Acoustic determiners of vocal attractiveness go well beyond apparent talker size

Molly Babel[1], Joseph King[2], Grant McGuire[2], and Teresa Miller[2]

[1]Department of Linguistics, University of British Columbia, [2]Department of Linguistics, University of California, Santa Cruz

Corresponding Author:
Molly Babel
2613 West Mall
University of British Columbia
Vancouver, British Columbia V6T 1Z4
Canada
mbabel@interchange.ubc.ca
Tel: 604 822 6322, Fax: 604 822 9687

ABSTRACT:

This study reports male and female Californians' ratings of vocal attractiveness for 30 male and 30 female voices reading isolated words. While ratings by both sexes were highly correlated, males generally rated fellow males as less attractive than females did, but both females and males had similar ratings of female voices. Detailed acoustic analyses were conducted on the voices and regression models were fitted to assess what acoustic properties predicted the attractiveness ratings. These models demonstrated that higher second formant frequencies for high vowels, breathier voice, reduced pitch variance, and longer durations were all important to the more attractive female voices. For the attractive male voices, shorter durations, higher vowels, and higher F2 in /u/ were characteristic. These results suggest that judgments of vocal attractiveness are more complex than previously thought.

1. Introduction

The voice is a rich source of information for listeners. In addition to functioning as the medium of communication in oral language, in its non-linguistic role, the human voice has the ability to convey biological information like sex (e.g., Lass, Hughes, Bowyer, Waters, & Bourne, 1976) and age (e.g., Ptacek & Sander, 1966); physiological details such as height and weight for men (van Dommelen & Moxness, 1995); social classifications such as race (Walton & Orlikoff, 1994); and emotional states (Scherer, Banse, & Wallnott, 2001). The attractiveness of a particular voice is potentially related to a number of these talker-specific qualities. Previous work on vocal attractiveness has used a small selection of acoustic-phonetic measures that are related simply to talker size to predict listeners' judgments of attractive voices. In this study, we employ a larger range of phonetic measures related to both the size of talkers' laryngeal source and supralaryngeal cavity and non-physiological stylistic aspects of spoken language measureable from the acoustic signal to study the subjective vocal attractiveness ratings of thirty talkers.

Attractiveness, as a general topic, is one of interest as human physical attraction drives the selection of mate partners (i.e., Grammer, Fink, Møller, & Thornhill, 2003; Rhodes, 2006). This theory is supported by research indicating that women's preferences for masculine-sounding men are enhanced during the fertile menstrual cycle phase (Feinberg et al., 2006). Previous research has found that listeners share high levels of agreement when rating the attractiveness of voices and that such ratings may relate to other qualities deducible from the voice such as dominance (Zuckerman & Driver, 1989). Additionally, some researchers have found evidence for a relationship between facial and vocal attractiveness (Collins & Missing, 2003; Saxton, Caryl, & Roberts, 2006; Riding, Lonsdale, & Brown, 2006). Beyond mate selection, judgments of attractiveness are important in everyday interaction as physically attractive people are judged to be more socially desirable and to get better jobs (Dion, Berscheid, & Walter, 1972), in addition to being more persuasive (Chaiken, 1979).

There has been a relatively small amount of research examining the specific acoustic properties on which listeners base their vocal attractiveness judgments. One acoustic feature, however, that is well known to correlate with vocal attractiveness is fundamental frequency. Fundamental frequency (f0) is an acoustic measure of the rate of vibration of the vocal folds; it is the acoustic property that listeners perceive as vocal pitch. Vocal folds with a larger mass tend to vibrate at a lower rate, typically giving males their lower pitch, whereas the vocal folds of females, being smaller and lighter, vibrate at a higher rate, providing a generally higher-pitched voice. Within English the general consensus is that an average or slightly higher-than-average overall f0 is considered more attractive for female voices and that an average or slightly lower-than-average voice is more attractive in male talkers (Tuomi & Fisher, 1979; Apple et al., 1979; Zuckerman & Miyake, 1993, Riding et al., 2006; Saxton et al., 2006). This preference for lower-pitched male voices holds for both adolescent and adult-aged females, although not female children (Saxon et al. 2006). This finding is assumed to be a reinforcing or an exaggeration of the average laryngeal differences between males and females and is thought to have cross-cultural relevance since the degree of size difference between males and females varies across cultures (Ohala, 1983, 1994; Ohara, 1992; see Johnson [2006] for a discussion of the role of male and female height differences across languages on male and female differences in formant frequencies). For example, van Bezooijen (1995) finds that pitch differences between males and females are greater in Japanese than in Dutch, but she confirms that the differences in gender stereotypes and expectations by culture are important mediators of such an effect. In a study of Hadza speakers living as hunter-gatherers in Tanzania, lower-pitched males had higher levels of reproductive success, suggesting the existence of selectional pressures for low-pitched male voices (Apicella, Feinberg, & Marlowe, 2007).

A few studies have examined the role of apparent larynx and vocal tract size in judgments of vocal attractiveness for male (Feinberg, Jones, Little, Burt, & Perrett, 2005) and female voices (Feinberg, DeBruine, Jones, & Perrett, 2008). The methodology involved taking acoustic measurements and making acoustic manipulations on English vowels produced in isolation. The acoustic manipulations involved making modifications to the fundamental frequency and the formant frequencies to modify apparent larynx and vocal tract size, respectively. With male voices, Feinberg et al. (2005) found that voices with larger apparent voices and larynxes were rated as older, more masculine, and larger sounding. Voices with lowered fundamental frequencies were also rated as more attractive, but no effect was found in manipulating formant frequencies. In a large-scale study of female voices, Feinberg et al. (2008) found that higher pitched voices were rated as more attractive than lower-pitched female voices. A subset (n=15) of the female voices – five from a low pitch group (200 Hz), an average group (220 Hz) and a high group (240 Hz) – were selected for Feinberg and colleagues' second study which involved manipulating f0 so as to modify apparent larynx size. Using a forced-choice paradigm, male listeners judged attractiveness, age, or femininity from the vocal samples. Listeners judged talkers with apparent smaller larynxes to sound more feminine and younger. Voices with raised pitch were rated as more attractive by male listeners than those without raised pitch; this effect was strongest for the voices in the low pitch group.

Some suggest that attractiveness ratings may relate to other qualities deducible from the voice, such as dominance (Zuckerman & Driver, 1989). Puts, Gaulin, and Verdolini (2007) examined the perception of social and physical dominance in male voices by independently manipulating f0 and the dispersion of formant frequencies. Both larger apparent vocal tract length and larger apparent larynx size resulted in higher dominance ratings, but the effect of apparent vocal tract length affected judgments of physical dominance more than social dominance. An important clarification is necessary here. In Puts and colleagues' study, formant dispersion was used as an estimate of apparent vocal tract length. While this measure would ideally prove an accurate proxy for vocal tract length if the vocal tract functioned as an unobstructed tube of uniform dimensions, such a tube is unrealistic in speech production and can only be somewhat obtained during the production of a schwa. Crucially, it is unclear from Puts et al. (2007) what type of speech sample was used to calculate the formant dispersion measurements. Moreover, González (2004) provides data from Spanish speakers that illustrates the relationship between formant frequencies and body size is weaker than previously assumed.

One additional phonetic characteristic has been shown relevant to judgments of vocal attractiveness: voice-onset-time (VOT). VOT is a temporal descriptor of oral stops when they are followed by a voiced sound (e.g., a vowel) that measures the duration, either positive or negative, of the lag interval between the onset of vocal fold vibration and the release of the oral closure. VOT has been demonstrated to vary during women's menstrual cycle such that those who are at their reproductive peaks have longer VOTs than when at their lowest fertility levels (Whiteside, Hanson, Cowell, 2004; Wadnerkar, Cowell, & Whiteside, 2006). This is presumably related to the similar observation that women at reproductive peaks of their cycle are rated as more vocally attractive (Pipitone & Gallup, 2008). These results suggest that, perhaps, measures of speech clarity influence attractiveness judgments as well.

To summarize, there is evidence that perceived pitch plays a significant role in judgments of vocal attractiveness and there is some indication that other measures, such as VOT, may also influence judgments. The voice spectrum is very complex and many other phonetic characteristics not previously included in studies of vocal attractiveness could contribute significantly to such judgments. Presumably any acoustic feature that may signal sex or social differences may be a significant predictor of vocal attractiveness. In the experiment described below, we present evidence that various acoustic qualities that are potentially related to a talker's size, health, and membership in a community contribute to judgments of vocal attractiveness. To do this, we used recordings of monosyllabic words, which we see as an improvement over studies that examined single vowels (e.g., Feinberg et al., 2005, 2006), as real word production is more ecologically valid for both the talkers and the listeners. In terms of pinpointing aspects of the acoustic signal that cue judgments of vocal attractiveness, single words are more appealing than full sentences, as they allow for a more controlled acoustic analysis.

2. Methods

2.1 Stimuli

As part of a previous study (Babel, 2009), 30 male and 30 female voices were recorded reading 50 low frequency monosyllablic words (Baayen, Piepenbrock, & van Rijn, 1993), which are shown in Table 1. Female talkers (mean age 24.2, range 18-57) and male talkers (mean age 24.1, range 18-47) did not differ significantly in age (χ (17) = 17.7, p = 0.4). Select words containing the vowels /i a u/ were chosen for the current experiment because these sounds typically represent the maximum dispersion of the first and second formant frequencies of a talker's acoustic-phonetic vowel space.

/u/	/i/	/a/	/æ/	/o/
bloom	breeze	clock	bask	close
boot	cheek	clot	bat	coat
doom	**deed**	**cot**	mask	comb
dune	freak	**pod**	nag	foal
glue	**key**	**sock**	smash	hone
hoop	**peel**	**sod**	snap	mote
pool	sneeze	spawn	tap	soap
tool	**teal**	stock	vat	toad
toot	teethe	**tot**	wag	tone
zoo	**weave**	wad	wax	woe

Table 1. Words recorded by each talker by vowel. Bolded words were used in the current experiment.

Sixty talkers read the words in Table 1. These were all native speakers of American English with no reported speech or hearing problems and were compensated $10. Words were presented in 36-point font on the middle of the screen in a random order. Recordings were made directly to a computer hard drive at 44.1kHz sampling rate with a head-mounted AKG C520 microphone positioned three inches from the talker's mouth in a sound-attenuated room. From the chosen vowels, a subset of 15 were used for this experiment (bolded in Table 1). All tokens normalized to have the same RMS amplitude and had silence trimmed by hand from the beginning and end of each file.

2.2 Procedure

Each trial consisted of each of the 15 tokens from a single voice presented sequentially in random order with 500ms between each sound file. Subjects listened to the voices over headphones at about 70dB in a sound-attenuated booth. After the presentation of the fifteenth token subjects were asked to rate the attractiveness of the voice on a scale from 1-9, where 1 is unattractive and 9 is very attractive. Subjects could only respond after all tokens were presented and had an unlimited time to respond. 1000ms after the response was logged, the tokens from the next voice were presented. The order of voices was randomized for each subject and the experiment lasted approximately 35 minutes.

2.3 Subjects

Twenty-five native speakers of Californian English (12 females, 13 males) served as raters and received course credit or $10 for compensation. All reported normal hearing and had lived in California from toddlerhood.

3 Results: ratings by sex

All listeners' judgments for each talker were averaged and two Spearman rank correlation coefficients were computed to assess the relationship between male and female ratings of male voices and male and female ratings of female voices. Results demonstrate here was a strong correlation between both sex's ratings for male voices (Spearman 1113, rho = 0.75, $p < 0.001$) and an even stronger one for female voices (Spearman = 582, rho = 0.87, $p < 0.001$). This relationship is shown in Figure 1.

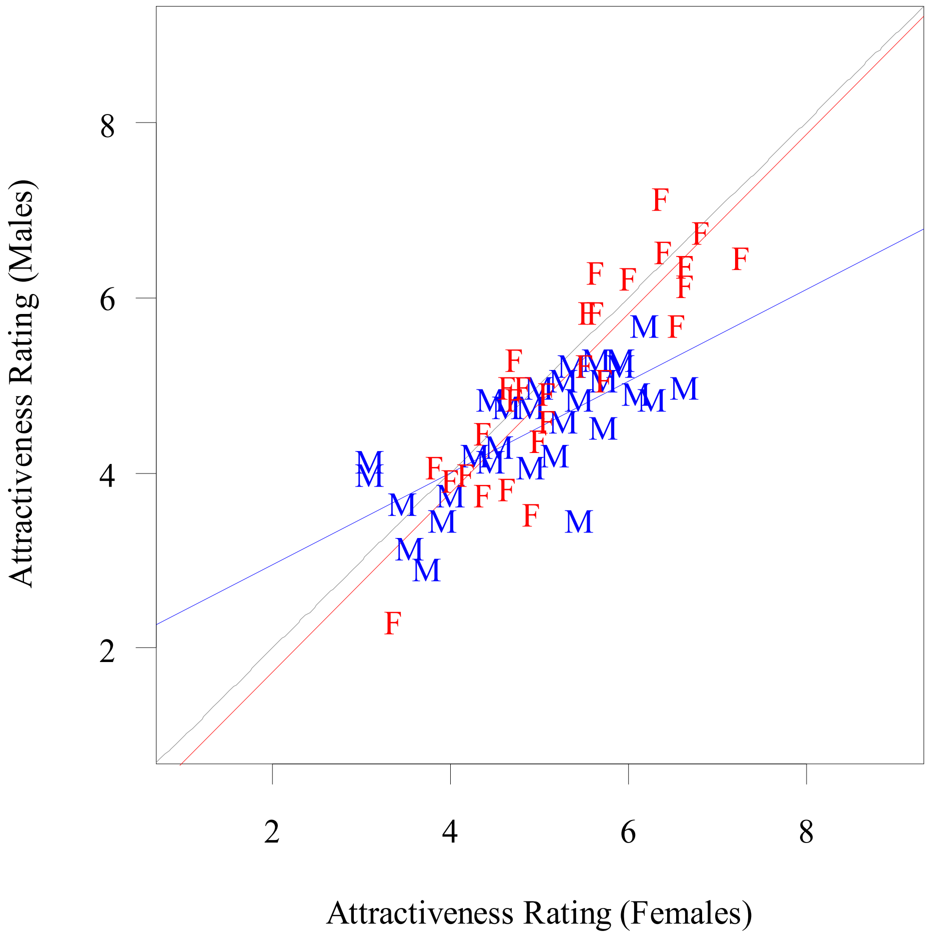

Figure 1. Correlations between Male and Female raters for Male (M) and Female (F) voices.

The correlation shown in Figure 1 demonstrates three main points. First, male and female raters agreed strongly on which voices are attractive and which are not for both sexes. Second, this agreement is strongest for the female voices; there is less agreement on the male voices. Finally, while males and females give female voices much the same attractiveness ratings, males rank fellow male voices as less attractive as a group than females do.

4 Results: acoustic features and attractiveness

Duration, fundamental frequency (F0), and formant frequency were measured for each subject using Praat (Boersma and Weenink, 2009). Measures of laryngeal setting were made using VoiceSauce (http://www.ee.ucla.edu/~spapl/voicesauce/index.html). Prior to analysis formant frequencies were normalized to the Bark scale. An exhaustive list of the acoustic measures is given in Table 2.

Factor	***Measure***	***Description***
Area	Area of the triangle in F1 / F2 space formed by /i a u/ for each talker (Euclidean distance)	Size of vowel space
Duration	Duration (ms) from onset of acoustic energy > 500Hz to end of acoustic energy >500Hz	Speech rate
meanF0	Mean F0 for all tokens (Hz)	Average pitch
varF0	Mean variance of F0 across all tokens from the talker (Hz)	Pitch range
if1bavg	Mean first formant (F1) for /i/ tokens (Bark)	Vowel quality; height of /i/; back cavity resonance
uf1bavg	Mean first formant (F1) for /u/ tokens (Bark)	Vowel quality; height of /u/; back cavity resonance
af1bavg	Mean first formant (F1) for /a/ tokens (Bark)	Vowel quality; height of /a/; front cavity resonance
if2bavg	Mean second formant (F2) for /i/ tokens (Bark)	Vowel quality; backness of /i/; front cavity resonance
uf2bavg	Mean second formant (F2) for /u/ tokens (Bark)	Vowel quality; backness of /u/; front cavity resonance
af2bavg	Mean second formant (F2) for /a/ tokens (Bark)	Vowel quality; backness of /a/; back cavity resonance
H1-H2	Amplitude of the first harmonic minus the second (dB)	Voice quality, open quotient; larger positive value = breathy, negative = creaky
H1-A3	Amplitude of the first harmonic minus the amplitude of the 3rd formant (dB)	Voice quality; abruptness of closure; larger positive value =

		breathy, low value = creaky

Table 2. Factor names and the acoustic measures used in the analysis.

In order to determine which factors were the most important in listeners' judgments of vocal attractiveness, two linear regression models were fitted with the predictor variables listed in Table 2. One model had listeners' attractiveness judgments of the male voices as its dependent variable and the other with the female voices. The sex of the rater was not taken into account in these models due to the strong agreement between sexes when rating each sex.

Variance inflation factors (VIFs) were calculated for each predictor. Most of the F1/F2 values as well as Area had VIFs above 5, a common cut-off value. A visual inspection of the correlations found that Area was highly collinear with several of the vowel measures, notably uf2bavg, consequently it was removed from further analyses. Recalculation of the VIFs showed that all were below 5 with the exception of if1bavg which was 5.3 in the female model. This was deemed to be sufficiently close to the criterion for continued inclusion in the model.

The variables for the final models were chosen using a backwards selection procedure with a criterion of $p < 0.15$. Following this procedure, the two final models were then run with the remaining predictors. Both the male voice and female voice models were significant (male: $F[7,22]=12.2$, $R^2 = 0.80$, $p < 0.001$; female: $F[5,24] = 9.9$, $R^2 = 0.67$, $p < 0.001$). The final predictors for each model and their values are listed in Table 3a,b.

MALES	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	6.3	2.24	2.82	0.010

Duration	-3.45	1.12	-3.08	0.005
uf2bavg	0.3	0.1	3.12	0.005
if1bavg	-1.73	0.58	-2.97	0.007
uf1bavg	-1.31	0.59	-2.22	0.037
af1bavg	0.53	0.19	2.78	0.011
H1-H2	-0.11	0.05	-2.08	0.050
H1-A3	0.1	0.03	3.73	0.001

Table 3a. Predictors for the regression model with male voices.

FEMALE	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	-9.92	3.59	-2.77	0.01
Duration	4.46	2.55	1.75	0.09
varF0	-0.05	0.02	-2.51	0.02
if2bavg	0.45	0.27	1.65	0.11
uf2bavg	0.46	0.16	2.92	0.01
H1A3	0.07	0.04	1.94	0.06

Table 3b. Predictors for the regression model with female voices.

The predictors in the model demonstrate several interesting characteristics. Both sexes had Duration and uf2bavg as predictors. For the males, shorter durations correlated with more attractive voices while the opposite was true for females. For both groups an/u/ with higher second formant frequencies was more attractive than /u/ vowels with lower F2 values; this was especially true for female voices.

For the male voices all the predictor variables for height were significant such that a lower F1 corresponded to a more attractive voice for /i u/ and the opposite for /a/. As the first formant for /i u/ is largely determined by the back cavity, the raters have preferred a larger back cavity for males. The results for the F1 of /a/ are somewhat more difficult to interpret as the model indicated a positive correlation, i.e. a smaller front cavity, yet the graphed correlation was clearly negative. We will address this conundrum in the conclusions.

Both voice quality measures were significant for males with H1-A3 being the stronger predictor. The negative correlation for H1-H2 indicates that a smaller open quotient was more attractive for males, while the large positive H1-A3 measure indicates that a lower velocity of vocal fold closure (i.e. larger difference between H1 and A3), or breathier phonation, is more attractive. The latter was also true for female voices.

Unique to the female voices, however, was the difference in F0. This was significant, but in an unexpected way—female voices with a smaller pitch range were perceived as having a more attractive voice. The final significant predictor unique to the female voices was higher F2 values for /i/, which patterned similarly to the F2 of /u/ measure.

5. Discussion

The first major finding of this study is that both sexes largely agree with each other when rating vocal attractiveness; that is, both females and males know what an attractive female and male voice sounds like. While this agreement is nearly one to one for female voices, males are reluctant to give fellow male voices high attractiveness ratings; we suggest that this is due to cultural sexual norms relating to masculinity and perceived sexuality.

The second major finding of this research relates to the acoustic predictors of what makes a voice attractive. In using a wide variety of acoustic measures, we found several parameters predict listener judgments. These parameters fit into measures that generally relate to perceived talker size, talker health, and talker membership in a speech community. We discuss our results with respect to each of these contributions in turn.

The attractiveness of males with lower first formant frequencies indicates that apparent vocal tract size matters in judgments of vocal attractiveness. For most vowels, the first formant frequency is an indicator of back cavity (i.e., the length of the portion of the vocal tract behind the articulatory constriction) length (Fant, 1970: 121), which is the portion of the vocal tract that differs most significantly across genders as a result of the lowering of the larynx in males during puberty (Fitch & Giedd, 1999). According to our regression model, lower first formant frequencies for /i/ and /u/ were judged as more attractive for males, but higher first formant frequencies for /a/ were more attractive, despite the fact that when F1 of /a/ was examined separately, lower formant values tended toward more attractive judgments. The switching of signs in a regression is classic symptom of colinearity in the model. Some degree of colinearity is inevitable in a model that incorporates formant frequency measurements, wherein that a single talker's formant values will always be somewhat correlated as they came from the same vocal tract, despite the fact that each vowel's exact positioning within the acoustic space is, in theory, independent. The issue is then, why would the sign switch for F1 of /a/ and not that of the other vowels? An explanation for this is that while first formant frequencies are generally a reliable indicator of back cavity length, this relationship is poorest with vowels like /a/ that have front and back cavities of near identical length (Fant, 1970).

For both males and females, the voice quality measures patterned in a uniform direction: modal to breathy phonation was rated as more attractive than creaky phonation. Creaky voice qualities can often be associated with excessive smoking or drinking habits, in addition to more temporary ailments such as the common cold or laryngitis (Laver, 1968). Talkers in this study whose voices exhibited healthy vocal qualities were rated as more attractive.

Unlike previous explorations of vocal attractiveness this study points to the importance of local socio-phonetic cues to dialect. Higher second formant frequencies for /u/ were more attractive for both male and female talkers. This pattern of /u/-fronting is characteristic of Californians, especially younger females (Hagiwara, 1997; Eckert, 2008; Aiello, 2010), and here it was found to be important to the attractiveness of the voice. A result reported by Feinberg et al. (2005) may be related. Despite the fact that Feinberg and colleagues did not find an overall relationship between apparent vocal tract length and attractiveness ratings, females with increased body size were more likely to rate men with longer apparent vocal tracts as more attractive. This finding is perhaps related the current finding that higher second formant frequencies in /u/ are perceived as more attractive to native speakers of California English who also have higher F2 values for this vowel, compared to most other varieties of North American English. We suggest that this finding of being attracted to talkers who are similar to your own speech patterns is akin to the recurrent finding that perceivers have a preference for average faces (Rhodes, 2006). Essentially, this is a measure of fitting into a community.

The divergent duration results make sense in the context of the literature on male ~ female differences in that males typically have shorter durations than females (Byrd, 1994; Simpson, 2009). Judging male voices with shorter durations and female voices with longer durations as attractive is, then, suggestive that attractiveness judgments are mediated by what is considered normal or stereotypical for a group.

Given previous findings, the lack of and effect for mean F0 is surprising and could be due to two possibilities. First, it may simply be only a minor factor in judgments of attractiveness for Californians. This is in line with other studies that the role of F0 in attractiveness is highly variable by culture (van Bezooijen, 1995). Alternatively, it could be a task effect; the talkers were reading isolated words with list intonation. A narrative or sentence reading task could be more appropriate for exploring the role of mean F0. While the overall F0 of the talkers was not significant, the variance was in an unexpected direction for females. This puzzling result could, again, indicate a difference among Californians and a shift in preference away from the stereotypical high-variance female voice.

Overall, this study demonstrates that male and female perceptions of attractiveness are closely related and are mediated by cultural factors, including dialect. Further research is needed to determine what kinds of characteristics, if any, are culturally universal. A related question for further research is the relationship between subjective judgments of attractiveness and gender stereotypicality — i.e. to what extent do attractive vocal features deviate from stereotypical ones (cf. Feinberg et al., 2008). Crucially, the results of this study suggest that vocal attractiveness, like other measures of attractiveness in other domains, is multi-dimensional in nature and involves the evaluation of multiple traits. The qualitative traits listeners' deemed attractive illustrate the role of a talker's physical size and health, as well as, crucially, community membership in evaluation of an attractive voice.

Acknowledgements

Author order is alphabetical. The voices used in this study were collected for the first author's dissertation (Babel 2009) and further information can be found there. Thanks are due to Devin Tankersly and Maia Spencer for help with coding data. Further laboratory assistance was provided by Cynthia Dorantes, Madeline Hannan-Leith, Wini Murphy, Jennifer Parra, and Sarah Rittenour.

References

Aiello, A. (2010). *A phonetic examination of California*. Unpublished MA Thesis, University of California, Santa Cruz.

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

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

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database (CDROM). Linguistic Data Consortium, University of Pennsylvania.

Babel, M. (2009). *Phonetic and Social Selectivity in Speech Accommodation*. Ph.D. Dissertation, University of California, Berkeley.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1.20) [computer program]. Institute of Phonetic Sciences, Amsterdam.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39-54.

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, 37, 8, 1387-1397.

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

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

Eckert, P. (2008). Where do Ethnolects Stop? *International Journal of Bilingualism*, 12.25–42.

Fant, G. (1970). *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations.* Mouton de Gruyter.

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

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

Fitch, W. F. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106 (3), 1511-1522.

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

Grammer, K., Fink, B., Møller, A.P., Thornhill, R. (2003). Darwinian aesthetics: sexual selection and the biology of beauty. *Biological Review*, 78, 385-407.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America 102*, 655-658.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics, 43*, 485–99.

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

Laver, J. (1968). Voice quality and indexical information. *International Journal of Language and Communication Disorders*, 3 (1), 43-54.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag. 189-216.

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

Ohara, Y. (1992). Gender dependent pitch levels: A comparative study in Japanese and English. In K. Hall, M. Bucholtz, & B. Moonwomon (Eds.), *Locating power: Proceedings of the Second Berkeley Women and Language Conference*, Vol. 2, 468-477.

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

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research, 9*(2), 273-277.

Puts, D. A., Gaulin, S.J.C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, 27, 283-296.

Puts, D. A., Hodges, C.R., Cárdenas, R.A., & Gaulin, S.J.C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions

Acoustics of vocal attractiveness

among men. *Evolution and Human Behavior*, 28, 340-344.

Rhodes, G. (2006). The Evolutionary Psychology of Facial Beauty. *Annual Review of Psychology*, 57, 199-226.

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, 55-61.

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

Scherer, K., Banse, R., & Wallbott, H. (2001). Emotional inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 1, 76-92.

Schnieder, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime: User's Guide, version 1.0, Psychology Software Tools.

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

Tuomi, S. K. & Fisher, J.E. (1979). Characteristics of Simulated Sexy Voice. *Folia phoniatrica,* 31, 4, 242-9

Van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38, 253–65.

van Dommelen, W. & Moxness, B. (1995). Acoustic parameters in Speaker Height and Weight Identification: Sex-Specific Behaviour. *Language and Speech*, 38, 3, 267-287.

Wadnerkar, M., Cowell, P., & Whiteside, S. (2006). Speech across the menstrual cycle: A replication and extension study. *Neuroscience Letters*, 408, 21-24.

Walton, J. & Orlikoff, R. (1994). Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech & Hearing Research*, 37, 4, 738-745.

Whiteside, S. Hanson, A., Cowell, P. (2004). Hormones and temporal components of speech: sex differences and effects of menstrual cyclicity on speech. *Neuroscience Letters*, 367, 44–47.

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