Body Height Perception from Connected Speech: Raising the Fundamental Frequency Increases Accuracy

ABSTRACT

Human listeners can perceive the speaker's body size from its voice to a certain degree. The voice pitch or speaking fundamental frequency (Fo) and the vocal formant frequencies are the voice parameters most intensively studied in research on the perception of body size (particularly height). Pisanki et al. (2014) found that artificially lowering Fo of isolated vowels pronounced by male speakers improved the perception of body height by part of listeners. The authors explained this effect as a denser harmonic spectrum provided by low pitch allowed for better resolution of formants, aiding formant-based size assessment.

In the present study, we tested and extended this research using connected speech (sentences, words) pronounced by speakers of both sexes. Unexpectedly, our results showed that raising Fo, not lowering it, increases the performance in two binary discrimination task of body size. This new finding is explained in the temporal domain by the dynamic and time-varying acoustic properties of connected speech. One possible reason is that an increase of Fo might increase the sampling density of the wave acoustic cycles and provide more detailed information, such as higher resolution, on the envelope shape.

KEYWORDS: speech perception, speaker body size, height, fundamental frequency, pitch.

Human voice is a multidimensional signal that conveys a great amount of information about the speaker: his/her sex, age, individual identity, emotional state and physical characteristics, such as her/his body size (height and weight). Among several voice parameters, both vocal formants and speaking fundamental frequency (pitch) predict the considerable variation in body size between humans of different ages and sexes (Peterson & Barney, 1952; Titze, 1989). However, when individual differences in body size are controlled by age and sex variables (e.g., adults of the same sex), only formants reliably predict speaker's size (González, 2004, 2006; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014a; Pisanski et al., 2014b). The formants are the resonant frequencies of the vocal tract and are dependent on the size of the vocal tube (Fant, 1960). As the vocal tract length (VTL) is larger, the formant frequencies are lower, and vice versa. Vocal tract is constrained to some extent by anatomical structures related to body size (e.g., individual's skull and body height). Consequently, a negative correlation exists between formant frequencies and speaker's body size, although its value is moderate in general terms (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; González, 2004, 2006; Pisanski et al., 2014b). Conversely, the correlation between pitch, or speaking fundamental frequency (Fo), and body size among adults of the same sex is very weak, or virtually null (Bruckert et al., 2006; González, 2007; Künzel, 1989; Pisanski et al., 2014b; van Dommelen & Moxness, 1995). Pitch is produced by vocal folds, which are composed of a soft tissue not subject to the same anatomical constraints than vocal formants.

Listeners can roughly infer the speaker's body size (specifically the body height) based (correctly) on the formant frequencies of his/her voice. Unexpectedly, research has shown that listeners incorrectly infer the speaker's height among adults of the same sex based on the pitch or

fundamental frequency (Fo) of her/his voice. Listeners consistently associate low speaking Fo with tall people, and, conversely, high Fo with short people (Feinberg, Jones, Little, Burt, & Perrett, 2005; Pisanski & Rendall, 2011; Pisanski et al., 2014b; van Dommelen & Moxness, 1995). Presumably, this association could be the result of an erroneous overgeneralization of the physical principle that large or long objects produce low frequency vibrations and small or short objects give rise to higher vibrations. This extrapolation is wrong because, as mentioned above, the size of the vocal cords is hardly related to the body size among adults of the same sex.

Given the absence of an objective link between voice pitch and body size at the within-sex level, pitch has traditionally been hypothesized that its presence in the human voice could make it difficult to correctly perceive the speaker's body size from the vocal formant frequencies. Pisanski, et al. (2014a) performed a direct test of how the accuracy of listeners' judgments is affected when voice pitch is present or absent from the acoustical signal. Counterintuitively, they found the listeners' accuracy increased in the presence rather than absence of voice pitch in the experimental stimuli. More specifically, height estimations based on relatively low voice pitch achieved higher accuracy. These authors hypothesized that the denser harmonic spectrum provided by low pitch allowed for better resolution of formants, aiding listeners in extracting reliable formant-based information from the voice. They tested this hypothesis (Exp. 3) and verified that artificially lowering the fundamental frequency of voice with the help of technology facilitated assessments of men's body height.

Pisanski, et al. (2014a, Exp. 3) employed the five Canadian English monophthong vowels (/ɑ/, /i/, /ɛ/, /o/, and /u/) recorded from thirty men as experimental stimuli, and subsequently, their fundamental frequency was manipulated using the PSOLA algorithm of Praat software (Boersma & Weenink, 2013) maintaining formants constant. The aim of the present work was to

extend this research in relation to several relevant aspects: a) using connected speech (sentences, words) as experimental study, beyond isolated vowels; b) using voices of men and women; and c) using materials from another linguistic context (Spanish).

EXPERIMENT 1

The purpose of Experiment 1 was to observe the effect of lowering or raising the fundamental frequency of a sentence spoken by men and women in assessments of speakers' body size.

METHOD

Participants.

Participants as listeners were 68 young adults of both sexes (57 females) whose age range was 18–30 years (M = 19.67; SD = 2.52). All participants were undergraduate students at the University Jaume I (Spain), who participated voluntarily with informed consent and were compensated with course credit.

Materials

The stimuli consisted of a Spanish interrogative sentence ('¿Cuántos años tiene tu primo de Barcelona?' [How old is your cousin from Barcelona?]) recorded from 40 young adult speakers, 20 men and 20 women, who had been university students several years ago, unknown to the participants. Voices were selected from a pool of voice recordings (González & Oliver, 2005), whose speakers had their height measured with a metric tape. The range of heights of the selected male speakers was 160-189 cm, with an average M = 176.9 cm (± SD = 6.3 cm), close to the general population of Spanish men (176.6 cm; NCDRISC, 2018). The range of heights of

the selected female speakers was 154-175 cm, with an average M = 163.4 cm (± SD = 7.5 cm), very close to the general population of Spanish women (163.4 cm; NCDRISC, 2018).

The sentence was recorded in a sound-attenuated booth with a Shure SM58 microphone at an approximate distance of 12 cm from the mouth, and a Sony-TCD D-8 digital audiotape (DAT) recorder with a sample frequency of 44.1 kHz. Then, the speech signal was digitally transferred to a PC computer and converted to 16-bit WAV files. Finally, the files were equated in RMS (root mean square) amplitude.

Manipulation of the Fundamental Frequency (Fo). We used three sets of stimuli: the natural speech recordings above mentioned ('Normal'), the natural recordings with their Fo lowered ('Low'), and the natural recordings with their Fo raised ('High'). Following the same procedure as Pisanski, et al. (2014a, Exp. 3), 'Low' stimuli were the natural recordings manipulated by means of PSOLA algorithm of Praat software (Boersma & Weenink, 2016) to subtract 0.5 ERBs of the baseline Fo, maintaining formants constant. 'High" stimuli were obtained in the same way adding 0.5 ERBs to the baseline Fo. PSOLA (Pitch Synchronous Overlap and Add) is a digital signal processing technique initially proposed by Moulines & Charpentier (1990) and used for manipulating the pitch (or time) of an acoustic speech signal, without altering other parameters. PSOLA decomposes the speech waveform in small overlapping segments, and moves them further apart to decrease Fo, or closer together to increase Fo. ERB or equivalent rectangular bandwidth is a measure used in psychoacoustics, which gives an approximation of the perceived voice pitch. For example, adding or subtracting 0.5 ERB to 120 Hz is roughly equivalent to adding or subtracting 20 Hz.

Pairing of speech stimuli. We created a total of 240 pairs of stimuli, 40 for each Fo condition (Low, Normal, High) x speaker's sex (males, females). Each pair was formed by two

different stimuli belonging to the same Fo condition and speaker's sex, separated by 800 ms of silence. Normal/male condition was formed by 40 different pairs of male speakers, whose heights differed in a range of 5 – 19 cm (M = 9.3 cm, SD = 3.9 cm); half of the pairs included the taller man in the first place, and half in the second place. Normal/female condition was formed by 40 different pairs of female speakers, whose heights differed in a range of 9 – 17 cm (M = 13.9 cm, SD = 3.0 cm); half of the pairs included the taller woman in the first place, and half in the second place. The pairs were the same for the other two Fo conditions, Low and High.

The total of 240 pairs was divided in four sets of 60 pairs (10 per each Fo x speaker sex condition), ensuring that neither pair of speakers was repeated through the Fo conditions within the same set.

Procedure

Each participant was randomly assigned to one of the four sets of stimuli, resulting in a total of 60 size assessment trials per each participant. Participants individually performed the experiment in six short sessions of ten stimuli of the same Fo/speaker sex condition. The order of the sessions was randomized through the participants. At the beginning of each session, participants were informed if the speakers were men or women.

On each trial, participants were presented through headphones with two speaker's voices of the same Fo condition (Low, Normal, High) and speaker's sex (males, females). Voices were played consecutively and separated by 800 ms of silence. After listening to the pair of voices, participants were asked to indicate which of the two voices belonged to the taller speaker by selecting the corresponding button on the screen. On each trial, participants could listen to the pair of voices again at will by clicking on a play button.

RESULTS AND DICUSSION

Contrary to expectations, it was raising Fo, rather than lowering Fo, which allowed increasing accuracy in estimating the speaker's body size. The overall percentages of successes of trials in which the taller speaker was correctly identified were the following for each Fo condition. For the 'Normal' condition (in which Fo was not modified), accuracy achieved 59.0%, 95% CI [55.5%, 62.6%], above chance level (50.0%; $p < .0001$). For the Low condition (Fo lowered - 0.5 ERBs), accuracy was 57.0% [53.5%, 60.5%], whereas for High condition (Fo raised + 0.5 ERBs) accuracy increased to 63.6% [60.5%, 66.6%]. Table 1 and Figure 1 shows the percentages separated by the sex of speakers. We can see the same trend for both sexes: accuracy increased for high-Fo sentences and roughly decreased for low-Fo sentences.

We carried out a 3 (Fo; Low, Normal, High) x 2 (Sex of Speakers) ANOVA. The analysis found a significant Fo effect, $F (2, 134) = 7.24$, $MS_e = 214.12$, $p = .001$, $\eta^2_p = .098$[1], and also a significant effect of the Sex of Speaker factor, $F (1, 67) = 9.54$, $MS_e = 230.21$, $p = .003$, $\eta^2_p = .125$, indicating that the female speakers received higher rates (62.2%, 95% CI [58.8%, 65.6%]) than the male speakers (57.5%, 95% CI [54.9%, 60.2%]). The Fo x Sex of Speaker interaction did not reach significance, $F (2, 134) < 1$. Posterior within-subject contrasts revealed that accuracy for the High Fo condition was significantly different (higher) from accuracy for the Normal condition, $F (1, 67) = 14.01$, $MS_e = 301.86$, $p < .0001$, $\eta^2_p = .173$.

Our data clearly suggested that by raising the fundamental frequency of the voice, listeners increased their accuracy in judging the speakers' body size. To examine this hypothesis, we performed a second experiment applying various degrees of Fo alteration and extending the research with stimuli consisted of words.

Please, Table 1 and Figure 1 about here

EXPERIMENT 2

The aim of Experiment 2 was to study the effect of lowering or raising two degrees (± 0.5, ± 1 ERBs) the pitch or speaking fundamental frequency (Fo) of a word in the perception of the speaker's body size by part of listeners. The use of various degrees of Fo multiplies the number of stimulus sets, so the experiment focalized on a sex of the speakers, specifically women, as Pisanski et al. 2014a employed only male speakers.

METHOD

Participants.

Participants as listeners were 64 young adults of both sexes (50 females) whose age range was 18–32 years (M = 19.56; SD = 2.67). All participants were undergraduate students at the University Jaume I (Spain), who participated voluntarily with informed consent and were compensated with course credit. None of the participants had participated in Experiment 1.

Materials

The stimuli consisted of a Spanish word ('importante' [important]) recorded from 20 young female speakers who had been university students several years ago, unknown to the participants. The speakers were the same as the female speakers in experiment 1. The words had been recorded and processed under the same conditions as the sentences from the experiment 1.

Manipulation of the Fundamental Frequency (Fo). We used five Fo conditions: the natural speech recordings ('Normal' condition), the natural recordings with their Fo lowered - 0.5 ERB ('- 0.5 ERB'), the natural recordings with their Fo lowered - 1.0 ERB ('- 1.0 ERB'), the natural recordings with their Fo raised + 0.5 ERB ('+ 0.5 ERB'), and the natural recordings with their Fo raised + 1.0 ERB ('+ 1.0 ERB'). The manipulation of Fo was carried out with the same technical procedure used in experiment 1 (PSOLA algorithm of Praat software, Boersma & Weenink, 2016).

Pairing of speech stimuli. We created a total of 150 pairs of stimuli, 30 for each Fo condition (- 1.0 ERB, - 0.5 ERB, Normal, + 0.5 ERB, + 1.0 ERB). Each pair was formed by two different stimuli belonging to the same Fo condition, separated by 800 ms of silence. Normal condition was formed by 30 different pairs of female speakers, whose heights differed in a range of 9 – 19 cm (M = 13.9 cm, SD = 2.9 cm); half of the pairs included the taller woman in the first place, and half in the second place. The pairs were the same for the other four Fo conditions.

Procedure

The total of 150 pairs was divided into three sets of 50 pairs (10 per each Fo condition). Each participant was randomly assigned to one of the three sets of stimuli, resulting in a total of 50 size assessment trials per each participant. Participants individually performed the experiment in two short sessions of twenty-five trials in random order.

On each trial, participants were presented through headphones with two women's voices of the same Fo condition. Voices were played consecutively and separated by 800 ms of silence. After listening to the pair of voices, participants were asked to indicate which of the two voices belonged to the taller woman by selecting the corresponding button on the screen. On each trial, participants could listen to the pair of voices again at will by clicking on a play button.

RESULTS AND DICUSSION

Table 1 and Figure 2 show the percentages of successes for each Fo condition. We found again that raising the pitch or Fo of an utterance (spoken word) tends to increase the accuracy of body size assessments, whereas lowering Fo tends to show the opposite effect, decreasing the accuracy in judging the body size, even though the effect size was smaller in this experiment than in the first experiment, and the consequences of Fo's manipulation become more evident in the most extreme degrees (- 1 ERB, + 1 ERB).

We performed a 5 (Fo condition; - 1.0 ERB, - 0.5 ERB, Normal, + 0.5 ERB, + 1.0 ERB) ANOVA. The analysis showed a marginal main effect of Fo, $F(4, 252) = 2.17$, $MS_e = 161.72$, $p = .073$, $\eta^2_p = .033$, and the polynomial test of within-subject contrasts was significant, $F(1, 63) = 6.87$, $MS_e = 193.75$, $p = .011$, $\eta^2_p = .098$. Posterior within-subject contrasts found significant differences between the - 1.ERB condition and the mean, and between the - 0.5 ERB vs. - 1.0 ERB conditions.

We performed an ANOVA with only the extreme Fo conditions (- 1.0 ERB, Normal, + 1.0 ERB). The analysis yielded a significant main effect of Fo, $F(2, 126) = 3.47$, $MS_e = 183.45$, $p = .034$, $\eta^2_p = .052$. A posterior within-subject contrast found a significant difference between the Normal vs. + 1.0 ERB conditions.

Please, Figure 2 about here

GENERAL DISCUSSION

The initial goal of the present study was to verify and extend the finding obtained by Pisanski et al. 2014a with isolated vowels to connected speech (sentences and words). Pisanski et al. (2014a) found in their third experiment that lowering the voice pitch or fundamental frequency (Fo) of a vowel helped listeners improve their accuracy in estimating the body size of the utterers (specifically male speakers). The authors explained this effect by means of the *Harmonic density hypothesis*. More specifically, increasing the density of harmonics in the speech spectrum (a consequence of lowering the fundamental frequency, because harmonics are multiples of Fo) would allow for better resolution of formants, aiding formant-based body size perception.

We tested this hypothesis employing connected speech as stimuli instead of isolated vowels. In the first experiment, we increased and decreased the Fo of a sentence pronounced by a set of male and female speakers. Unexpectedly, listeners perceived better the speaker's body size (height) when Fo was raised (not lowered), contradicting to the prediction made by the Harmonic density hypothesis. This effect was roughly corroborated in a second experiment based on a word spoken by female speakers, although the effect size was smaller. Apparently, this finding was counterintuitive and contrary to the prediction of the mentioned hypothesis.

One possible explanation could be in the temporal domain of stimuli rather than in the frequency domain. González and Oliver (2004) compared size estimations from connected speech (sentences) versus sustained vowels and found that connected speech yielded more accurate judgments than vowels. Logically, connected speech comprises a much larger voice sample than isolated vowels. However, connected speech is more than the sum of different phonemes and constitutes a dynamic and coarticulated succession of consonants and vowels

forming syllables as units of articulation (particularly in Spanish). There may be other acoustic predictors of body size beyond the mean frequencies of Fo and the formants (F1-F4). González and Oliver (2004) found that actual height was the strongest predictor of perceptual judgments from a spoken sentence, in both male and female speakers. This was more to be expected in female voices, given the low correlation between the acoustic parameters studied (Fo, F1-F4) and height, but it was also evident in male voices, in which the partial correlation between actual height and judgments of "taller" in a binary discrimination task reached 0.51 after removing the influence of Fo and F1-F4. These authors concluded that “partial correlations and multiple regression analysis showed that listeners efficiently rely on other features present mainly in connected speech (sentences as opposed to a sustained vowel)” (p. 29). They hypothesized that such features could be related to dynamic and time-varying acoustic properties of speech that could serve as sources of inferences about the biomechanical properties – mass, mobility and inertial characteristics – of the articulator apparatus and its possible relationship with body size. In this sense, the envelope of the acoustic wave of speech in the temporal domain could play a role in providing information on such dynamic and time-varying properties. In this case, an increase in the fundamental frequency would increase the sampling density of the wave acoustic cycles and provide more detailed information, such as higher resolution, on the envelope shape.

Irino, Aoki, Kawahara, and Patterson (2012) studied the interaction between Fo (glottal pulse rate, GPR) and mean-formant-frequency (MFF) with synthetized words in the perception of speaker characteristics and speech recognition. They observed that, within a certain range of MFF, increasing the Fo had the effect of improving the performance. The authors interpreted this result in the temporal domain of speech signal because as the glottal pulse rate (Fo) increases, the number of cycles per unit time also increases. They related this physical fact to the functioning

of the auditory system. “If the analysis performed by the auditory system is pitch synchronous and anchored to the glottal pulses as described in Irino and Patterson (2002), the internal representation will become much more stable as GPR increases from 0.25 to 4.0, and it seems reasonable to assume that speech features would be better defined in a more stable representation” (Irino et al., 2012, p. 1009). Irino and Patterson (2002) considered that human listeners hear, for example, vowels spoken by men and women as approximately the same vowel although the length of the vocal tract varies considerably from group to group. Simultaneously, listeners can identify the speaker group. They suggested that the auditory system can extract and segregate information about the size of the vocal tract from information about its shape. In accordance with Irino and Patterson (2002), the auditory system performs a transform (Mellin transform) creating an auditory image, in which the information on the vocal tract size is represented apart from the information corresponding to its particular shape. According with Irino et al. (2012), this auditory image will become more stable and efficient when Fo increases within a certain range.

Nevertheless, further research will be necessary to confirm the present finding and test the effect that lowering versus raising the speaking fundamental frequency could have on perception of speaker’s body size (and vocal tract length) using different speech materials beyond isolated vowels, and compare voices of male versus female speakers, along with children.

Footnotes

[1] The effect size interpretations for η^2_p values are: .01 = small, .06 = medium, and .14 = large.

Ethical Compliance Section

This work was completed with resources provided by the University Jaume I (Spain).

All procedures performed in studies involving human participants were in accordance with the Deontological commission and of the Ethical Committee of Animal Welfare of the University Jaume I (Spain) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

The authors declare they have no conflict of interest.

Written informed consent was obtained from all individual adult participants included in the study.

References

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.2.15). [Software]. Retrieved from http://www.praat.org.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.19). [Software]. Retrieved from http://www.praat.org.

Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 83-89.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal behaviour*, *69*(3), 561-568.

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

González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Developments in Applied Physics*, *9*, 1–15.

González, J. (2007). Correlations between speakers' body size and acoustic parameters of voice. *Perceptual and motor skills*, *105*(1), 215-220.

González, J., & Oliver, J. C. (2004). Percepción a través de la voz de las características físicas del hablante: identificación de la estatura a partir de una frase o una vocal [Perception through the voice of speaker physical characteristics: identification of height from a sentence or a vowel]. *Revista de psicología general y aplicada*, *57*(1), 21-34.

Gonzalez, J., & Oliver, J. C. (2005). Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *The Journal of the Acoustical Society of America*, *118*(1), 461-470.

Irino, T., Aoki, Y., Kawahara, H., & Patterson, R. D. (2012). Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination. *Speech Communication*, *54*(9), 998-1013.

Irino, T., & Patterson, R. D. (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. *Speech Communication*, *36*(3-4), 181-203.

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, *46*(1-3), 117-125.

NCDRISC (2018). NCD Risk Factor Collaboration (NCD-RisC). https://www.ncdrisc.org/height-mean-map.html.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech communication*, *9*(5-6), 453-467.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, *24*(2), 175-184.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., & Feinberg, D. R. (2014a). Return to Oz: Voice pitch facilitates assessments of men's body size. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1316.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., Röder, S., Andrews, P. W., ... & Feinberg, D. R. (2014b). Vocal indicators of body size in men and women: a meta-analysis. *Animal Behaviour*, *95*, 89-99.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201-2212.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*(4), 1699-1707.

van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and speech*, *38*(3), 267-287.

FIGURE CAPTIONS

Figure 1. Experiment 1 (sentences). Percentages of successes in a binary discrimination task on the speakers' body height, separated by experimental conditions according to the manipulation of Fo and the speaker sex.

Figure 2. Experiment 2 (words). Percentages of successes in a binary discrimination task on the female speakers' height, separated by experimental conditions according to the manipulation of Fo. ERB: Equivalent rectangular bandwidth.

TABLES

Table 1. Data from Experiments 1 (sentences) and 2 (words). Percentages of successes in a binary discrimination task on the speakers' body height, separated by experimental conditions according to the manipulation of Fo. ERB: Equivalent rectangular bandwidth.

Table 1. Data from Experiments 1 (sentences) and 2 (words). Percentages of successes in a binary discrimination task on the speakers' body height, separated by experimental conditions according to the manipulation of Fo. ERB: Equivalent rectangular bandwidth.

	Fo manipulation				
	-1.0 ERB	-0.5 ERB	Normal	+0.5 ERB	+1.0 ERB
Sentences (1st Exp.)					
male speakers		55.6 (17.8)	55.9 (14.8)	61.2 (14.9)	
female speakers		58.4 (17.6)	62.2 (19.7)	66.0 (16.0)	
Words (2nd Exp.)					
female speakers	61.0 (16.0)	63.1 (15.5)	65.0 (18.2)	65.1 (15.0)	67.3 (12.3)

Figure 1. Experiment 1 (sentences). Percentages of successes in a binary discrimination task on the speakers' body height, separated by experimental conditions according to the manipulation of Fo and the speaker sex.

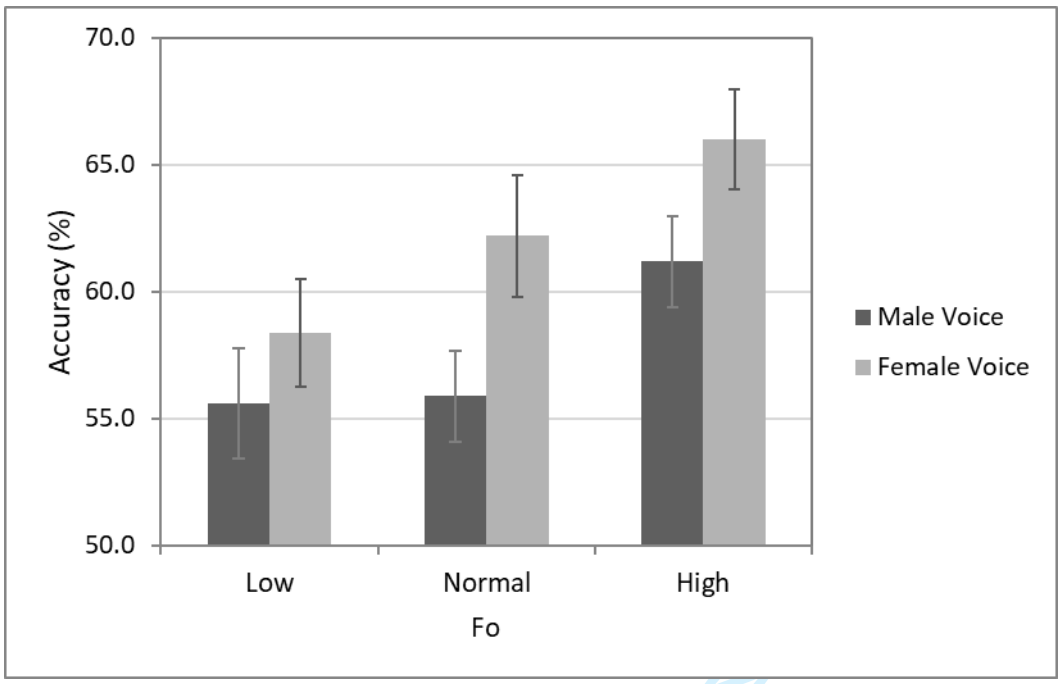

Figure 2. Experiment 2 (words). Percentages of successes in a binary discrimination task on the female speakers' height, separated by experimental conditions according to the manipulation of Fo. ERB: Equivalent rectangular bandwidth.