

Voice attractiveness: Influence of stimulus duration and type

C. Ferdenzi · S. Patel · I. Mehu-Blantar · M. Khidasheli ·
D. Sander · S. Delplanque

Published online: 13 December 2012
© Psychonomic Society, Inc. 2012

Abstract Voice attractiveness is a relatively new area of research. Some aspects of the methodology used in this domain deserve particular attention. Especially, the duration of voice samples is often neglected as a factor and happens to be manipulated without the perceptual consequences of these manipulations being known. Moreover, the type of voice stimulus varies from a single vowel to complex sentences. The aim of this experiment was to investigate the extent to which stimulus duration (nonmanipulated vs. normalized) and type (vowel vs. word) influence perceived voice attractiveness. Twenty-seven male and female raters made attractiveness judgments of 30 male and female voice samples. Voice samples included a single vowel /a/, a three-vowel series /i a o/, and the French word “bonjour” (i.e., “hello”). These samples were presented in three conditions: nonmanipulated, shortened, and lengthened duration. Duration manipulation was performed using the pitch synchronous overlap and add (PSOLA) algorithm implemented in Praat. Results for the effect of stimulus type showed that word length samples were more attractive to the opposite sex than vowels. Results for the effect of duration showed that the nonmanipulated sound sample duration was not predictive of perceived

attractiveness. Duration manipulation, on the other hand, altered perceived attractiveness for the lengthening condition. In particular, there was a linear decrease in attractiveness as a function of modification percentage (especially for the word, as compared with the vowels). Recommendations for voice sample normalization with the PSOLA algorithm are thus to prefer shortening over lengthening and, if not possible, to limit the extent of duration manipulation—for example, by normalizing to the mean sample duration.

Keywords Voice perception · Attractiveness · PSOLA algorithm · Duration normalization

Introduction

There is growing evidence that the human voice conveys important, socially relevant information about the speaker, independently of any linguistic and emotional content (Latinus & Belin, 2011). Especially, in the last 10 years, a number of studies have focused on the link between the voice acoustic features and its perceived attractiveness (e.g., Bruckert, Lienard, Lacroix, Kreutzer, & Leboucher, 2006; Collins, 2000; Feinberg, Jones, Little, Burt, & Perrett, 2005; Hodges-Simeon, Gaulin, & Puts, 2011). Some of these studies revealed, for example, that lower-pitched male voices and higher-pitched female voices are generally more attractive to opposite-sex listeners (Collins, 2000; Collins & Missing, 2003; Vukovic et al., 2011) when pitch values are not extreme (Borkowska & Pawlowski, 2011). Formant characteristics (frequency, dispersion) also play a role in voice attractiveness (Collins & Missing, 2003; Feinberg et al., 2011, 2006; Puts, Barndt, Welling, Dawood, & Burriss, 2011). Around the fertile phase of their menstrual cycle, women display stronger preference for masculine, low-pitch voices (Feinberg et al., 2006; Puts, 2005) and have themselves a higher-pitched voice (at least for some types of speech; Bryant & Haselton, 2009) and a more perceptually attractive voice (Pipitone &

Electronic supplementary material The online version of this article (doi:10.3758/s13428-012-0275-0) contains supplementary material, which is available to authorized users.

C. Ferdenzi (✉) · S. Patel · I. Mehu-Blantar · M. Khidasheli ·
D. Sander · S. Delplanque
Swiss Center for Affective Sciences, University of Geneva,
7 rue des Batoirs,
1205 Geneva, Switzerland
e-mail: camille.ferdenzi@unige.ch

C. Ferdenzi
e-mail: cferdenzi@hotmail.com

C. Ferdenzi · I. Mehu-Blantar · D. Sander · S. Delplanque
Laboratory for the Study of Emotion Elicitation and Expression
(E3 Lab), Department of Psychology, FPSE, University of Geneva,
Geneva, Switzerland

Gallup, 2008). This suggests that voice perception has a role in guiding human mate choice. This idea is supported further by the fact that in some studies, voices are rated more attractive in individuals bearing signals of mate quality, such as body symmetry (Hughes, Harrison, & Gallup, 2002; Hughes, Pastizzo, & Gallup, 2008), reproductive/mating success (Hughes, Dispenza, & Gallup, 2004), and face attractiveness (Lander, 2008; Saxton, Caryl, & Roberts, 2006).

This is a relatively new area of research. As such, the methods used are variable, and some methodological questions remain unanswered. The main question we chose to address in this study concerns the duration of the sound excerpts. Many studies do not mention the range of their sample durations, and the studies that mention them reveal substantial variations between experiments. For example, the vowel duration was 250–380 ms (mean = 290 ms) in Collins and Missing (2003), 640 ms on average in Feinberg, Jones, Little, et al. (2005), and 201–477 ms in Bruckert et al. (2010). Studies using sentences also sometimes have mentioned the average duration, but only for information purposes (e.g., Lander, 2008; Puts, Apicella, & Cárdenas, 2012). The only study investigating the effect of voice sample duration on listeners' ratings found no relationship between the duration of 1-to-10 counting sequences and attractiveness of male and female voices (Hughes et al., 2008). The authors found only a significant negative relationship between duration and estimated intelligence in male voices, but this could as well have been due to speech rate (Feldstein, Dohm, & Crown, 2001) rather than to sound duration (only the total duration of the sequence was taken into account). Since previous research in speech has shown that stimulus length influences speech perception (e.g., Diehl, Lotto, & Holt, 2004) and emotion recognition (e.g., Pell & Kotz, 2011), it remains to be seen whether sample duration affects perceived attractiveness.

Furthermore, some studies have chosen to standardize sound duration, using an algorithm developed initially by Moulines and Charpentier (1990): the pitch synchronous overlap and add (PSOLA) algorithm. To obtain an expanded voice sample, for example, the algorithm first analyzes and segments the sound signal. Then it synthesizes a new time-stretched version by overlapping and adding time segments extracted from the input sound. Using the Praat implementation of this algorithm (Boersma & Weenink, 2011), authors have normalized the duration of individual vowels (e.g., to 500 ms, Feinberg, DeBruine, Jones, & Perrett, 2008, and Feinberg et al., 2006; to 350 ms, Saxton, DeBruine, Jones, Little, & Roberts, 2009). Their aim was to “to control for variation in spoken vowel duration between individuals” (Feinberg et al., 2006, p. 217), and they did not investigate the possible impact of this manipulation on subsequent attractiveness ratings, certainly because this would not affect their results given the design they used (comparison of two versions, masculinized and feminized, of the same length-

modified voices; Feinberg et al., 2006; Saxton et al., 2009). However, it is possible that such a manipulation makes the output voices sound less natural and, consequently, less attractive than the original. It is also likely that such perceptual consequences would be more pronounced for samples with durations that are the most distant from the target duration chosen for normalization. There are some designs where this could be detrimental—for instance, when the attractiveness of an individual's voice is put in relation to other characteristics of that individual or when brain correlates of attractiveness are studied. Consequently, knowing whether (and to what extent) duration contraction and expansion change voice attractiveness would be valuable for future voice perception studies.

Additionally, in voice research, samples vary considerably according to their content. Some authors use short sounds with neutral content, such as numbers (e.g., from 1 to 10; Hughes et al., 2002, and subsequent work), while others use neutral sentences such as the Rainbow passage (see Puts, Gaulin, & Verdolini, 2006, and subsequent work) or the time of day (e.g., “it's fifteen minutes to three,” used by Lander, 2008). Connoted sentences have also been chosen, such as the equivalent of “hello” (Apicella & Feinberg, 2009), “I really like you/I really don't like you” (Jones, Feinberg, DeBruine, Little, & Vukovic, 2008; Vukovic et al., 2008), and even free speech sentences (Fischer et al., 2011; Hodges-Simeon et al., 2011; Puts, 2005). Still others use monophthong vowel sounds (e.g., /a/ and /i/ in English). These stimuli are most common in studies on voice preference (Bruckert et al., 2010; Collins, 2000; Feinberg, Jones, DeBruine, et al., 2005; Ferdenzi, Lemaître, Leongómez, & Roberts, 2011). The use of vowels is beneficial in two regards. First, these samples enable perceptual judgments of pitch and voice quality without being colored by contextual factors (co-articulation, emphasis, and semantic meaning). Second, vowel stimuli are often preferred for acoustics measures, such as for voice quality and formant analysis (Patel, Scherer, Björkner, & Sundberg, 2011). To date, little is known about differences in the perceived attractiveness among several voice sample types (e.g., word vs. vowel) recorded from the same individual. Most important with regard to our main question, it is not known whether some voice sample types are more sensitive to duration and duration manipulation effects on attractiveness (providing there are any).

In this experiment, we tested two different questions: (1) Do speech type and speech segment length influence attractiveness judgments of voices, and (2) does duration manipulation of a voice stimulus affect its perceived attractiveness. To answer these questions, we used different types of stimuli commonly used in attractiveness studies—namely, a single vowel, a three-vowel sequence, and a word, with varied sample durations. We manipulated duration to compare attractiveness ratings of different types of original stimuli of varied durations versus the same stimuli with normalized short and long durations.

Method

Participants

Twenty-seven Caucasian participants (15 men, 12 women) 22.1 ± 4.5 years of age (range, 17–34) were recruited from students and members of the staff of the University of Geneva, Switzerland. These participants served as raters in the main experiment. To avoid unwanted variability in attractiveness ratings due to language (Bresnahan, Ohashi, Nebashi, Liu, & Morinaga Shearman, 2002) and, possibly, sexual orientation, participants were required to be native speakers of French and to report being heterosexual. Participation was voluntary, and participants gave their informed written consent before starting the experiment.

Voice stimuli

Participants

Voice recordings of 30 Caucasian participants were used. These participants were distinct from the raters. Half of the speakers were males, and half were females (age: 22.9 ± 3.8 years; range 18–34 years). These individuals were also recruited mostly from students and members of the staff of the University of Geneva. All were French native speakers, heterosexual, and nonsmokers and declared not having a cold or illness on the day of voice recording or any speech impediment that would affect the way they naturally spoke. The voice recordings were a part of a larger experiment in which participants were also videotaped to create a database of voices and faces, the Geneva Attractiveness Database (GEAD; currently under development and soon to be released for academic research). Therefore, participants received compensation, either financially or by credits for the psychology course at the University, for their participation in the study.

Stimulus recordings

The voices were recorded with a BCM 104 condenser studio microphone with cardioid directional characteristics (Neumann, Berlin; www.neumann.com) in a quiet room at a constant distance from the microphone. The recording sessions were led by one of three female experimenters,¹ but the participants were alone in the experimental room during the recording and were in contact with the experimenter through a speaker-microphone device. The voices were recorded onto a computer hard disk using Cubase v.5.5.0 (www.steinberg.fr)

¹ The three experimenters did the recording for 1, 9, and 20 of the 30 voices used in this experiment. Mean attractiveness of the voices (nonmanipulated duration) did not differ between the two latter experimenters [one-way ANOVA with experimenter as between-subjects factor: $F(2,27) = 1.16, p = .330$].

at a sampling rate of 44.1 kHz with 24-bit quantization and were saved as uncompressed wav files. The amplitude of the recorded signal was adjusted for each participant with a mixing table. Participants were required to pronounce the sentence “Bonjour. Il est deux heures moins dix” (Hi. It’s ten to two) and a series of six monophthong vowels /ɛ/, /i/, /a/, /o/, /u/ and /y/ (International Phonetic Alphabet). The sentence and the series of vowels were each pronounced twice. The audio samples used for the ratings were extracted from the second repetition (when participants are expected to be more relaxed). We only used “bonjour” and three of the vowels in the middle of the series—/i/, /a/, and /o/—as a three-vowel series (using vowels in the middle of the sequence limits intonation variations; see Collins, 2000). The vowel /a/ was also used independently as a single vowel. The audio samples were isolated using Praat v.5.2 (Boersma & Weenink, 2011), and normalization of sound intensity was performed by matching the average absolute amplitude of all recordings using MATLAB v.7.12.

Stimulus selection

While a total of 30 samples were used (each from a different individual), these were selected from a database of samples from 86 individuals. These samples were subdivided into three groups (low-, middle-, and high-pitched) on the basis of f_0 . To do this, the f_0 was computed in Praat within the range of 100–600 Hz for female voices and 65–300 Hz for male voices (as performed in Feinberg et al., 2006). It was decided to keep five samples in each pitch category (lowest, middle, high) for each sex, to favor a wide range of frequencies and, therefore, potentially a wide range of attractiveness levels (Collins, 2000; Vukovic et al., 2011). A number of steps were performed to obtain this final sample set and reduce its variability, due to issues in the recording or duration manipulation processes. Note that in the following steps, when a sample was discarded, all the other samples of the same participant were removed too. First, the extreme duration values were identified by visual inspection of the distribution of duration values (computed in MATLAB; see Fig. 1). The samples with the highest and lowest durations were removed (i.e., samples of 22 participants). The new minimum and maximum durations of the remaining samples were used as the limits for duration manipulation: 200 and 400 ms for each of the individual vowels and 420 and 820 ms for the word samples. Second, we removed recordings that were noisier than the others (samples of 6 participants) and samples that sounded too unnatural after duration transformation (see *Stimulus manipulation*) (samples of 11 participants). Also, we eliminated samples of 17 participants for whom the word “bonjour” had a more abrupt ending than for others (undesired effect due to the fact that this word was initially part of a whole sentence).

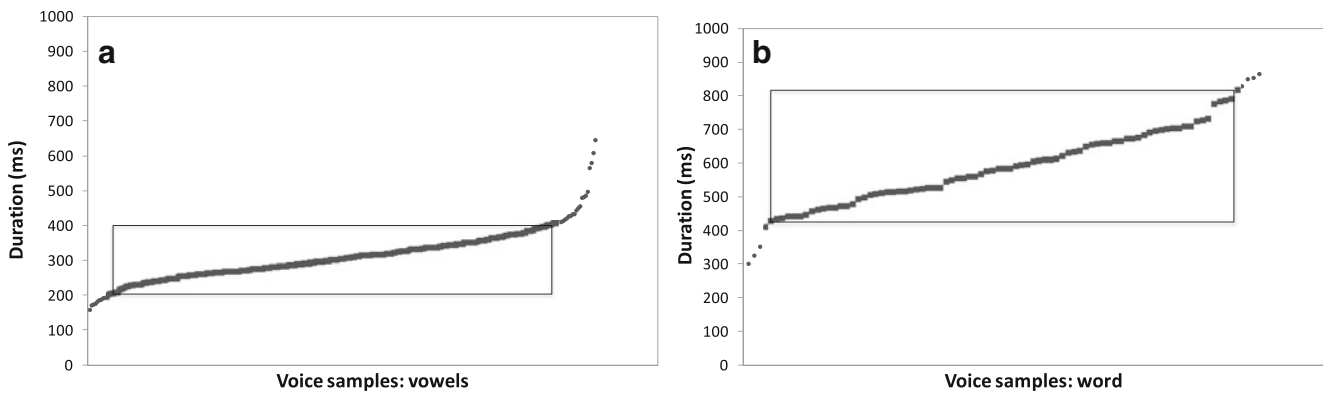


Fig. 1 Distribution of the durations (in milliseconds) of the nonmanipulated audio samples from 86 individuals, from which the 30 samples used in this study were selected: **a** the vowels /i/, /a/, /o/ and **b** the word “bonjour.” Samples within the rectangle have been included in the stimuli selection

Stimulus manipulation

Each audio sample (/i/, /a/, /o/ within the three-vowel sequence and “bonjour”) was used in three versions: non-manipulated duration (ranging from 209 to 400 ms for the vowels and from 436 to 791 ms for the word), duration shortened (200 ms for the vowels and 420 ms for the word), and duration lengthened (400 ms for the vowels and 820 ms for the word). Reduction of duration included a manipulation ranging from -4% to -50% of the original length for the vowels and from -4% to -47% for the word. Extension of duration represented a manipulation ranging from $+0\%$ to $+92\%$ for the vowels and from $+4\%$ to $+88\%$ for the word. Duration lengthening and shortening were performed using the PSOLA algorithm in Praat (Boersma & Weenink, 2011). From these nonmanipulated and manipulated samples, a total of 270 samples were presented to the raters: samples from 30 participants \times 3 types of stimuli (single vowel /a/, three-vowel sequence /i a o/, word “bonjour”) \times 3 duration (nonmanipulated, short, long) conditions. Examples of stimuli are provided in the supplementary materials, and all stimuli are available from the authors upon request for nonprofit research.

Procedure

The 270 stimuli were presented in random order with an E-Prime interface, v.2.0 (Psychology Software Tools), through headphones at constant amplitude. The experiment was divided into 10 blocks of 27 voices, separated by a break of at least 15 s (at the end of this fixed duration, participants could decide to continue as soon as they were ready). The total duration of the rating session was 30–40 min. Each stimulus was preceded by a 1-s black screen and 1-s “listen” instructions announcing the stimulus. For the three-vowel sequence, each individual vowel was presented every 600 ms, including the length of the sample. The rating

screen appeared 500 ms after the end of the sound sample (see Fig. 2).

Participants were instructed to rate attractiveness on a scale of 1–7 as accurately and as quickly as possible. The response options were evenly placed in a semicircle in the center of the screen. Both attractiveness ratings and response times (time between apparition of the rating scale and the click of the mouse on the chosen response option) were recorded. The maximum allowed response time was 7 s. A null rating was given if no response was provided within this time frame, and the test automatically stopped if more than 10 % of the trials were missed. All participants complied with the instructions, and none had to restart the experiment. The maximum number of answers that were missed per participant was 2 % of 270 trials. To minimize possible biases in the response time measure, the mouse cursor was automatically repositioned at a central point equidistant from the answer options (squares 1–7) after each trial (see Fig. 2 for a summary of the procedure and a view of the user interface). Before performing the analyses, outliers in the response time were removed. The outliers were defined as values greater or less than three standard deviations of the participant’s mean (2 % of the trials). Since the distribution was skewed, the remaining response time values were then log-transformed (cf. Whelan, 2008).

Results

The analyses described below were performed on attractiveness and response time scores averaged by stimulus (i.e., voice; sample size, $N = 30$). All post hoc analyses were Tukey HSD tests ($\alpha = .05$), unless otherwise noted.

Effect of stimulus type and duration

Two 2-way Greenhouse–Geisser corrected repeated measures ANOVAs were run to investigate the effect of stimulus type (single vowel, three-vowel sequence, word) and duration

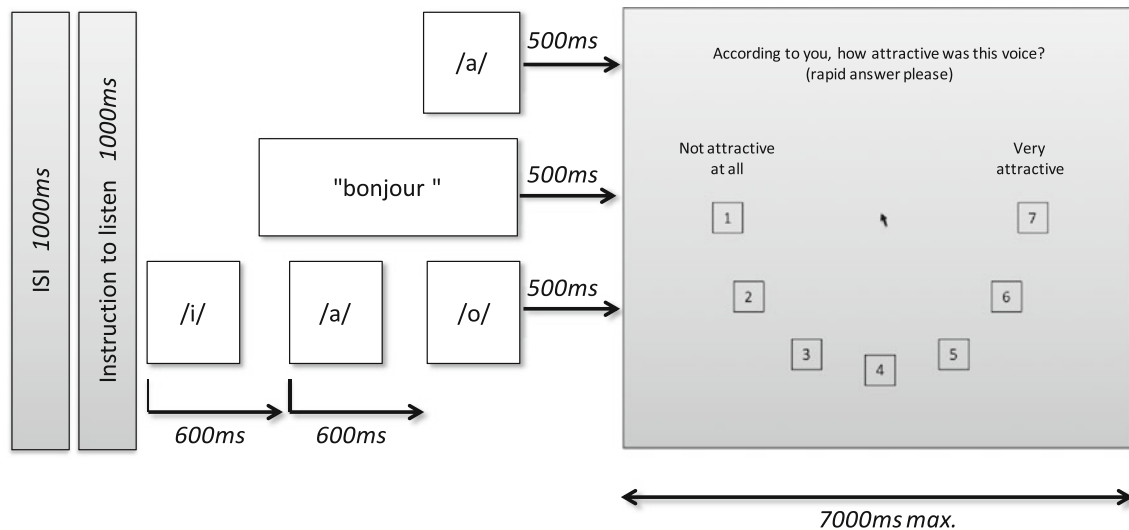


Fig. 2 Procedure used for attractiveness ratings and reaction time recording of the three types of stimulus (single vowel, three-vowel sequence, word “bonjour”). ISI = interstimulus interval

(nonmanipulated, short, long) on attractiveness ratings and response time, respectively. Results on attractiveness ratings revealed no main effect of stimulus type, $F(2, 58) = 2.23, p = .119$, but a significant effect of duration, $F(2, 58) = 60.23, p < .001$, and a significant interaction between stimulus type and duration, $F(4, 116) = 19.26, p < .001$. Post hoc analyses showed that the lengthened stimuli were significantly less attractive than the shortened and nonmanipulated ones on average, which, in turn, did not differ. Furthermore, the word “bonjour” was more affected by this effect than were the two other types of stimuli, the three-vowel sequence being the least affected (see Fig. 3a). Similar ANOVAs were performed with average attractiveness calculated for opposite-sex and same-sex raters separately and brought similar results, with one piece of additional information. Opposite-sex voice attractiveness varied as a function of stimulus type, $F(2, 58) = 3.62, p < .05$; the word “bonjour” was significantly more attractive than the single vowel /a/. The three-vowel sequence was in between and not significantly different from the other two stimulus types. Results for response time revealed a main effect of stimulus type, $F(2, 58) = 33.93, p < .001$ (see Fig. 3b). The three types of stimuli significantly differed from each other; the single vowel resulted in the longest response time, and the three-vowel sequence resulted in the shortest time (post hoc). There was a significant effect of duration on response time, $F(2, 58) = 6.44, p < .01$. Post hoc testing showed that responses were faster for long samples than for short samples, unchanged samples being not significantly different from either long or short samples. There was no significant interaction between stimulus type and duration, $F(4, 116) = 1.69, p = .166$.

Effect of percentage of duration manipulation

Linear regressions were performed to investigate whether the amount of duration manipulation (for the single vowel /a/ and

“bonjour”) could predict modifications in attractiveness ratings and response times. The two variables used were the percentage of modification, as compared with the nonmanipulated duration, and the difference in attractiveness or response time between the nonmanipulated and the normalized sounds (normalized minus nonmanipulated). Attractiveness was significantly decreased as a function of percentage of lengthening for both the single vowel, $r = -.48, F(1, 28) = 8.40, p < .01$, and the word “bonjour,” $r = -.81, F(1, 28) = 51.76, p < .001$. Shortening was not a significant predictor of attractiveness modification (single vowel, $r = .03$; “bonjour,” $r = .16, p > .392$; see Fig. 4). However, it can be speculated that duration transformation is more drastic for lengthening than for shortening, since no sample is shortened more than 50 % of its original size, whereas some of them are lengthened more than 50 % their size. Therefore, the regressions were performed again after removing the samples that underwent more than 50 % duration modification ($N = 6$ for the single vowel /a/ and $N = 9$ for the word “bonjour”). The outcomes remained unchanged, with significant relationships only between percentage of lengthening and attractiveness alteration [for /a/, $r = -.59, F(1, 22) = 11.86, p < .01$; for “bonjour,” $r = -.80, F(1, 19) = 34.97, p < .001$]. Regarding response time, only lengthening of the single vowel /a/ was marginally associated with a decrease in response time, $r = -.36; F(1, 28) = 4.13, p = .052$ (Fig. 4c).

Effect of nonmanipulated duration

The original nonmanipulated duration did not significantly predict attractiveness (linear regressions between /a/ and “bonjour” original durations and attractiveness: $r_s = -.16, p_s > .406$). It predicted response time, but for the single vowel /a/ only [$r = -.37, F(1, 28) = 4.40, p > .05$; longer durations triggered shorter response times].

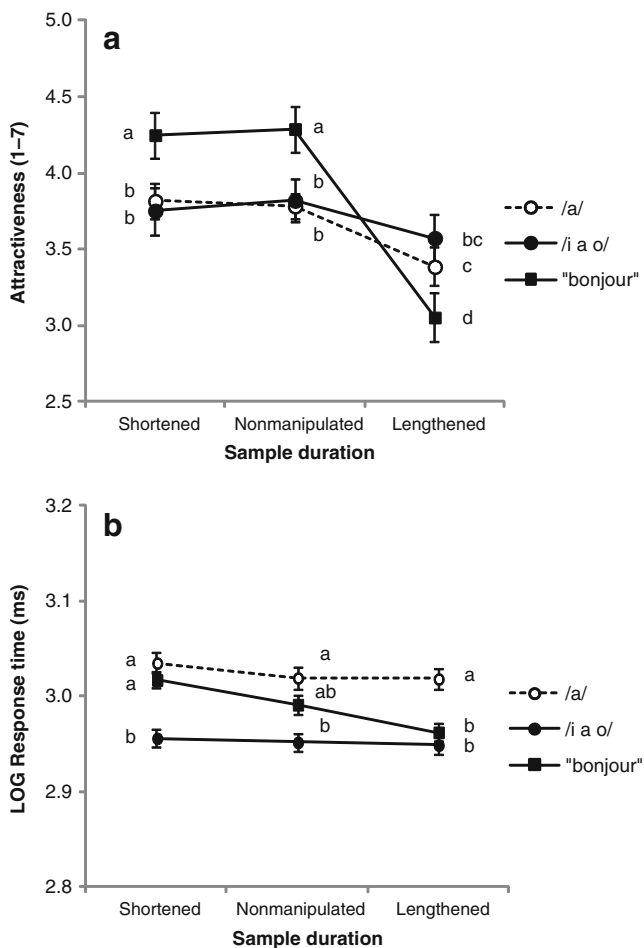


Fig. 3 **a** Attractiveness ratings and **b** response time (log-transformed) to attractiveness ratings of voice samples from 30 participants, rated by 27 participants, as a function of their duration (nonmanipulated or normalized to a short and a long duration) and their type (single vowel, three-vowel sequence, or the word “bonjour”). Letters a, b, c, d indicate significantly different groups according to a Tukey HSD post hoc test

Interrater agreement

Cronbach’s alphas were computed on raw (nonaveraged) data, allowing us to quantify the level of agreement between participants for the attractiveness ratings. Agreement was high for all conditions, since alphas were $>.70$ (Kline, 1993; ranging from .84 to .93). To compare the different conditions (type, duration), we used a bootstrapping procedure performed with MATLAB software. Instead of using a single alpha by condition (e.g., shortened “bonjour”), this method is based on a repeated resampling procedure: It uses a randomly chosen N -size subsample of the initial participant sample to compute Cronbach’s alpha. Resampling was performed 1,000 times, thus providing a distribution of 1,000 alphas per condition. To test the difference between the alphas of two conditions (e.g., shortened vs. nonmanipulated “bonjour”), we computed the differences for all iterations of the two distributions, which provided a distribution of the differences. Alphas of the two

conditions were considered as significantly different when the distribution of the differences, based on the confidence interval (determined with a chance level of $p < .001$), did not include the zero value (Davison & Hinkley, 1997). Two-by-two comparisons showed no significant difference as a function of stimulus type (for a given duration condition) or as a function of duration (for a given stimulus type).

Discussion

The main aim of this experiment was to investigate whether duration of a voice sample would have an effect on its perceived attractiveness and, more specifically, whether duration normalization would affect attractiveness judgments. We showed in this experiment that the nonmanipulated sound sample duration was not predictive of perceived attractiveness. Between-rater agreement for attractiveness rating was high, regardless of the duration condition (nonmanipulated, shortened, lengthened). Shortening the voice samples in Praat (up to almost 50 % of the original duration) did not significantly modify their perceived attractiveness. On the contrary, lengthening had detrimental effects on attractiveness ratings. Furthermore, lengthening was linearly related to a decrease in attractiveness. This was also true when we limited the analyses to samples stretched no more than 50 % of their original duration. The deleterious effect of lengthening could be due to the algorithm procedure itself, which, although a very powerful tool, might alter the signal integrity, making the sample sound more unnatural than the nonmanipulated one.

Some experimental designs might require equalization of the voice samples’ duration, and manipulation with the PSOLA algorithm might be useful in these cases (note that taking only a fixed-length portion of the signal is not recommended for vowels, because of sharp cuts, and is even impossible for words). Reaction time or brain activation studies are good examples of such designs, where stimulus duration may by itself affect the outcome variables. If, in addition, the individual level of attractiveness is important—for example, if brain responses to a given voice are meant to be linked with other characteristics of that voice or of the person producing that voice—then it is imperative to be aware that voice duration manipulation may introduce some unwanted noise into the results by altering attractiveness. Some recommendations to researchers in voice attractiveness can thus be formulated from our results. First, duration normalization of male and female voice samples does not have to be systematic, since there is no relationship between natural duration and perceived attractiveness (at least for the duration ranges used in the present study—namely, 209 – 400 ms for vowels and 436 – 791 ms for the word “bonjour”). Second, for designs where normalization is required,

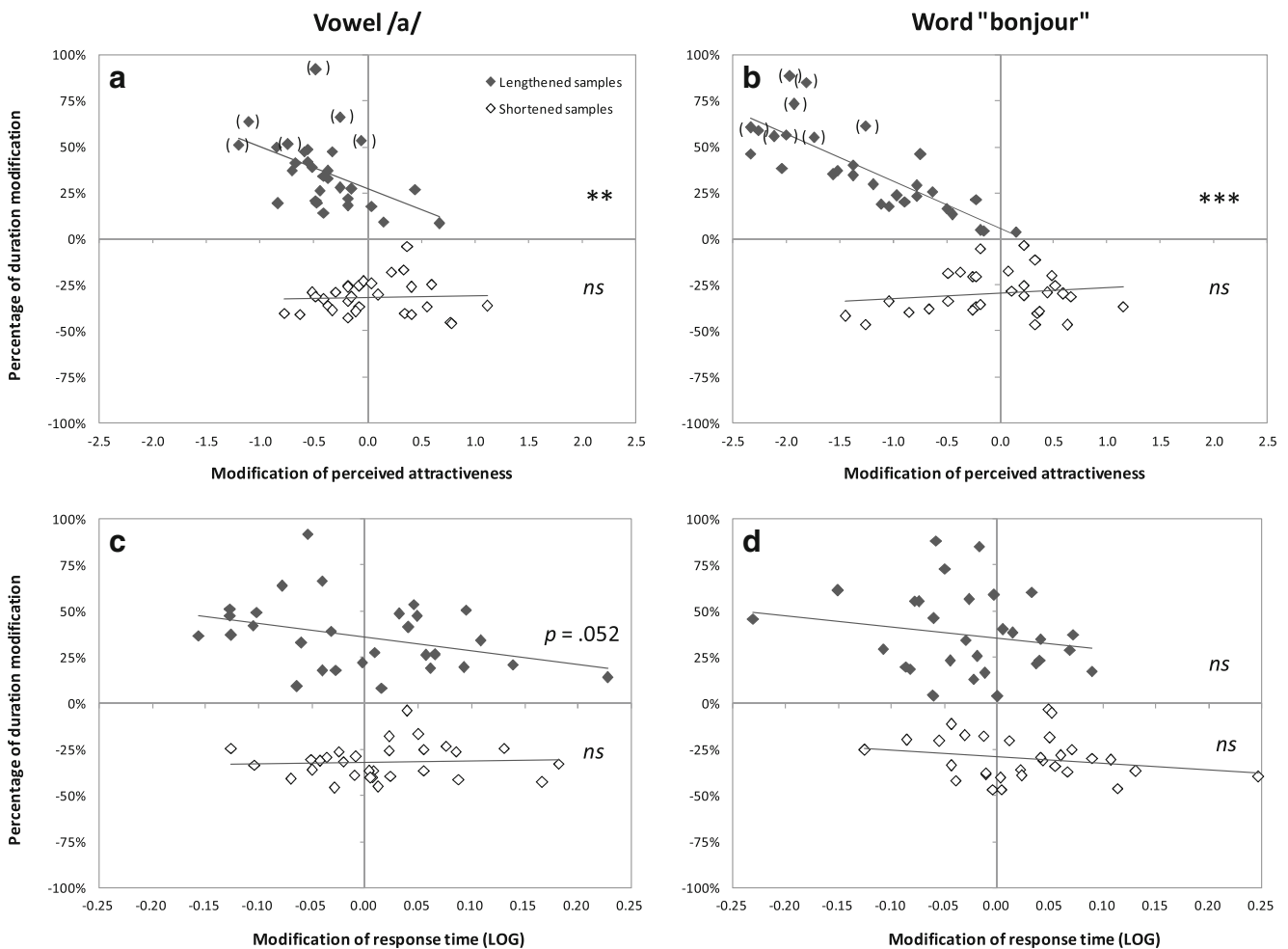


Fig. 4 Linear relationships between the percentage of duration manipulation, as compared with the nonmanipulated duration (positive values, lengthening; negative values, shortening), and modification of perceived attractiveness of **a** the single vowel /a/ and **b** the word “bonjour”(attractiveness of the normalized sample minus the nonmanipulated one) and modification of response time (log-transformed) for **c** the single vowel /a/ and **d** the word “bonjour” (response time for the

normalized sample minus the nonmanipulated one). Result of the linear regressions: *** $p < .001$; ** $p < .01$; ns, not significant ($p > .10$). Regressions were performed again without the cases within brackets, due to the discrepancy of manipulation amplitude between the lengthening and shortening conditions (significance levels of the resulting regressions remained unchanged)

as was mentioned above, our results suggest that duration manipulation with PSOLA should be performed preferably in the direction of shortening, rather than lengthening. If the samples are long enough, normalization to the shortest duration should be applied. If not, we recommend limiting the amount of manipulation (i.e., the percentage of duration change, as compared with the natural duration). One way to operationalize this would be to normalize to the mean duration of the samples. It must be kept in mind that we chose the duration normalization parameters as a function of the naturally occurring durations of the voice samples we collected in our given settings. This might vary from one laboratory to another. For example, in another study, vowel durations were normalized to 500 ms, which would be too long for our samples (Feinberg et al., 2008, 2006). Differences in the natural duration of produced sounds may be due

to the instructions given to the participants—for example, whether they are required or not to sustain the sound and whether/how the experimenter pronounces the sound to the participant (which we believe should be avoided). Finally, adding stimulus type as a variable in our design revealed that the duration lengthening procedure was more deleterious for the word “bonjour” than for the vowels. The above-described recommendations are thus even more relevant for researchers willing to normalize voice samples that are more complex than plain vowels.

As a secondary aim of our experiment, we were also able to determine whether the voice sample type influenced attractiveness ratings. Using three different types of stimuli commonly used in voice attractiveness studies (single vowel /a/, three-vowel sequence /i a o/, and the word “bonjour”), this experiment provided evidence that when pronounced by

the opposite sex, the word is rated as more attractive than one or several vowels. One reason for this may be the inclusion of prosodic and potentially emotional information in word length samples, but not vowel or vowel series samples; however, we cannot conclude this with certainty in the present study. Indeed, it has been shown before that cues of social interest displayed by the speaker—which might be better indicated in a word than in a vowel—positively affect the listener’s evaluation of the voice (Jones et al., 2008). Response time was inversely proportional to sample durations, suggesting that the greater the amount of information (in terms of sound duration), the easier the judgments. It might also be that attractiveness judgments require only a small amount of information and that raters, therefore, have more time to prepare their motor response during the longer samples. Finally, between-rater agreement for attractiveness ratings was high and did not differ according to stimulus type, suggesting a good reliability of participants’ judgments for vowels as well as for words. Consequently, all stimulus types investigated can be confidently used, and the choice of a given stimulus type (when several are available, as in our GEAD) should depend on whether priority is given to acoustics measurements (then vowels are good candidates; e.g., Patel et al., 2011; Shrivastav, Camacho, Patel, & Eddins, 2011) or to prosody and content (then a word such as “bonjour” may be more appropriate).

This experiment provides elements to answer methodological questions that many researchers in voice attractiveness might have raised at some point when designing their experiments. We chose a particular experimental design to answer those questions, but some limitations to our approach should be mentioned. Additional parameters possibly influencing attractiveness should be investigated—for example, speech rate. In the three-vowel sequences, we used vowel presentation every 600 ms, but rate of speech might be influential and should therefore be studied. Several studies report standardizing the rate of sound excerpts presentation (e.g., one numeral per second, Hughes et al., 2004, and Saxton et al., 2006; one vowel per 0.5 s, Feinberg, Jones, Little, et al., 2005), but because it was not the aim of the studies, potential effects of rate on perceptual variables were not described. Additionally, although we measured between-subjects consistency (Cronbach’s alpha), within-subjects reliability of attractiveness ratings was not investigated, due to time constraints of the test. Recent evidence suggests that several repetitions are needed to obtain reliable responses on rating scales (Shrivastav, Sapienza, & Nandur, 2005). Repeated ratings are rarely done in voice attractiveness experiments, and this question would be worth being explored further.

Conclusion

This experiment investigated some methodological questions related to the manipulation of voice samples’ duration

and to the choice of a stimulus type in voice attractiveness studies. Although more of these aspects still need to be studied (e.g., presentation rate of multiple sounds, within-rater consistency, etc.), our study provides evidence for formulating recommendations to voice attractiveness researchers. No effect of duration on attractiveness perception was shown for the range of our samples. Therefore, if a similar range of durations is obtained, no normalization is necessary. Nevertheless, if duration normalization with the PSOLA algorithm is applied, one must be cautious of the perceptual consequences. Our results showed that lengthening the samples affected attractiveness perception. Therefore, shortening the sample duration is preferred to lengthening. A more practical implementation may be to normalize the duration to the mean of sample duration to limit manipulation range. Although the different stimulus types investigated triggered reliable attractiveness judgments, it must be kept in mind when designing an experiment that words such as “bonjour” are more sensitive to the deleterious effects of duration lengthening than are more simple sounds like vowels.

Acknowledgments The authors wish to thank Benoît Bediou, Sophie Jarlier, Christophe Mermoud, and Lucas Tamarit for technical help, Mylena Da Paz Cabral and Olga Vorontsova for their help in stimulus collection, and Juan David Leongómez for useful discussions on the design. This research was financed by a grant from the Swiss National Science Foundation (100014_130036) and was supported by the National Center of Competence in Research “Affective Sciences” (51NF40-104897), hosted by the University of Geneva.

References

- 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.
- Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.2.46). [Computer program]. Retrieved from <http://www.praat.org/>
- Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1), 55–59.
- Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Morinaga Shearman, S. (2002). Attitudinal and affective response toward accented English. *Language & Communication*, 22(2), 171–185.
- Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., et al. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20(2), 116–120.
- Bruckert, L., Lienard, J.-S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men’s characteristics. *Proceedings of the Royal Society B: Biological Sciences*, 273(1582), 83–89.
- Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters*, 5(1), 12–15.
- Collins, S. A. (2000). Men’s voices and women’s choices. *Animal Behaviour*, 60(6), 773–780.

- Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, *65*(5), 997–1004.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179.
- Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, *37*(4), 615–623.
- Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., Tiddeman, B. P., et al. (2005a). The voice and face of woman: one ornament that signals quality? *Evolution and Human Behavior*, *26*(5), 398–408.
- Feinberg, D. R., Jones, B. C., DeBruine, L. M., O'Connor, J. J. M., Tigue, C. C., & Borak, D. J. (2011). Integrating fundamental and formant frequencies in women's preferences for men's voices. *Behavioral Ecology*, *22*(6), 1320–1325.
- Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G., et al. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior*, *49*(2), 215–222.
- Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005b). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561–568.
- Feldstein, S., Dohm, F. A., & Crown, C. L. (2001). Gender and speech rate in the perception of competence and social attractiveness. *Journal of Social Psychology*, *141*(6), 785–806.
- 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.
- Fischer, J., Semple, S., Fickenscher, G., Jürgens, R., Kruse, E., Heistermann, M., & Amir, O. (2011). Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime. *PLoS ONE*, *6*(9), e24490.
- 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.
- Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, *25*(5), 295–304.
- Hughes, S. M., Harrison, M. A., & Gallup, G. G. (2002). The sound of symmetry: voice as a marker of developmental instability. *Evolution and Human Behavior*, *23*(3), 173–180.
- Hughes, S. M., Pastizzo, M. J., & Gallup, G. G. (2008). The sound of symmetry revisited: subjective and objective analyses of voice. *Journal of Nonverbal Behavior*, *32*(2), 93–108.
- 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), 192–194.
- Kline, P. (1993). *Handbook of psychological testing*. London: Routledge.
- Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, *75*(3), 817–822.
- Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145.
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communications*, *9*(5–6), 453–467.
- Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: the role of voice production. *Biological Psychology*, *87*(1), 93–98.
- Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, *6*(11), e27256.
- Pipitone, R. N., & Gallup, G. G. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, *29*(4), 268–274.
- Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, *26*(5), 388–397.
- Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 601–609.
- Puts, D. A., Barndt, J. L., Welling, L. L. M., Dawood, K., & Burriss, R. P. (2011). Intrasexual competition among women: vocal femininity affects perceptions of attractiveness and flirtatiousness. *Personality and Individual Differences*, *50*(1), 111–115.
- Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296.
- Saxton, T. K., Caryl, P. G., & Roberts, S. C. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: the ontogeny of mate choice. *Ethology*, *112*, 1179–1185.
- Saxton, T. K., DeBruine, L. M., Jones, B. C., Little, A. C., & Roberts, S. C. (2009). Face and voice attractiveness judgments change during adolescence. *Evolution and Human Behavior*, *30*(6), 398–408.
- Shrivastav, R., Camacho, A., Patel, S., & Eddins, D. A. (2011). A model for the prediction of breathiness in vowels. *The Journal of the Acoustical Society of America*, *129*(3), 1605–1615.
- Shrivastav, R., Sapienza, C. M., & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *Journal of Speech, Language, and Hearing Research*, *48*(2), 323–335.
- Vukovic, J., Feinberg, D. R., Jones, B. C., DeBruine, L. M., Welling, L. L. M., Little, A. C., & Smith, F. G. (2008). Self-rated attractiveness predicts individual differences in women's preferences for masculine men's voices. *Personality and Individual Differences*, *45*(6), 451–456.
- Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L. M., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, *102*, 37–48.
- Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record*, *58*(3), 475–482.