy.
- It's more fun for my partner to get our child something new than to get himself/herself something new.
- Most of the time when my partner goes out of the house he/she takes our child with him/her.
- I believe my partner gives our child a lot of attention.
- My partner spends a lot of time with our child.
- My partner is/was/will be involved with our child's schoolwork.
- My partner is the best possible parent.

APPENDIX D

Supplementary Methods

Table D1. Independent ratings of target attributes: images, employee testimonials, names, and job titles.

		Facial image							
		Attractiveness		Dominance		Prestige		Age	
GLM	Pairwise Comparisons	***F***	**p**	***F***	**p**	***F***	**p**	***F***	**p**
		5.10	**0.010**	**19.59**	**<0.001**	**6.70**	**0.003**	**14.48**	**<0.001**
	A vs D		0.394		**<0.001**		0.383		**<0.001**
	A vs P		0.085		**0.006**		**0.020**		**0.001**
	D vs P		**<0.001**		**0.003**		**0.006**		0.144

		Testimonial			
		Dominance		Prestige	
GLM	Pairwise Comparisons	***F***	**p**	***F***	**p**
		40.60	**<0.001**	**41.41**	**<0.001**
	A vs D		**<0.001**		0.189
	A vs P		**<0.001**		**<0.001**
	D vs P		**<0.001**		**<0.001**

		Name			
		Dominance		Prestige	
GLM	Pairwise Comparisons	***F***	**p**	***F***	**p**
		10.94	**<0.001**	**22.90**	**<0.001**
	A vs D		**<0.001**		0.611
	A vs P		**<0.001**		**<0.001**
	D vs P		0.737		**<0.001**

		Job Title			
		Dominance		Prestige	
GLM	Pairwise Comparisons	***F***	**p**	***F***	**p**
		30.86	**<0.001**	**23.03**	**<0.001**
	A vs D		**<0.001**		0.535
	A vs P		**<0.001**		**<0.001**
	D vs P		**0.008**		**<0.001**

A = Average target, D = Dominant target, P = Prestigious target. Results are from repeated-measures generalized linear models (d.f. = 2, 42 in each case) for each rated attribute. Significant effects are in bold.

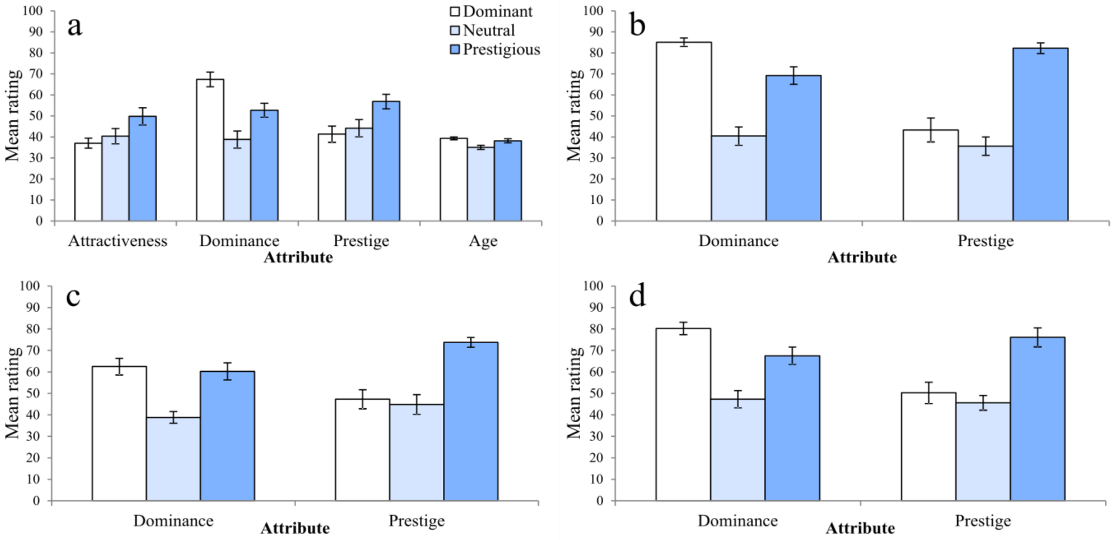


Figure D1. Ratings of targets' attributes split by target (average: white bars; dominant: light grey bars; prestigious: dark grey bars) and attribute rated. (a) Facial images; (b) Employee testimonials; (c) Names; (d) Job titles. Bars represent mean ± 1 s.e.m.

Experimental procedure

At the end of the experiment, participants were asked to write what they thought the purpose of the study was. 39.6% of the participants (7 men, 12 women) had a relatively accurate idea about my manipulation (i.e. they understood I was manipulating the dominance and prestige of the targets). However, it is important to highlight that because this was the last part of the experiment, participants responded to this question after rating all the targets for both dominance and prestige (which should have given them a good idea of what the real purpose of the experiment was). No participant, however, realised that the main focus of the study was to perform an acoustic analysis of their voices.

Target stimuli

21 raters (14 men, mean age ± SD = 30.7 ± 9.6; 7 women, 35.4 ± 10.1) independently judged each attribute of the targets (names, job titles, testimonials, and faces).

Supplementary Results

Table D2. Context-dependent variation in vocal parameters: full models including intensity parameters and between-subject effects.

Effect	**Vocal Parameter**														
	Mean intensity			Intensity SD			Mean F0			F0 SD			Min F0		
Within-subject Effect	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**
T	0.596	1.25, 51.45	0.480	0.006	2, 82	0.994	2.484	2, 82	0.090	1.195	1.65, 67.48	0.302	0.076	2, 82	0.927
T * PD	0.608	1.25, 51.45	0.474	0.130	2, 82	0.878	**4.560**	**2, 82**	**0.013**	1.136	1.65, 67.48	0.319	2.331	2, 82	0.104
T * PP	0.105	1.25, 51.45	0.804	0.035	2, 82	0.965	0.214	2, 82	0.808	0.631	1.65, 67.48	0.506	0.427	2, 82	0.654
T * PS	0.318	1.25, 51.45	0.626	1.013	2, 82	0.368	1.526	2, 82	0.223	3.078	1.65, 67.48	0.062	0.088	2, 82	0.916
Q	0.044	1.46, 59.72	0.912	**4.327**	**1.65, 67.76**	**0.023**	**6.775**	**1.50, 61.61**	**0.005**	3.078	2, 82	0.051	0.462	2, 82	0.632
Q * PD	0.637	1.46, 59.72	0.484	1.270	1.65, 67.76	0.283	0.404	1.50, 61.61	0.611	1.390	2, 82	0.255	2.644	2, 82	0.077
Q * PP	0.067	1.46, 59.72	0.882	1.811	1.65, 67.76	0.177	**6.439**	**1.50, 61.61**	**0.006**	2.606	2, 82	0.080	0.024	2, 82	0.976
Q * PS	1.674	1.46, 59.72	0.201	0.506	1.65, 67.76	0.570	**14.394**	**1.50, 61.61**	**0.000**	**15.310**	**2, 82**	**0.000**	1.443	2, 82	0.242
T * Q	0.212	1.61, 65.83	0.760	0.885	3.27, 134.02	0.458	**3.379**	**2.72, 111.33**	**0.025**	0.841	4, 164	0.501	2.428	2.52, 103.13	0.080
T * Q * PD	0.353	1.61, 65.83	0.657	0.930	3.27, 134.02	0.435	**3.984**	**2.72, 111.33**	**0.012**	1.265	4, 164	0.286	2.416	2.52, 103.13	0.081
T* Q * PP	0.159	1.61, 65.83	0.806	0.654	3.27, 134.02	0.595	1.412	2.72, 111.33	0.245	0.532	4, 164	0.712	0.717	2.52, 103.13	0.521
T * Q * PS	1.095	1.61, 65.83	0.329	2.179	3.27, 134.02	0.088	2.015	2.72, 111.33	0.122	1.358	4, 164	0.251	1.763	2.52, 103.13	0.168
Between-subject Effect	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**	***F***	***d.f.***	**p**
PD	0.974	1, 41	0.329	0.001	1, 41	0.977	1.339	1, 41	0.254	0.367	1, 41	0.548	1.289	1, 41	0.263
PP	0.147	1, 41	0.703	3.013	1, 41	0.090	1.388	1, 41	0.246	0.466	1, 41	0.499	0.579	1, 41	0.451
PS	1.914	1, 41	0.174	3.405	1, 41	0.072	**454.429**	**1, 41**	**0.000**	**69.667**	**1, 41**	**0.000**	**128.703**	**1, 41**	**0.000**

T = Target (neutral, dominant, prestigious), Q = Question, PD= Participant Dominance, PP = Participant Prestige, PS = Participant Sex (male, female), A = Average Target, HS = High Status Targets (dominant, pretigious), D = Dominant Target, P = Prestigious Target. Results are from repeated measures generalized linear models for each vocal parameter . Significant effects are in bold.