

Can you hear my body size?: Estimating speaker height and body build from speech*

Dayeon Yoon**

Konkuk Research Institute for Multilingualism and Multiculturalism, Konkuk University, Seoul, Korea

Abstract

This study aims to assess listeners' ability to estimate speakers' body size from speech. Twelve acoustic correlates related to resonance (F1, F2, F3, F4, and formant dispersion), pitch (f0), and breathiness/roughness (H1*-H2*, HNR05, HNR15, HNR25, HNR35, and CPP) were measured on /a/s in sentences spoken by 33 male and 37 female French and Korean speakers. Subsequently, 35 male and 70 female French and Korean listeners evaluated speakers' body height and build from speech, each on a 6-point scale. Results indicated that French listeners generally outperformed Korean listeners in estimating body height from speech, although overall accuracy was low to modest. Acoustic analyses showed that lower values of f0, formant structure (F1-F4), and harmonic measures (HNR05, CPP) are associated with larger body size; however, the direction and strength of these associations varied across speaker sex as well as speaker and listener language. These patterns suggest that listeners are sensitive to pitch and spectral information linked to apparent vocal fold and vocal tract size, yet such cues provide only coarse and often inconsistent inferences about actual height and build. These findings are interpreted as evidence that speech-body associations are shaped by sex-dependent and language-specific perceptual stereotypes, rather than serving as a reliable vocal indicator of physical dominance.

Keywords: body size perception, speaker sex, language-specific cue weighting, speech acoustics

1. Introduction

The differences in acoustic characteristics between male and female voices depend in part on anatomical characteristics. Fundamental frequency (f0) is one of the most discriminating acoustic correlates between men and women. The vocal folds,

generally shorter in women than men [about 10 mm in women compared to 15 mm in men (Filho et al., 2005)], allow faster vibrations in the female voice (207 Hz) than in the male voice (118 Hz; Boë et al., 1975). Vocal tract resonances are another discriminating characteristic between men and women related to morphological differences. Women have higher resonance

* This work, conducted as part of the author's doctoral dissertation research in the Laboratoire de phonétique et phonologie (CNRS/Univ. Sorbonne Nouvelle, Paris, France), was supported by the Global Korea Scholarship (National Institute for International Education, NIIED) and the Pony Chung Humanities Scholarship (PONY CHUNG Foundation).

** dayeon.yoon25@gmail.com, Corresponding author

Received 31 October 2025; Revised 9 December 2025; Accepted 10 December 2025

© Copyright 2025 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

frequencies due to a shorter vocal tract (on average 14 cm) than that of men (on average 17 cm; Fant, 1966). Harmonic amplitude differences such as H1-H2, H1-A3, etc., are often mentioned to differentiate between male and female voices because the differences in harmonic amplitude depend on the degree of closure of the glottis, which is influenced by the thickness of vocal folds (Södersten & Lindestad, 1990). Because female vocal folds are thinner [approximately 7 mm compared to 9 mm in males (Hollien, 1962)] and therefore more prone to incomplete closure that allows continuous airflow to escape, female voices tend to be perceived as more breathy than male voices (Hanson & Chuang, 1999; Iseli et al., 2007; Södersten & Lindestad, 1990). This characteristic of the vocal folds is reflected by other acoustic parameters such as harmonic/noise ratio (HNR) (Ambreen et al., 2019) and cepstral peak prominence (CPP) (Choi & Choi, 2016; Hillenbrand et al., 1994) in terms of spectral noise and periodicity.

According to earlier studies, a correlation is often found between acoustic characteristics of voice and body size in mammals. Fitch (1997) reports that the height, weight, vocal tract length (VTL), and skull length are all correlated with formant dispersion in both male and female macaque monkeys. Rendall et al. (2004) show that differences in f_0 and body size are twofold each between male and female baboon, and baboons discriminate males and females through voice alone at a significant level. Moreover, animals associate this link between acoustic characteristics of voice and body size to physical dominance (Morton, 1977; Ohala, 1980; Reby et al., 2005). This is because physical dominance is important within same-sex competition. In animals, lower pitch and resonances with harsh voice quality are associated with a larger body proportional to larger vocal folds and a larger vocal tract, which gives a physically more aggressive and dominant impression. Conversely, if an animal wants to avoid fighting and giving an aggressive impression, it produces a high-pitched clearer vocalization and shortens its VTL to convey a baby-like impression reminding of smaller vocal folds and vocal tract (Morton, 1977; Ohala, 1980). These findings in animals provide a useful framework for investigating whether similar acoustic cues may convey body size and dominance in humans, given that the basic principles of vocal production are shared across mammals.

However, the relationship between body size and acoustic characteristics of voice is controversial in human compared to other animal species. Many studies have analyzed correlations between voice characteristics and body size but no consensus has yet been reached. Some have found strong negative correlations between height and f_0 (Dusan, 2005; Hatano et al., 2012) and between weight and f_0 (Evans et al., 2006), as well as between height and formants (Dusan, 2005; Evans et al., 2006; Johnson, 2006). Strong positive correlations have also been reported between body mass index (BMI) and jitter, and between BMI and noise-to-harmonic ratio (NHR) (Da Cunha et al., 2011). In contrast, other studies have found only weak correlations between the same bodily indices and acoustic measures (González, 2004; Pisanski et al., 2014; van Dommelen & Moxness, 1995). This discrepancy is observed similarly at the perceptual level: Rendall et al. (2007) found that Canadian English listeners could estimate speakers' height using formants and f_0 when hearing isolated words, while van Dommelen & Moxness (1995) observed that Norwegian listeners were able to estimate only male speakers' height and weight, possibly relying on f_0 , F2-based estimated vocal tract length, and duration, in

experiments using isolated words and two paragraphs of text. The findings of van Dommelen & Moxness (1995) also suggest that humans' ability to estimate the speaker's body dimensions from acoustic cues may depend on both speaker and listener sex. Similarly, Charlton et al. (2013) found that male listeners showed better estimation ability than female listeners in experiments using synthesized vocalizations of animals of different sizes. However, some studies indicate that female listeners' performance depends on the specific body dimension being estimated. According to Collins (2000) and Brucker et al. (2006), Dutch and French female listeners, respectively, could estimate male speakers' weight from isolated vowels, but not their height.

Another interesting issue here is that whether listeners succeed or fail in estimating speakers' body size from voice, they exhibit a consistent tendency in using specific acoustic information during estimation. For example, a lower f_0 and lower formants give information that the speaker's body size is relatively larger and heavier. Conversely, a higher f_0 and higher formants give impression that the speaker's body size is relatively smaller and thinner (Cartei et al., 2014; Charlton et al., 2013; Rendall et al., 2007; Uchida, 2022; van Dommelen & Moxness, 1995). In terms of voice quality, which is understudied compared to f_0 and formant frequencies, breathy voice with increased H1*-H2*, H1*-A1*, and H1*-A3* reminds of a smaller body, together with a higher f_0 and wider formant dispersion (Xu et al., 2013).

Finally, previous research highlights cross-cultural variation in body size perception. For instance, Koreans are known to hold particularly stringent aesthetic norms regarding body shape (Jung & Forbes, 2007). Similarly, in many Western societies, thinness is widely regarded as a key standard of physical attractiveness (McCabe et al., 2013). In contrast, preferences for thinness appear to be less pronounced in several African cultures (Szabo & Allwood, 2006) and among French populations (Maezono et al., 2018) compared to other Western contexts. Taken together, these findings suggest that listeners from different cultural backgrounds may rely on distinct visual or auditory cues when judging physical attributes from speech.

The present study further examines listeners' ability to estimate speakers' body size from speech and investigates acoustic characteristics of speech related to perceived body size, comparing French and Korean participants. For acoustic characteristics, some dimensions associated to phonation, resonance, and voice quality are analyzed: f_0 , formant frequencies, and formant dispersion are studied because they show clear differences between men and women due to sexually dimorphic anatomy (Boë et al., 1975; Fant, 1966); for sex-dependent voice quality, we observe descriptors linked to spectral tilt [H1*-H2* (Hanson & Chuang, 1999; Iseli et al., 2007)] and to spectral noise [HNR (Ambreen et al., 2019) and CPP (Choi & Choi, 2016; Hillenbrand et al., 1994)]. The main hypotheses formulated in this study are as follows: 1) Listeners can perceive speakers' height and build from speech, with performance varying depending on the sex and language of the listener and the speaker; 2) Speakers are perceived as taller and heavier when their vocal pitch, resonance frequencies, and breathiness are lower, and roughness is higher—that is, when f_0 , formants, H1*-H2*, CPP, and HNR values are lower.

2. Methods

2.1. Recording

Thirty-three males and 37 female speakers, aged 18 to 42, were recorded. These speakers were recruited in France and in Korea as part of a wider project on the effect of language and culture on sex-dependent voice properties. A two-way ANOVA revealed a significant main effect of sex on height [$F(1, 66)=106.39, p<.001$] and weight [$F(1, 66)=67.54, p<.001$], with males taller and heavier than females in our data. French males ($M=177.5$ cm, $SD=5.80$) and Korean males ($M=175.35$ cm, $SD=6.63$) were taller than French ($M=163.06$ cm, $SD=6.30$) and Korean females ($M=160.85$ cm, $SD=4.72$). Similarly, French males ($M=68.44$ kg, $SD=7.97$) and Korean males ($M=74.86$ kg, $SD=12.48$) were heavier than French ($M=55.53$ kg, $SD=6.98$) and Korean females ($M=53.48$ kg, $SD=6.96$). A two-way ANOVA on weight revealed a marginal language \times sex interaction [$F(1, 66)=4.01, p=.049$], with Korean male speakers being heavier than French male speakers ($F=6.42, p=.041$). Most French speakers speak Northern Metropolitan French, and all Korean speakers speak the Seoul/Gyeonggi dialect. None of the speakers reported hearing or speech impairments.

The speakers were recorded in quiet rooms with professional equipment (Shure SM10A microphone placed at a 45-degree angle and 7 to 8 centimeters from the speakers' mouth and PreSonus Studio 24c sound card).

Speakers read aloud the logatomes *mama*, *mamama*, and *mamamama*, embedded in carrier sentences of comparable length and structure in either French or Korean. Korean speakers read the phrase /onil ohue, mama, mamama, mamamama, seei kojwane kagirotta/, while French speakers read the phrase /set apɛ midi, mama, mamama, e mamamama, nuz iɔ̃ o paɔk də lil sɛ̃t ida/, both of which correspond in meaning to *This afternoon, mama, mamama, and mamamama, we will go to the park* (with *Sainte Ida park* specified for the French sentence).

2.2. Acoustic Measurements

In the recorded samples, the vowels /a/ of target words were analyzed. For the measurement of vibration mode, the open vowel like /a/ is most commonly used in the literature (Storck & Drinnan, 2008; Kreiman et al., 2021; Wagner & Braun, 2003), as for closed vowels, there is a risk of f_0 approaching F_1 , thus causing detection errors. For the vowel /a/, whose F_1 value is relatively higher compared to other vowels, there is less risk of f_0 and F_1 approaching each other. Thus, the distinction between source and filter is well preserved.

Acoustic analyses focused on resonance, phonation, and voice quality. Resonance was quantified using formant frequencies (F_1 - F_4) and formant dispersion, extracted with the Burg LPC algorithm using a 25 ms Gaussian window, detecting up to five formants within 0–5 kHz for males and 0–5.5 kHz for females, and measured at five equidistant points across the selected interval. Phonation was assessed via f_0 within 70–300 Hz for males and 100–600 Hz for females, using the STRAIGHT algorithm (Kawahara et al., 1999). Breathiness and roughness were evaluated using formant-corrected spectral shape measures ($H1^*-H2^*$) calculated with the Snack Sound Toolkit (Hanson, 1997; Iseli et al., 2007; Sjölander, 2004), while spectral noise

was characterized with HNR (de Krom, 1993) across multiple frequency bands (0–500 Hz, 0–1,500 Hz, 0–2,500 Hz, 0–3,500 Hz) and CPP (Hillenbrand et al., 1994). Formant frequencies were measured with Praat (Boersma & Weenink, 2022), and all the other acoustic variables were measured with VoiceSauce (Shue et al., 2011) every five milliseconds over the vowel target intervals.

Regarding the values of acoustic measurements, considering that the human auditory system is not linear but rather logarithmic, the f_0 and formants measured in Hertz were converted into the scales that best correspond to perception, before the correlation was calculated: f_0 into semitones and formants into Bark (Traummüller, 1990). All measures extracted within each vowel interval were then averaged across frames prior to statistical analysis.

2.3. Listening Test

Thirty-five males (19 Korean and 16 French; M age=30.6 years, $SD=10.7$) and seventy females (20 Korean and 50 French; M age=24.4 years, $SD=5.6$) participated in the listening test. The words *mama*, *mamama*, and *mamamama* extracted from the sentences were used as test stimuli. A fade-in and a fade-out of 3 ms each were inserted at the beginning and end of each target word. 200 ms of silence was also inserted between the words. The participants were given 70 recorded samples and 8 additional ones to evaluate intra-rater reliability. These 8 samples were composed of 2 French males, 2 French females, 2 Korean males, and 2 Korean females. They were separately arranged before and after the break session so as not to be presented consecutively with the same sample, and then presented randomly along with the other 70 samples.

The test was programmed using PsyToolkit and distributed and conducted online (Stoet, 2010, 2017). Before the start of the experiment, participants practiced with 5 other samples not included in the analysis. During the experiment, the participants listened to the voice samples and estimated speakers' height within six levels from 'very short' to 'very tall' and speakers' body build within six levels from 'very thin' to 'very fat'. The 6-point scale was used to prevent participants from choosing only the middle value. Each voice sample was presented only once.

Right after the participants responded to the questions about height and weight in order, the next sound sample was automatically played with a new answer screen. At the end of the test, participants entered their sex, mother tongue, age, and educational backgrounds in phonetics or speech language and hearing.

2.4. Data Analysis

To check the intra-rater reliability of individual responses, we compared participants' responses for the 8 pairs of stimuli provided twice in the test. Inspired by the percent agreement method, which assesses the proportion of ratings that are either identical or fall within one adjacent level (Altman, 1991; Gisev et al., 2013), responses were considered consistent if a participant's responses showed a difference of 0 or 1 between the two responses for each pair of stimuli (the perceptual test using a six-level scale, the maximum possible difference for a pair being five). Thus, for each pair of stimuli, 1 point was

awarded for consistent responses between the stimuli in a pair, and 0 points were awarded for inconsistent responses. Ultimately, the sum of all participants' scores was calculated, then the average was established to obtain the intra-rater reliability. Since 1 point was awarded for consistent responses, the result would be 1 if all participants had provided consistent responses for all pairs of stimuli.

To assess the inter-rater reliability, a mutual comparison of each participant's responses to the 78 stimuli provided in the test was conducted. Using a percentage agreement approach, as in the intra-rater reliability test (Altman, 1991; Gisev et al., 2013), a participant's response to a particular stimulus was considered consistent if it differed by 0 or 1 from the responses of the other 104 participants for the same stimulus. In this manner, 1 point was awarded for consistent responses between two participants, and 0 points were awarded for inconsistent responses. Then, the average of the scores each participant obtained was calculated. Since 1 point was awarded for consistent responses with each of the other participants for each stimulus, the result would be 1 if a participant had provided consistent responses with others for all stimuli. Finally, the mean values of scores from all participants were averaged again as the criterion for inter-rater reliability. If all participants had provided consistent responses with each other for all stimuli, the result would be 1.

For intra-rater reliability, participants in our study demonstrated a reliability score of 0.81 for questions regarding the height and build of the speakers, respectively, indicating consistency in their responses. For inter-rater reliability, participants in our study showed a consistency score of 0.7 in their responses to questions about speaker height and a consistency score of 0.71 in their responses to questions about speaker body build. Therefore, it was considered that response consistency among participants was ensured.

3. Results

3.1. Effects of Sex and Language of Speakers and Listeners

Regardless of the listener's sex, the body height of female speakers is widely estimated between scores 2 and 4 on the height rating scale, with the highest density at score 3 for female speakers measuring between 158 and 163 cm (which represents approximately 51% of the female population). The height of male speakers is primarily estimated between scores 3 and 5 by both male and female listeners, with maximum density at score 4 for male speakers measuring approximately between 173 and 180 cm (which represents approximately 64% of the male population).

For the estimated body build, regardless of the listener's sex, the build of female speakers is largely estimated between scores 2 and 4 on the rating scale, with a peak density at score 3 for female speakers weighing approximately 53–55 kg (representing about 30% of the female population). The build of male speakers is primarily estimated between scores 3 and 5 by both male and female listeners, with a peak density at score 4 for male speakers weighing between 65 and 80 kg (representing approximately 64% of the male population).

The influence of speaker sex, speaker and listener language,

and their interactions on perceived height and build scores were thus examined using a linear mixed model with the *lmer* function in the *lme4* package (Bates et al., 2014) for R (R Core Team, 2021). Random intercepts for speaker and listener were included in the model. Results showed significant main effects of speaker sex on both perceived height [$F(1, 71)=125.81, p<.001$] and perceived body build [$F(1, 70.6)=47.99, p<.001$], with male speakers receiving higher ratings (perceived height: female speakers, $M=3.20$; male speakers, $M=3.96$; perceived body build: female speakers, $M=2.93$; male speakers, $M=3.48$). Speaker language also had a significant effect on perceived height [$F(1, 71)=34.42, p<.001$], with French speakers receiving higher scores (Korean speakers, $M=3.38$; French speakers, $M=3.78$). Moreover, listener language had significant effects on perceived body build [$F(1, 104.9)=4.80, p=.031$], with higher scores for French listeners (Korean listeners: $M=3.13$; French listeners: $M=3.29$). Additionally, there were significant interactions between speaker sex and listener language [$F(1, 7164)=42.92, p<.001$ for perceived height; $F(1, 7164)=8.55, p=.003$ for perceived build] and speaker language and listener language [$F(1, 7164)=29.10, p<.001$ for perceived height; $F(1, 70)=74.13, p<.001$ for perceived build].

Post-hoc test revealed that, for perceived height, the difference between male and female speakers ($E=-0.6, p<.001$ in KR vs. $E=-0.91, p<.001$ in FR) and the difference between Korean and French speakers ($E=-0.27, p=.0003$ vs. $E=-0.53, p<.001$ in FR) were more pronounced in French listeners than in Korean listeners. For perceived build, the difference between male and female speakers was more pronounced in French listeners ($E=-0.49, p<.001$ in KR vs. $E=-0.62, p<.001$ in FR), while the difference between Korean and French speakers was more marked in Korean listeners ($E=-0.30, <.001$ vs. $E=0.10, p=.23$ in FR). As listener language modulated the effects of speaker language and sex, subsequent analyses were conducted separately for each listener language.

To assess listeners' ability to estimate speaker height and body build from speech, Pearson correlation coefficients were calculated between actual and estimated height, separately for male and female speakers of each language in each language of listeners. Similarly, correlations between actual and estimated weight were computed for male and female speakers of each language in each listener group. To examine whether acoustic measurements were associated with listener judgments, the effects of acoustic cues, speaker sex, speaker language, and their interactions on perceived height and build scores were investigated separately for each listener language using linear mixed models, with random intercepts for both speaker and listener.

3.2. Listeners' Ability to Estimate Speakers' Body Size

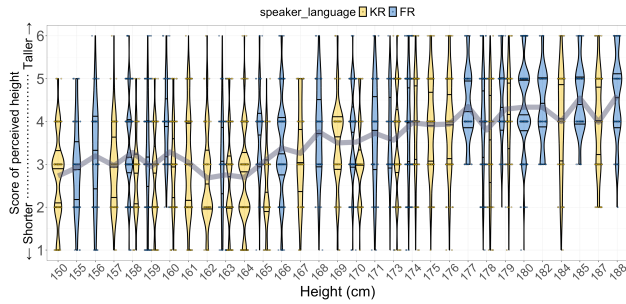


Figure 1. Violin plot of the distribution of estimated height (1: small, 6: tall) for each actual height rated by French listeners. The gray trend line connects the average perceived height scores for each speaker (KR, Korean speakers; FR, French speakers).

Results revealed that French listeners showed a moderate positive correlation between actual and estimated height of speakers ($r=0.37$, $p<.001$ for FR speakers; $r=0.38$, $p<.001$ for KR speakers), as shown in Figure 1. The gray trend line indicates that the average perceived height score increases with actual height, suggesting a moderate association indicating that French listeners are able to detect speaker height to some extent from speech. However, French listeners exhibited a weak correlation between actual and estimated height in terms of speaker sex ($r=0.12$, $p<.001$ for female speakers; $r=0.20$, $p<.001$ for male speakers). Korean listeners showed only weak correlations between actual and estimated height of speakers both in terms of speaker sex ($r=0.12$, $p<.001$ for female speakers; $r=0.20$, $p<.001$ for male speakers) and speaker language ($r=0.28$, $p<.001$ for FR speakers; $r=0.32$, $p<.001$ for KR speakers).

For body build, the results show a weak relationship between actual weight and perceived build. French listeners showed only weak correlations between actual weight and estimated body build of speakers both in terms of speaker sex ($r=0.11$, $p<.001$ for female speakers; $r=0.08$, $p<.001$ for male speakers) and speaker language ($r=0.26$, $p<.001$ for FR speakers; $r=0.26$, $p<.001$ for KR speakers). Similarly, Korean listeners showed only weak correlations between actual weight and estimated body build of speakers both in terms of speaker sex ($r=0.13$, $p<.001$ for female speakers; $r=0.07$, $p<.001$ for male speakers) and speaker language ($r=0.23$, $p<.001$ for FR speakers; $r=0.25$, $p<.001$ for KR speakers).

3.3. Relationships Between Acoustic Parameters and Judgments

To provide an overview of the main patterns observed in the data, Table 1 summarizes the effects of key acoustic cues on perceived height and body build across listener language (Korean vs. French). The table highlights the direction of each effect, indicating whether higher or lower values of a given cue tended to increase perceptions of height or build, as well as any notable listener- or speaker-specific variations.

3.3.1. Perceived height and related acoustic cues

As expected, speakers with smaller F1 values were generally perceived as taller. For Korean listeners, significant main effects were found for speaker sex [$F(1, 69.68)=18.59$, $p<.001$] and F1 [$F(1, 69.68)=27.86$, $p<.001$]. Importantly, F1 interacted significantly with speaker sex [$F(1, 69.68)=15.32$, $p<.001$] and with both sex and language in a three-way interaction [$F(1, 69.68)=4.91$, $p=.030$]. Smaller F1 values were associated with taller perceptions for

Table 1. Main effects of acoustic cues on Korean and French listeners' perception of body dimensions, with a summary of key results

Acoustic cue	Body dimension	Korean listeners	French listeners	Interpretation
F1	H	O	O	Higher F1 → shorter (steeper slope for KR males; FR females deviate positively)
	B	O	O	Higher F1 → thinner (steeper slope for males)
F2	H	O	X	Higher F2 → shorter (steeper slope for males)
	B	X	O	Higher F2 → thinner (for FR listeners)
F3	H	X	X	No effect
	B	O	X	Higher F3 → thinner (for KR listeners)
F4	H	O	X	Higher F4 → shorter (for KR listeners; steeper slope for KR)
	B	O	O	Higher F4 → thinner (negative slope for KR; FR males deviate with positive slope in FR listeners)
Formant dispersion	H	X	X	No effect
	B	X	X	No effect
f0	H	O	O	Lower f0 → taller across groups
	B	O	O	Lower f0 → heavier (For FR males, the slope is positive in KR listeners, and negative but shallower in FR listeners)
H1*-H2*	H	X	X	No effect
	B	X	X	No effect
HNR05	H	X	X	Higher HNR05 → shorter (KR females show positive slope)
	B	X	O	Higher HNR05 → thinner (for FR listeners; KR females deviate positively)
CPP	H	X	X	Higher CPP → shorter in males, taller in females (for KR listeners)
	B	X	O	Higher CPP → thinner (for FR listeners; KR females deviate positively)

H, height; B, body build; O, significant; X, not significant; FR: French; KR, Korean

Korean male ($\beta=-0.437$) and female ($\beta=-0.256$) speakers and French male speakers ($\beta=-0.544$), while smaller F1 corresponded to slightly shorter perceived height for French female speakers ($\beta=0.110$). These results indicate that the role of F1 in height perception varies by speaker sex and depends on speaker language. For French listeners, F1 also exerted a strong effect on height perception, with significant main effects of speaker sex [$F(1, 69.58)=10.91, p=.0015$] and F1 [$F(1, 69.58)=26.28, p<.001$]. The speaker sex×F1 interaction was significant [$F(1, 69.58)=5.78, p=.0188$], indicating a stronger negative relationship in male speakers ($\beta=-0.394$) compared to female speakers ($\beta=-0.142$). No other F1 interactions were significant.

Smaller F2 values were also associated with taller perceptions, with the magnitude varying by speaker sex. For Korean listeners, F2 had a significant main effect [$F(1, 69.73)=9.30, p=.003$] and interacted with speaker sex [$F(1, 69.73)=5.11, p=.027$], showing a stronger negative relationship for male speakers ($\beta=-0.465$) than female speakers ($\beta=-0.069$). No language-related effects were observed. Similarly, French listeners showed a significant main effect of speaker sex [$F(1, 69.62)=7.49, p=.0079$] and a speaker sex×F2 interaction [$F(1, 69.62)=5.36, p=.0236$], with a negative relationship for male speakers ($\beta=-0.355$) and a slight positive relationship for female speakers ($\beta=0.033$).

F3 did not predict height judgments for either listener group. F4 significantly affected height ratings for Korean listeners [$F(1, 69.81)=8.63, p=.004$], and interacted with language [$F(1, 69.81)=5.40, p=.023$], with a stronger negative effect for Korean speakers ($\beta=-0.765$) than French speakers ($\beta=-0.089$). No reliable effects of F3 or F4 were observed for French listeners. Formant dispersion did not significantly influence height perception in either group.

Regarding f0, significant main effects were found for both Korean [$F(1, 69.60)=5.29, p=.024$] and French listeners [$F(1, 69.58)=20.74, p<.001$], with higher f0 generally associated with lower perceived height, regardless of speaker sex or language.

H1*·H2* did not significantly predict height ratings. However, strong main effects of speaker sex [$F(1, 69.64)=22.65, p<.001$ for KR; $F(1, 69.65)=65.46, p<.001$ for FR] and speaker language [$F(1, 69.64)=6.87, p=.011$ for KR; $F(1, 69.65)=27.97, p<.001$ for FR] were observed, with male speakers perceived as taller than female speakers ($E=3.92$ vs. 3.38 for KR; $E=3.96$ vs. 3.13 for FR).

Across HNR measures (HNR05, HNR15, HNR25, HNR35), none showed significant main effects. HNR05 showed a trend in which voices with greater harmonic structure were judged as smaller, with effects depending on speaker sex or language. For Korean listeners, a significant interaction between HNR05 and speaker sex emerged [$F(1, 69.73)=6.68, p=.012$], with higher HNR05 associated with smaller perceived height for French male ($\beta=-0.014$) and female ($\beta=-0.001$) speakers and Korean male speakers ($\beta=-0.043$), while lower HNR05 corresponded to smaller height for Korean female speakers ($\beta=0.022$). For French listeners, too, a significant three-way interaction [$F(1, 69.62)=5.54, p=.021$], indicated similar patterns, with $\beta=-0.006$ for French male speakers, $\beta=-0.025$ for French female speakers, $\beta=-0.032$ for Korean male speakers and $\beta=0.015$ for Korean female speakers. Higher HNR bands (HNR15, HNR 25, and HNR35) showed no significant effects in either listener group.

CPP did not significantly predict height estimation for Korean listeners, although speaker sex strongly affected ratings [$F(1, 69.56)=13.27, p<.001$], with a significant interaction between sex and CPP [$F(1, 69.57)=8.43, p=.005$]. Negative relationships were observed for male speakers ($\beta=-0.053$) and positive relationships for female speakers ($\beta=0.046$). For French listeners, only speaker sex significantly affected perceived height [$F(1, 69.66)=8.05, p=.006$], with male speakers perceived as taller ($E=4.00$) than female speakers ($E=3.06$).

3.3.2. Perceived body build and related acoustic cues

For Korean listeners, F1 had a main effect on perceived body build [$F(1, 69.60)=48.63, p<.001$], with speakers exhibiting higher F1 consistently judged as lighter. A smaller but significant main effect of speaker sex was also observed [$F(1, 69.61)=4.34, p=.041$], with male speakers rated as heavier overall, independent of acoustic manipulation. Importantly, F1 interacted with speaker sex [$F(1, 69.60)=4.26, p=.043$], showing a negative relationship in male speakers ($\beta=-0.598$) than in female speakers ($\beta=-0.325$). For French listeners, only the main effect of F1 was significant [$F(1, 69.58)=30.29, p<.001$].

For Korean listeners, F2 did not significantly affect weight ratings, nor did any interaction involving F2 reach significance. In contrast, French listeners showed a main effect of F2 [$F(1, 69.63)=7.44, p=.008$], with higher F2 associated with thinner perceived body, but no significant interactions were observed.

Higher formants also influenced perceived body build. In Korean listeners, F3 [$F(1, 69.65)=11.98, p=.001$] and F4 [$F(1, 69.78)=7.75, p=.007$] both significantly predicted body build perception of Korean listeners, with increases in these formants generally associated with lighter ratings. No significant interactions involving F3 and F4 were found. For French listeners, F3 had no significant effects, whereas F4 showed a main effect [$F(1, 69.64)=8.97, p=.004$], and a significant speaker language×speaker sex interaction [$F(1, 69.64)=5.56, p=.021$] as well as a three-way interaction with F4 [$F(1, 69.64)=5.56, p=.021$]. Lower F4 was perceived as heavier for Korean male ($\beta=-1.157$) and female ($\beta=-0.522$) speakers and French female speakers ($\beta=-0.609$), but as thinner for French male speakers ($\beta=0.210$). Formant dispersion did not significantly affect perceived build in either group.

f0 showed a significant main effect on perceived build for Korean listeners [$F(1, 69.57)=29.17, p<.001$], with lower f0 associated with heavier ratings, consistent with established pitch-body size associations. Speaker sex [$F(1, 69.57)=6.84, p=.011$] and the interaction between speaker sex and f0 [$F(1, 69.57)=6.29, p=.014$] were also significant, indicating differential weighting between male and female voices. A significant three-way interaction [$F(1, 69.57)=15.49, p<.001$] further indicated lower f0 perceived as heavier for Korean male ($\beta=-0.144$) and female ($\beta=-0.90$) speakers and French female speakers ($\beta=-0.213$), while lower f0 was associated with thinner perception in French male speakers ($\beta=0.034$). For French listeners, f0 had a robust main effect [$F(1, 69.55)=36.00, p<.001$], along with a significant speaker language×speaker sex interaction [$F(1, 69.55)=8.44, p=.005$] and a significant three-way interaction [$F(1, 69.55)=8.46, p=.005$]. The difference in pitch effect between Korean males and females ($\Delta=0.053$,

with female $\beta=-0.075$ and male $\beta=-0.128$) was smaller than that for French speakers ($\Delta=0.134$, with female $\beta=-0.158$ and male $\beta=-0.024$), suggesting that both Korean and French listeners interpret pitch differently depending on the sex of the speaker, reflecting speaker sex-specific expectations in their judgments of body build from speech.

H1*H2* significantly affected body build ratings, with main effects of speaker language [$F(1, 69.69)=4.84, p=.031$ for KR; $F(1, 69.65)=4.42, p=.039$ for FR] and speaker sex [$F(1, 69.69)=11.18, p=.001$ for KR; $F(1, 69.65)=42.78, p<.001$ for FR] in both listener groups.

HNR05 revealed a significant three-way interaction among speaker language, speaker sex, and HNR for Korean listeners [$F(1, 69.74)=5.80, p=.019$], with lower HNR05 perceived as heavier for Korean male ($\beta=-0.033$) and French female ($\beta=-0.023$) speaker, and higher HNR05 perceived as heavier for French male ($\beta=0.015$) and Korean female ($\beta=0.023$) speakers. For French listeners, HNR05 showed a main effect [$F(1, 69.62)=6.29, p=.014$], along with a significant speaker language \times speaker sex interaction [$F(1, 69.62)=7.28, p=.009$] and a significant three-way interaction [$F(1, 69.62)=6.15, p=.016$]. Lower HNR05 was perceived as heavier for Korean male ($\beta=-0.045$) and French male ($\beta=-0.013$) and female ($\beta=-0.031$) speakers, but as lighter for Korean female speakers ($\beta=0.013$).

CPP showed a similar pattern. For Korean listeners, an interaction between speaker language and speaker sex [$F(1, 69.69)=5.22, p=.025$] and a significant three-way interaction [$F(1, 69.70)=4.45, p=.038$] indicated a negative relationship between CPP and perceived build for Korean male ($\beta=-0.068$) and French male ($\beta=-0.004$) and female ($\beta=-0.063$) speakers, and a positive relationship for Korean female speakers ($\beta=0.060$). For French listeners, CPP had a strong main effect [$F(1, 69.60)=8.20, p=.006$], a significant speaker language \times speaker sex interaction [$F(1, 69.60)=10.85, p=.002$], and a significant three-way interaction [$F(1, 69.60)=9.87, p=.002$], showing a negative relationship for Korean male ($\beta=-0.077$) and French male ($\beta=-0.032$) and female ($\beta=-0.122$) speakers, and a positive relationship for Korean female speakers ($\beta=0.041$). As with height perception, higher HNR bands (HNR15, HNR 25, HNR35) did not significantly influence body build perception in either listener group.

4. Discussion and Conclusion

Overall, the results indicate that both speaker-related factors (sex, language) and listener language systematically shape how body size is perceived from speech. For perceived height, male and French speakers tended to be judged as taller, with these differences especially pronounced for French listeners, suggesting that French listeners are more sensitive to height-related vocal cues than Korean listeners. For perceived body build, sex differences were again stronger for French listeners, whereas differences between Korean and French speakers' build are more salient for Korean listeners, implying that listeners' linguistic-cultural background modulates which aspects of speakers' sex and language they rely on when inferring body size.

Regarding the ability to estimate body size from speech, our results suggest that body size is only weakly recoverable from vocal cues. Although French listeners showed moderate ability

to track speaker height, correlations between actual and perceived height and body build were generally low across speaker sex and language. While some studies propose that evolution may have favored men's ability to signal their body size and physical dominance more clearly through the voice (Puts et al., 2006; Sell et al., 2010; Watkins et al., 2010), our data indicate that sex-specific accuracy in body-size estimation from speech is overall weak. Possible explanations include the probabilistic nature of acoustic cues to body size, as well as large within-group variability and overlap in acoustic profiles between taller and shorter or heavier and lighter speakers, which may reduce the strength of the mapping. For example, when compared with actual size, only F1 ($r=-.43, p=.02$) and F3 ($r=.36, p=.04$) showed significant but only moderate correlations with the height and weight of male speakers, respectively, consistent with previous studies (González, 2004, 2006; Pisanski et al., 2014; van Dommelen & Moxness, 1995). In addition, listener judgments may be influenced by cultural stereotypes or task strategies rather than fine-grained acoustic detail, limiting the precision with which body size can be inferred from speech.

The acoustic parameters to which the listeners were most sensitive in judging speakers' body size were the formants. Listeners generally judged that lower formant values were corresponded to taller and heavier speakers, whereas higher formant values were associated with smaller and thinner perceived body size. In the vowel /a/, F1 relates to the length of the pharyngeal cavity, F2 to the oral cavity, and F3 to the front part of the tongue constriction (Fant, 1970: 120-121). This likely explains why listeners perceived speakers as taller when formant values decreased with a longer vocal tract length, despite formants being only weakly related to actual body size (González, 2004, 2006; Pisanski et al., 2014; van Dommelen & Moxness, 1995). This "vocal stereotype" (González, 2006) mirrors patterns in animal behaviors, in which higher formant frequencies are associated with smaller bodies, and lower formants with larger bodies. Consequently, listeners may base judgments on formant values even when these judgments are incorrect. Our results align with van Dommelen (1993), van Dommelen & Moxness (1995), and González (2003), who reported that listeners' judgments are generally consistent but often misguided by incorrect stereotypes.

However, these strategies were not necessarily applied equivalently across sexes and languages. Steeper slopes of F1 were observed for perceived height and build of Korean male speakers compared to females. Only French listeners relied on F2 for estimating body build, whereas only Korean listeners used F3. This suggests that listeners' reliance on specific formant cues is sex-dependent and shaped by their linguistic background, leading French and Korean listeners to adopt partially different cue-weighting strategies. One plausible explanation lies in cross-linguistic differences in the articulation of /a/: French /a/ is fronter compared to Korean /a/, resulting in differently structured vowel spaces in the two languages. Consequently, French listeners, who are accustomed to fine contrasts among front vowels, treat F2 as a salient cue to anterior vocal-tract shape, whereas Korean listeners tend to rely more on global resonance patterns, including higher formants such as F3. These differences in vowel-space organization provide a coherent account for the asymmetric cue-weighting patterns observed.

For f_0 , as expected, lower pitch was perceived as heavier across both listener groups. According to Morton (1977), animals use sound codes whereby larger phonatory and articulatory apparatus produce harsher, deeper sounds to signal larger body dimensions, while smaller apparatus produce tonal, higher sounds, evoking smaller size. Similarly, Ohala (2010) notes that humans perceive low f_0 as produced by a speaker with larger body dimensions, often expressing anger, whereas high f_0 conveys smaller body size or a childish impression. Numerous studies also confirm that, along with formants, f_0 is a primary cue for listeners' impressions of speaker height and weight (González, 2006; Rendall et al., 2007; Smith & Patterson, 2005; Uchida, 2022; van Dommelen & Moxness, 1995), and our findings corroborate these earlier observations.

For the variables related to vocal roughness, higher CPP and HNR05 generally corresponded with smaller perceived body size. Higher HNR05 was associated with shorter and thinner bodies, although Korean female speakers showed the opposite trend, being judged as taller and heavier. Similarly, higher CPP tended to be linked to shorter men but taller women for Korean listeners, and to thinner bodies for both French and Korean listeners, again with Korean female speakers showing a reversed, positive association with perceived build. Vocal roughness, typically produced in aggressive contexts, would likely project an impression of a larger body, while vocal clearness in distress contexts would suggest a smaller body in both animals (Morton, 1977; Ohala, 1980) and humans (Raine et al., 2019). Our Korean and French listeners appeared to employ a similar strategy as observed in these previous studies, except for Korean female speakers. This exception may reflect cultural and perceptual expectations: within Korean culture, clearer voices in women, being more resonant and stable, may convey health, vigor, and physical presence, leading listeners to associate them with taller and heavier bodies. Thus, voice quality cues appear to interact with speaker sex and cultural norms in shaping body size perception. Further perceptual experiments will be necessary to determine what kinds of impressions or social images listeners – both Korean and French – associate with the clearer voice quality observed specifically in Korean female speakers.

Finally, our listeners did not seem sensitive to variables related to vocal breathiness. This contrasts with the findings of Xu et al. (2013), who showed that breathy voice is one of the acoustic characteristics associated with a smaller body. Breathiness is certainly more common in women (Hanson & Chuang, 1999; Iseli et al., 2007; Södersten & Lindestad, 1990), whose vocal folds are generally thinner and smaller, and who generally have smaller body size than men. However, previous studies have reported that voice quality also varies depending on a speaker's language or cultural background (Pépiot & Arnold, 2020; Šebesta et al., 2017), which may explain why breathiness was not a reliable cue for estimating body size from speech.

In conclusion, the present study has revealed that body size is only weakly encoded in speech and that any cues listeners use are probabilistic and strongly shaped by language background and gendered expectations. While certain acoustic parameters, such as f_0 , formant structure, and harmonic energy, systematically influenced judgments, they did not support reliably accurate estimates of height or build. Moreover, the direction and strength of these effects varied across French and Korean listeners as well as male and female speakers. These findings suggest that vocal

impressions of body size reflect a complex interplay between modest anatomical correlates and culturally learned stereotypes rather than a straightforward readout of physical dominance from speech. Recognizing this complexity is important not only for theoretical models of speech perception but also for applied contexts. Misinterpretations arising from culturally shaped perceptual biases may influence voice-based eyewitness descriptions in forensic settings, and biometric systems may benefit from incorporating probabilistic and culturally variable cues. Thus, a better understanding of how listeners infer body size from speech can improve both theoretical accounts and practical applications involving human speech perception.

5. Limitations of the Study

Future research should extend the analysis to a wider range of speech segments. The present study examined only the vowel /a/, chosen because it yields reliable acoustic correlates related to spectral shape and high first formant values. However, voice and speech characteristics observed in a single vowel may not generalize to other speech segments (Choi & Choi, 2016; Henton & Bladon, 1985; Iseli et al., 2007). In addition, increasing the number of participants and examining additional listener populations would be beneficial, as would further investigation of cross-cultural variability in perceptual strategies. Including various segments and broader participant groups in future work will enable a more comprehensive understanding of how body dimensions shape the acoustic manifestations of speech.

Acknowledgement

This work was supported by the Global Korea Scholarship (National Institute for International Education, NIIED) and the Pony Chung Humanities Scholarship (PONY CHUNG Foundation). I am sincerely grateful to my advisors, Cécile Fougeron and Nicolas Audibert, whose invaluable guidance and insightful comments throughout my doctoral studies greatly contributed to the present research.

References

- Altman, D. G. (1991). *Practical statistics for medical research* (reprint 1999). Boca Raton, FL: CRC Press.
- Ambreen, S., Bashir, N., Tarar, S. A., & Kausar, R. (2019). Acoustic analysis of normal voice patterns in Pakistani adults. *Journal of Voice*, 33(1), 124.e49-124.e58.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Boč, L. J., Contini, M., & Rakotofringa, H. (1975). Etude statistique de la fréquence laryngienne. *Phonetica*, 32(1), 1-23.
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (version 6.2.07) [computer program]. Retrieved from <http://www.praat.org/>
- Cartei, V., Bond, R., & Reby, D. (2014). What makes a voice masculine: physiological and acoustical correlates of women's ratings of men's vocal masculinity. *Hormones and Behavior*, 66(4), 569-576.

- Charlton, B. D., Taylor, A. M., & Reby, D. (2013). Are men better than women at acoustic size judgements? *Biology Letters*, 9(4), 20130270.
- Choi, S. H., & Choi, C. H. (2016). The effect of gender and speech task on cepstral- and spectral-measures of Korean normal speakers. *Audiology and Speech Research*, 12(3), 157-163.
- de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, 36(2), 254-266.
- Dusan, S. (2005, September). Estimation of speaker's height and vocal tract length from speech signal. *Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH 2005)* (pp. 1989-1992). Lisbon, Portugal.
- Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. *Biological Psychology*, 72(2), 160-163.
- Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transactions Laboratory Quarterly Progress and Status Report*, 7(4), 22-30.
- Fant, G. (1970). *Acoustic theory of speech production* (2nd ed.). Berlin, Germany: De Gruyter Mouton.
- Filho, J. A. X., Christiano, E., de Melo, M., de Giacomo Carneiro, C., Tsuji, D. H., & Sennes, L. U. (2005). Length of the human vocal folds: Proposal of mathematical equations as a function of gender and body height. *Annals of Otolaryngology Rhinology Laryngology*, 114(5), 390-392.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 102(2), 1213-1222.
- Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3), 330-338.
- González, J. (2003). Estimation of speakers' weight and height from speech: A re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills*, 96(1), 297-304.
- González, J. (2004). Formant frequencies and body size of speaker: A weak relationship in adult humans. *Journal of Phonetics*, 32(2), 277-287.
- González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Development in Applied Physics*, 9, 1-15.
- Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America*, 101(1), 466-481.
- Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America*, 106(2), 1064-1077.
- Hatano, H., Kitamura, T., Takemoto, H., Mokhtari, P., Honda, K., & Masaki, S. (2012). Correlation between vocal tract length, body height, formant frequencies, and pitch frequency for the five Japanese vowels uttered by fifteen male speakers. *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH 2012)* (pp. 402-405). Portland, OR, USA.
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769-778.
- Hollien, H. (1962). Vocal fold thickness and fundamental frequency of phonation. *Journal of Speech and Hearing Research*, 5(3), 237-243.
- Iseli, M., Shue, Y. L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, 121(4), 2283-2295.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485-499.
- Jung, J., & Forbes, G. B. (2007). Body dissatisfaction and disordered eating among college women in China, South Korea, and the United States: Contrasting predictions from sociocultural and feminist theories. *Psychology of Women Quarterly*, 31(4), 381-393.
- Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 27(3-4), 187-207.
- Maezono, J., Hamada, S., Sillanmäki, L., Kaneko, H., Ogura, M., Lempinen, L., & Sourander, A. (2018). Cross-cultural, populationbased study on adolescent body image and eating distress in Japan and Finland. *Scandinavian Journal of Psychology*, 60(1), 67-76.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855-869.
- Ohala, J. J. (1980). The acoustic origin of the smile. *The Journal of the Acoustical Society of America*, 68(S1), S33.
- Ohala, J. J. (2010). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge, UK: Cambridge University Press.
- Pépiot, E., & Arnold, A. (2020). Cross-gender differences in English/French bilingual speakers: A multiparametric study. *Perceptual and Motor Skills*, 128(1), 153-177.
- Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., Fink, B., ... Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour*, 95, 89-99.
- 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.
- R Core Team. (2021). R: A language and environment for statistical computing (version 4.1.2) [computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Raine, J., Pisanski, K., Bond, R., Simmer, J., & Reby, D. (2019). Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLOS ONE*, 14(3), e0213034.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T.,

- & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, 272(1566), 941-947.
- Rendall, D., Owren, M., Weerts, E., & Hienz, R. D. (2004). Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners. *The Journal of the Acoustical Society of America*, 115(1), 411-421.
- Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1208-1219.
- Šebesta, P., Kleisner, K., Tureček, P., Kočnar, T., Akoko, R. M., Třebický, V., & Havlíček, J. (2017). Voices of Africa: Acoustic predictors of human male vocal attractiveness. *Animal Behaviour*, 127, 205-211.
- Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., Krauss, A., & Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, 277, 3509-3518.
- Shue, Y. L., Keating, P., Vicenik, C., & Yu, K. (2011, August). VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)* (pp. 1846-1849). Hong Kong.
- Sjölander, K. (2004). Snack sound toolkit. KTH Stockholm, Sweden. Retrieved from <http://www.speech.kth.se/snack>
- Södersten, M., & Lindestad, P. Å. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research*, 33(3), 601-611.
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096-1104.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.
- Trautmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, 88(1), 97-100.
- Uchida, T. (2022). Voice pitch illusion and perception of speaker's body size: Relationship with the spectral tilt in speech sound. *Acoustical Science and Technology*, 43(1), 73-76.
- van Dommelen, W. A. (1993). Speaker height and weight identification: A re-evaluation of some old data. *Journal of Phonetics*, 21(3), 337-341.
- 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.
- Watkins, C. D., Fraccaro, P. J., Smith, F. G., Vukovic, J., Feinberg, D. R., DeBruine, L. M., & Jones, B. C. (2010). Taller men are less sensitive to cues of dominance in other men. *Behavioral Ecology*, 21(5), 943-947.

Multilingualism and Multiculturalism, Konkuk University
 120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Korea
 Tel: +82-2-446-6914
 Email: dayeon.yoon25@gmail.com
 Areas of interest: Phonetics, Sociophonetics, Sound change, Emotional speech

• **Dayeon Yoon**, corresponding author
 Postdoctoral researcher, Konkuk Research Institute for