Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Attractiveness of male speakers: effects of voice pitch and of speech tempo

Hugo Quené, Geke Boomsma, Romée van Erning

Utrecht inst of Linguistics OTS, Utrecht University, the Netherlands

h.quene@uu.nl

Abstract

Men with lower-pitched voices tend to be rated as more attractive by female listeners; this tendency has been attributed to female sexual selection. Males do not only speak with a lower pitch than females, however, but they also tend to speak at a faster tempo. Therefore this study investigates whether speech tempo also affects the subjective attractiveness of male speakers for female listeners. To this end, sentences read by 24 male speakers were changed in relative tempo (factors 0.85, 1.00 and 1.15) and in overall pitch (-1.5, 0, $+1.5$ semitone). Ratings of attractiveness by female heterosexual listeners show significant effects of both tempo and pitch. The pitch effect interacts with the relative attractiveness of a speaker's voice, both within listeners and between listeners. The tempo effect however is unrelated to the speaker's overall attractiveness, which suggests that the effects of pitch and of tempo may arise through different causal mechanisms. In conclusion, female listeners rate a male speaker as more attractive if his pitch is lower and his tempo faster. Therefore both tempo and pitch may be relevant for speech-based sexual selection of males by females.

Index Terms: pitch; f0; tempo; speech rate; attractiveness; sexual selection

1. Introduction

Male and female speakers differ in their average fundamental frequency (f_0, perceived as pitch), viz. typically about 110 Hz for males and 205 Hz for females [1, 2, 3]. This large and significant difference in f_0 develops during puberty, which in itself suggests that it may serve a sexual function. Adult males' voice pitch is reportedly related to the speaker's level of testosterone [4, 3] and his self-reported number of children [5]. Because the testosterone level is related to masculinity, a male speaker's pitch may indicate his health and physical dominance. Female listeners may therefore use voice pitch to assess the speaker's physical suitability for producing and protecting offspring, i. e., in sexual selection via female choice of mate [6]. Indeed, ratings of attractiveness by female listeners are correlated with the male speaker's f_0 [7], and experiments have confirmed that manipulations of f_0 influence these attractiveness ratings [8].

Males do not only speak with a lower f_0 than females, however, but they also tend to speak at a faster speech rate or tempo than females (about 5% faster) [9, 10]. This difference too may be related to male dominance, because the faster tempo presumably indicates the speaker's cognitive abilities and motor skills through his speaking. The faster tempo requires more physical energy [11], even more so because the male speech organs have somewhat more mass than the females', and it also requires more cognitive effort in linguistic planning and motor control. Indeed, faster speakers tend to be rated as more convincing, reliable, empathic, serious, active and competent [12, 13]. Presumably, then, female listeners also use a male speaker's tempo, to assess his motor skills and cognitive suitability as a potential mate.

This study aims to replicate previous findings on female preference for male voices with lower *pitch*, and to extend that work by investigating the presumed female preference for male speakers speaking at a faster *tempo*. In addition, we are interested in the interaction between the two factors. From a sexual selection perspective, a speaker who combines a low pitch with a fast tempo may be most attractive (and vice versa), because this combination would suggest a healthy physique as well as good motor and cognitive capabilities.

The experiment reported below addresses these questions by manipulating Dutch sentences in tempo and in pitch, and then asking female listeners to rate the attractiveness of the speaker. A speaker's voice is presented together with a portrait photo, in order to distract listeners from the phonetic manipulations of the speech stimuli. Although the experiment focuses on ratings of attractiveness of male speech by heterosexual female listeners, this was obfuscated in the actual experiment by recruiting both male and female listeners, and by presenting both male and female speech to them, with subsequent selection of targeted stimuli and participants.

2. Methods

2.1. Participants

Participants were 166 students at Utrecht University, from 8 undergraduate course groups. In order to conceal the research topic (knowledge of which might have increased response bias), both targeted participants and others were tested and subsequently presented with a questionnaire. Targeted participants were heterosexual women of the approximate age range of the selected speakers (see §2.2), not suffering from hearing problems. Based on participants' responses to the questionnaire (see §2.4), data from 34 men and 12 women (5 based on lesbian or bisexual orientation, 3 based on age, 2 based on hearing problems, 2 based on non-Dutch language background) were discarded. Thus 120 female, self-identified heterosexual listeners remained for further analysis (average age 20.3, s.d. 2.3, range 18–29). Responses by teachers of the course groups were also collected but not used for further analysis.

2.2. Materials

Stimulus sentences were taken from Dutch spontaneous monologues by 24 male speakers (average age 18.0, s.d. 0.7, range 16–19), who spoke about an informal topic of their own choice [14, 15]. Two sentences were selected from each speaker's interview. Selected sentences were between 2.5 and 3.5 s in duration, were spoken fluently and without a long pause, with neutral content, comprehensible without context, and not elliptic (i. e. contained both a subject and an inflected verb). Filler sen-

10.21437/SpeechProsody.2016-223

tences were taken from 24 female speakers (each contributing one sentence) using the same criteria.

For each of the 48 selected stimulus sentences, average syllable duration (excluding pauses) and average f_0 (over voiced portions) were measured using `Praat` [16]. These measurements were then analyzed by means of linear mixed models [17, 18, 19, 20] with only the intercept as fixed predictor, and with speakers as a random effect. The estimated average syllable duration was 0.188 s ($s_u = 0.015, s_e = 0.026$, ICC $= .25$, i. e. with most variance between sentences within speakers), and the estimated average f_0 was 116 Hz ($s_u = 16, s_e = 7$, ICC $= .82$, i. e. with most variance between speakers).

Each speaker (voice) was matched to a unique portrait photo. Photos were taken from 3 public databases of facial portraits [21, 22, 23] and do not portray the actual speakers. The selected photos of 24 males and 24 females each showed one person in the target age range (18–25) with a neutral facial expression. All selected photos were cropped and/or resized to the same size.

2.3. Speech manipulations

One of the two sentences of each male speaker was retained as a baseline stimulus with unchanged tempo and unchanged pitch. The other sentence of each male speaker was varied in tempo (factors $0.85, 1.00, 1.15$) and in overall pitch ($-1.5, 0, +1.5$ semitone), yielding 8 manipulated versions of each sentence. The changes are well above the respective JND [24, 25] and correspond to approx. $\pm 1 s_e$ for both manipulations, while the resulting sentences still sound very natural. Filler sentences by female speakers were not varied. Tempo and pitch were manipulated by means of `sox` [26]. Finally, stimulus and filler sentences were all scaled to -0.5 dB relative to the maximum amplitude.

2.4. Procedure

The 8 manipulated versions of each sentence were distributed over 8 experimental lists, counterbalanced over the 24 male speakers. The 24 unchanged male-spoken sentences and 24 female-spoken filler sentences were added to each experimental lists. Hence the unchanged sentences of all speakers were presented to all listeners, whereas the changed sentences were partitioned over lists so that each listener heard only a single changed version of a particular sentence. This design allowed subsequent within-speaker and within-listener comparisons of baseline and changed versions. The 72 sentences were presented in random order (the same for each list).

The experiment was conducted in a classroom setting, with each experimental list presented to a separate undergraduate course group. Portraits and speech stimuli were presented simultaneously (using PowerPoint) over the classroom projector and sound system. Participants were instructed to rate the attractiveness of the speaker on a 7-point scale (1=extremely unattractive, 7=extremely attractive) on a printed response sheet (see App.A).

After the rating experiment, participants were invited to answer a brief questionnaire about their sex, age, nationality, native language(s), hearing problems, speech problems, dexterity, and sexual orientation as heterosexual or homosexual or bisexual or unknown (including unwilling to answer); see §2.1.

3. Results

The average ratings by the targeted listeners observed in the listening experiment are summarized in Table 1. The lower standard error in the baseline condition is due to the larger number of responses in this condition, as all listeners judged the unchanged sentences of all speakers (see §2.4). Five listeners who had a response range of only 2 (i. e., who had responded with only 2 adjacent scale values to all 72 stimuli) were excluded from further analyses, with 115 listeners remaining.

Table 1: *Mean responses (by targeted listeners only) of subjective attractiveness on an 7-point scale, broken down by manipulations of tempo and pitch, with standard errors in parentheses. Within each cell, the first row summarizes the raw responses and the second row summarizes the log-transformed responses.*

		pitch		
		lower	unchanged	higher
tempo	slower	2.55 (.07)	2.49 (.07)	2.20 (.06)
		0.806 (.028)	0.788 (.028)	0.650 (.028)
	unch'd	2.66 (.07)	2.64 (.02)	2.27 (.07)
		0.847 (.028)	0.844 (.010)	0.675 (.029)
	faster	2.71 (.07)	2.64 (.07)	2.38 (.07)
		0.871 (.028)	0.845 (.028)	0.728 (.029)

The log-transformed responses of the remaining listeners to the *changed* sentences were fed into a linear mixed-effects model (LMM) [17, 18, 19], with listeners ($n = 115$) and speakers ($n = 24$) as two crossed random effects, using maximum likelihood estimation. Fixed predictors in the LMM were (i) the log-transformed baseline response for the same listener and same speaker for the *unchanged* condition (centered), plus the slopes of (ii) tempo (coded as -1=slower, $+1$=faster) and (iii) pitch (-1=lower, $+1$=higher), plus all two-way interactions. The main effects of baseline and pitch were also included as random slopes at the levels of listeners and of speakers. Tentative models with higher-order polynomials, with the three-way interaction term, and with random slopes of tempo, were also explored, but none of these performed better than the optimal LMM reported below. An alternative model with course group as an additional random effect was also explored (with participants nested in these course groups, and with random slopes of baseline, tempo and pitch over groups), but this alternative model too did not perform better than the optimal model ignoring these groups (with restricted maximum likelihood estimation of both models, and using a Likelihood Ratio test, $\chi^2(6) = 1.75, p = .941$). The fixed regression coefficients, variances and correlations estimated by the LMM described above are listed in Table 2.

The LMM shows a significant positive effect of tempo (faster speech tempo is more attractive) as well as a significant negative effect of pitch (higher voice pitch is less attractive). The effects of the prosodic manipulations, although small (cf. Table 1), are clearly significant if the baseline response for the unchanged sentence (for the same listener and same speaker), as well as the random effects of listeners and speakers, are taken into account.

Interestingly, the LMM also shows a significant interaction in the fixed part. The negative interaction of baseline and pitch suggests that the negative effect of pitch becomes larger (more negative) as the speaker is rated as more attractive, as illustrated in Figure 1. In other words, pitch manipulations yield larger effects for a more attractive speaker (the slope of pitch is more

Table 2: *Estimated coefficients of the LMM. Random intercepts and random slopes are reported in variance units, with standardized correlations among random effects. Coefficients and correlations are marked with an $*$ asterisk if $p < .05$, based on percentiles of bootstrapped estimates, over 400 bootstrap replications of the LMM.*

Fixed effects:	estim	std.err	t
intercept	0.720	0.033	21.76$*$
(i) baseline	$+0.371$	0.027	$+13.70*$
(ii) tempo	$+0.036$	0.008	$+\ 4.54*$
(iii) pitch	-0.059	0.018	$-\ 3.24*$
baseline $\times$ tempo	-0.027	0.015	$-\ 1.79$
baseline $\times$ pitch	-0.072	0.017	$-\ 4.21*$
tempo $\times$ pitch	-0.009	0.009	$-\ 1.00$
Random effects:	var	correlations	
listeners ($n = 115$)	0.0393		
baseline|listeners	0.0116	-0.10	
pitch|listeners	0.0039	$+0.54*$	$+0.34$
speakers ($n = 24$)	0.0165		
baseline|speakers	0.0073	$+0.36$	
pitch|speakers	0.0057	$+0.16$	$+0.41$
residual ($n = 2744$)	0.1172		

negative on the righthand side in Fig.1), and smaller effects for a less attractive speaker (the slope of pitch is approximately zero on the lefthand side).

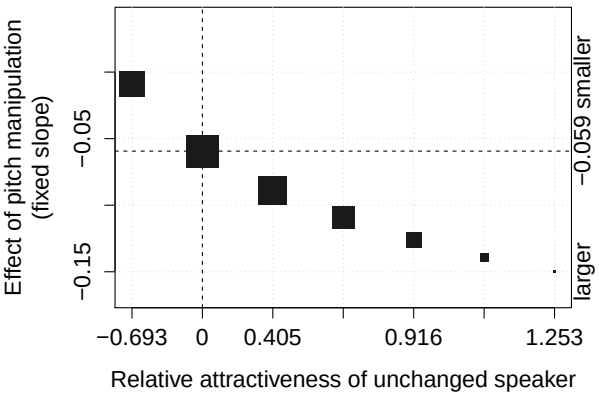

Figure 1: *Interaction of relative baseline ratings (for unchanged stimuli, in log units, centered) and effect of pitch (fixed slope). Symbol size corresponds to the number of responses.*

Between listeners, however, the pattern is somewhat different, as indicated by the positive correlation between a listener's random intercept and her random slope of pitch (see Figure 2, $r = +0.54, p < .0025$). Here, the pattern suggests that pitch manipulations tend to have a larger (negative) effect for individual listeners who give generally lower ratings of attractiveness to the stimuli. Conversely, a listener who is generally more positive, also tends to have a less negative (i. e. shallower) slope of tempo. This correlation between random intercepts and random slopes may suggest a ceiling effect, in that listeners who give higher ratings tend to do so anyway, and are less sensitive to manipulations of voice pitch.

Responses in the unchanged and changed conditions were correlated in the fixed part of the LMM ($\beta = +0.371, p <$

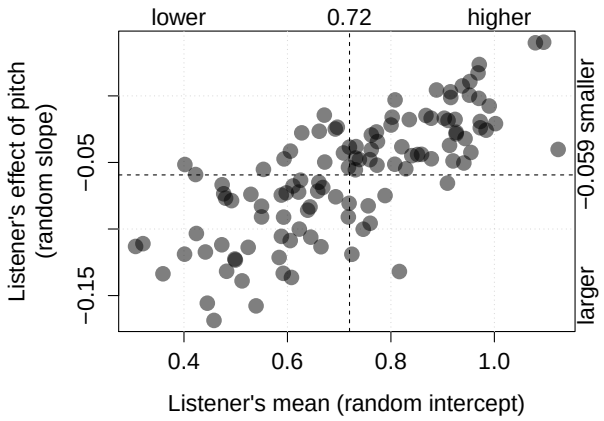

Figure 2: *Individual differences between $n = 115$ listeners in their random intercept and random slope of pitch on log-transformed attractiveness ratings.*

.0025) and perhaps also in the random part at the speakers' level ($r = +0.36$, bootstrapped $p = .095$). A speaker who is rated as overall more attractive in the changed conditions also tends to have a higher average rating in the unchanged baseline condition. This correspondence between a speaker's ratings in unchanged and changed conditions captures individual differences in the speaker's attractiveness, presumably due to the unchanged individual characteristics of his speech (voice quality, vocal tract properties, segmental properties, articulatory behavior, etc) as well as to the photo presented simultaneously with the speech stimulus (cf. §2.4). The baseline effect and the speaker-dependent correlation therefore suggest that listeners performed their task reliably.

4. Discussion

First, the results confirm previously reported effects of pitch manipulations on attractiveness [7, 8], with lower voices being more attractive. While these previous studies used only short vowel stimuli, these findings are replicated here with sentence-length stimuli. This result further corroborates the evidence for the role of male voice pitch in sexual selection through female choice of mate.

Second, the results confirm our prediction that manipulations of tempo also affect the speaker's attractiveness, with faster speech being more attractive. Faster speakers may be regarded as more attractive because speech tempo may indicate the speaker's motor skills and cognitive capabilities.

The significant effects of pitch as well as tempo on the speakers' attractiveness are remarkable, because these effects may well have been weakened by the accompanying portraits. These photos were included in the experimental procedure in order to make the task more realistic for the participants (we typically assess speakers whom we also see). The portrait was presented with the speech stimulus as if it represented the speaker, and the same photo was presented for the changed and unchanged stimuli by the same speaker, in order to stabilise ratings. Presumably, participants' ratings were affected to some extent by the visual properties of the pretended speaker (e.g. a participant may like or dislike particular visual characteristics of the portrayed person). This stabilisation of ratings for the unchanged and changed versions of a speaker may also have

limited the effects of prosodic manipulations on those ratings. Arguably, then, participants' ratings would have been affected more strongly by the prosodic properties of the speech stimuli, if there would have been no visual cues about the speaker's attractiveness.

A third finding is that the effects of pitch, but not of tempo, are modulated by the baseline attractiveness of the speaker and by the individual listener's average rating. Within listeners, pitch manipulations yield larger effects for the more attractive speakers than for the less attractive speakers (Fig. 1). Thus pitch is more important for assessing a speaker whom the listener also considers as more attractive if his prosody is unchanged. Between listeners, however, pitch manipulations yield larger effects for lower-rating listeners than for higher-rating listeners (Fig. 2). This ceiling effect entails that pitch is more important for listeners who generally consider speakers as less attractive (as compared to higher-rating listeners). These modulating effects were only observed for pitch manipulations, however, and not for tempo manipulations. This could be due to the fact that pitch varies more between speakers (and less within speakers) than tempo does (cf. §2.2 for these variations in our stimuli), so that pitch may constitute a more reliable indicator of the speaker's individual characteristics than tempo. In any case, the different modulation patterns may correspond to different functions and causal mechanisms of voice pitch vs. speech tempo in sexual selection.

5. Conclusions

Female listeners rate a male speaker as more attractive if his voice pitch is lowered and his speech tempo is increased, and as less attractive if his pitch is increased and his tempo is decreased, relative to a baseline sentence with unchanged pitch and tempo. These effects suggest that both pitch and tempo play a role in speech-based sexual selection of males by females, although the underlying mechanisms may well be different. Voice pitch indicates the speaker's health and physical dominance [4, 3, 7, 8], while speech tempo may indicate his motor skills and cognitive competence.

6. Acknowledgements

Part of this study was conducted as a joint BA thesis project in Linguistics by the second and third author. Our thanks are due to the teachers for their cooperation in collecting data, and to Nivja de Jong for helpful comments.

7. References

[1] Holmberg, E.B., Hillman, R.E. and Perkell, J.S., "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", *J. Acoust. Soc. Am.* 84, 511–529, 1988.

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

[3] Puts, D.A., Apicella, C.L. and Cárdenas, R.A., "Masculine voices signal men's threat potential in forager and industrial societies," *Proc. Royal Soc. B: Biological Sciences*, 279, 601–609, 2012.

[4] Dabbs, J.M., and Mallinger, A. "High testosterone levels predict low voice pitch among men," *Personality and Individual Differences*, 27, 801–804.

[5] Apicella, C.L, Feinberg, D.R. and Marlowe, F.W., "Voice pitch predicts reproductive success in male hunter-gatherers," *Biology Letters*, 3, 682–684, 2007.

[6] Andersson, M.B., *Sexual selection*. Princeton: Princeton Univ Press, 1994.

[7] Collins, S.A., "Men's voices and women's choices," *Animal Behaviour*, 60, 773–780, 2000.

[8] Feinberg, D.R., Jones, B.C., Little, A.C., Burt, D.M. and Perrett, D.I., "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices," *Animal Behaviour*, 69, 561–568, 2005.

[9] Quené, H., "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," *J. Acoust. Soc. Am.*, 123, 1104–1113, 2008.

[10] Jacewicz, E., Fox, R.A. and Wei, L., "Between-speaker and within-speaker variation in speech tempo of American English," *J. Acoust. Soc. Am.*, 128, 839–850, 2010.

[11] Moon, S.-J., and Lindblom, B. "Two experiments on oxygen consumption during speech production: vocal effort and speaking tempo," in *XVth Int. Congress of Phonetic Sciences, Barcelona, Spain, 2003, Proceedings*, pp.3129–3132.

[12] Apple, W., Streeter, L.A. and Krauss, R.M., "Effects of pitch and speech rate on personal attributions," *J. Personal. and Soc. Psy.*, 37, 715–727, 1979.

[13] Smith, B.L., Brown, B.L., Strong, W.J. and Rencher, A.C., "Effects of speech rate on personality perception," *Language & Speech*, 18, 145–152, 1975.

[14] Orr, R., Quené, H., van Beek, R., Diefenbach, T., van Leeuwen, D.A. and Huijbregts, M., "An International English speech corpus for longitudinal study of accent development", in *InterSpeech 2011, 27–31 Aug, Florence, Italy, Proceedings*, pp.1889–1892.

[15] Quené, H. and Orr. R., "Long-term convergence of speech rhythm in L1 and L2 English", in *Speech Prosody 2014, 20–23 May, Dublin, Ireland, Proceedings*, pp.342–345.

[16] Boersma, P. and Weenink, D., Praat: Doing phonetics by computer, version 6.0, 2015. Online: http://www.praat.org

[17] Quené, H. and van den Bergh, H., "On Multi-Level Modeling of data from repeated measures designs: A tutorial", *Speech Comm.*, 43(1–2):103–121, 2004.

[18] Quené, H. and Van den Bergh, H., "Examples of mixed-effects modeling with crossed random effects and with binomial data", *J. Memory and Language*, 59(4):413–425, 2008.

[19] Bates, D., Maechler, M., Bolker, B. and Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9, 2015. Online: http://CRAN.R-project.org/package=lme4

[20] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, version 3.2.2, 2015. Online: http://www.R-project.org/.

[21] Hancock, P., Utrecht ECVP: Psychological Image Collection at Stirling, 2008. Online: http://pics.stir.ac.uk/.

[22] Nefian, A.V., Georgia Tech face database, 1999. Online: http://www.anefian.com/research/face_reco.htm

[23] Spacek, L., Collection of facial images (faces94, faces95), 2008. Online: http://cswww.essex.ac.uk/mv/allfaces/.

[24] Quené, H. "On the just noticeable difference for tempo in speech", *J. Phonetics*, 35, 353–362, 2006.

[25] 't Hart, J., Collier, R., and Cohen, A. *A Perceptual Study of Intonation: An experimental- phonetic approach to speech perception*. Cambridge: Cambridge Univ Press, 1990.

[26] Bagwell, C. Sound eXchange (SOX), version 14-4-1, 2013. Online: http://sourceforge.net/projects/sox/.

A. Excerpt of instructions

... In a moment you will see 72 photos of people. With every face you will also hear a sound fragment. We'd like to ask you to indicate for every person how attractive you find that person. You have about 3 seconds to respond for each person.