Gender attitudes affect the strength of the Frequency Code

Sasha Calhoun, Paul Warren, Jemima Agnew, Joy Mills

Te Herenga Waka - Victoria University of Wellington (THW-VUW)
firstname.lastname@vuw.ac.nz

Abstract

Following the Frequency Code, certain intonational meanings have a biological basis: high vs. low pitch are physically linked to small vs. large body size and to female vs. male gender (via sexual dimorphism), leading to affective meanings like submissiveness vs. dominance. While such associations appear widespread, the code assumes culture- and individual-specific ideological links, e.g., between submissiveness and femininity. We present Implicit Association Test experiments measuring associations between voice pitch and body size/gender. All participants showed these associations, however, their strength varied according to listeners' genders and gender beliefs. We discuss implications for theories of pitch iconicity.

Index Terms: pitch perception, Frequency Code, sociophonetics, sound symbolism, iconicity, gender and language

1. Introduction

Since pitch is one of the few language features we share with other mammals and birds, it has been argued that an important source of intonational meaning is biological. In particular, the *Frequency Code* links high and low f0 with body size and sex, based on animal behaviour, and in turn to affective meanings such as submissiveness versus dominance [1, 2, 3]. Here we probe the Frequency Code from a sociophonetic perspective, where such meanings can be taken to be socially constructed. That is, they can stem from speakers' and listeners' experiences and beliefs in relation to users of different linguistic features, such as an ideological link between masculinity and dominance, which influences linguistic meanings related to low pitch (see [4, 5, 6]). Our proposal seeks to reconcile these approaches: the availability of iconic pitch associations depends not only on physical associations, but also on listeners' beliefs, particularly in relation to gender. We investigate this using Implicit Association Tests (IATs) [7], looking at associations between voice pitch and each of body size and gender, and how these vary by the gender and gender beliefs of the listener.

Biological codes have been proposed to "explain what is universal about the interpretation of pitch variation" ([3], p. 74), linking physiological properties associated with pitch production with informational and affective interpretations of pitch features in language. For example, larger animals have larger vocal apparatus, producing vocalisations with lower f0. Morton [1] observed mammals and birds tended to use lower f0 when acting aggressively, and higher f0 when acting submissively, regardless of their actual size, i.e. f0 conveyed *apparent* size. Following this, Ohala ([2], p. 327) proposes the Frequency Code: affective meanings such as "deference, politeness, submission, lack of confidence, are signaled by high and/or rising F0 whereas assertiveness, authority, aggression, confidence, threat are conveyed by low/or falling F0". Ohala [2] extended this to human sex differences, because of sexual dimorphism: since males tend to be larger than females, the relationship between low f0 and dominance is particularly associated with males, and high f0 and submissiveness with females. Importantly, while studies show that f0 may not be a reliable cue to human body size (e.g. [8, 9]), listeners strongly associate low pitch with larger body size [10]. Further, while gender and its performance are complex, and the links to biological sex indirect, listeners are very sensitive to vocal cues to gender in speech, including f0, from a very young age (e.g. [11, 12]). The Frequency Code has considerable support in perception studies, with high pitch associated with uncertainty, weakness, dependency and submissiveness across listeners of different cultural and linguistic backgrounds (e.g. [13, 14, 15, 16]). In particular, the link between low f0 and dominance is widely accepted in social psychology (e.g. [10, 17, 18]).

Biological code theory proposes that meanings of pitch features can stem from sound symbolism or iconicity, which may then be phonologized within a language [19, 2]. However, despite cross-linguistic commonalities in these seemingly natural links, the Frequency Code assumes culturally-specific ideological links, e.g. between submissiveness and femininity, which have largely not been accounted for in prior research. This runs contrary to decades of sociolinguistic and sociophonetic research showing that links between linguistic features and particular groups and affects are socially constructed (e.g. see [4, 5]). Nonetheless, recent work suggests how these approaches can be reconciled. For example, Eckert ([4], p. 754) discusses how iconic associations are "products of cultural constructions of resemblance to things in the natural world". Thus, of the multiple potential resemblances pitch (and other linguistic features) could have, those that align with existing listener 'world views' and the affordances of the context, are most likely to become established social meanings [5, 4, 20, 6]. For example, Walker et al. [20] note that the link between high pitch features and politeness depends on the link between high f0 and weakness or submissiveness, i.e. the speaker signals politeness by appearing weaker and non-threatening to their interlocutor. In a perception study on Korean and English, they show Korean listeners have similar pitch associations for submissiveness to English listeners, but not politeness, showing these are not linked in Korean; rather, politeness is usually signalled by low pitch features (see also [6]). Such an approach could also explain other apparently contradictory findings in Frequency Code research, e.g. dominance signalled by high pitch for female speakers [21].

Our research explores the idea that pitch iconicity can provide a shared 'extra-linguistic' basis for linguistic meanings. However, we propose that in language cognition the availability of different physical pitch associations, and thus different derived meanings, varies according to listeners' experiences and beliefs, and the context. In relation to the Frequency Code, the most important of these relate to listeners' gender and gender beliefs, as the core affective associations of the Frequency Code (e.g. low pitch and dominance) align closely with normative gender. As a starting point, we look at associations of voice pitch with the core physical associations which are said to underlie the Frequency Code, i.e. body size and binary gender. We investigate whether the strengths of these associations differ between individuals based on their gender and gender beliefs. We predict that associations will be stronger for listeners of male gender and those with more normative, or less egalitarian, gender beliefs.

We investigate physical associations of voice pitch using IATs [7], a well-established task that measures implicit association strength between paired concepts and attributes. We measure associations of small/large body size or female/male with high/low pitch, using participants' speed at categorising stimuli in each category. Most previous studies reviewed above have used rating tasks. The IAT is arguably better to tap into subconscious associations which listeners may not have metacognition of [22]. The IAT has been shown to effectively measure subconscious associations of other linguistic features and, importantly, how these vary given social factors (e.g. [23, 24]).

2. Method

IAT experiments were run to find out firstly if participants showed the implicit associations between voice pitch and each of gender and body size predicted by the Frequency Code, and secondly, whether the strength of these associations were affected by the participants' gender and their experiences and beliefs around gender, as measured in gender attitude surveys.

2.1. Participants

Data is reported from 239 participants, recruited on Prolific (www.prolific.co), evenly distributed across experiment versions by participant gender. There were 113 female, 120 male and 6 non-binary/other gender, all with English as their first language and living in New Zealand, Australia or the United Kingdom. For the statistical modelling, female and non-binary/other genders were grouped (labelled 'non-male'). Median age was 33 years (range 18-69 years). Participants received Prolific credits for participation. This study was approved by the Human Ethics Committee at THW-VUW (#29710).

2.2. Materials

For the IAT experiments we created two sets of concept stimuli (gender and size) and one set of attribute stimuli (pitch). For gender classification, twelve names were chosen with strong associations to either female (*Anna, Claire, Jane, Julie, Rachel, Sarah*) or male (*Andrew, Daniel, David, James, Mark, Michael*) gender. Names were chosen using the New Zealand Department of Internal Affairs register of baby names (www.dia.govt.nz/diawebsite.nsf/wpg_URL/Services-Births-Deaths-and-Marriages-Most-Popular-Male-and-Female-First-Names). We selected names that were among the most frequent across the period 1960-2000, so they would be familiar to participants.

For size, twelve animals were chosen which would usually be classified as either small (*frog, mouse, blackbird, guinea pig, squirrel, rabbit*) or large (*elephant, giraffe, bear, crocodile, panda, camel*) in size. Black-and-white photographs of the animals were taken from the Animal Images database [25]. Pictures were all 300*225 pixels. The small and large sets were matched for familiarity and valence.

The pitch stimuli involved nonwords similar in phonetic make-up to discourse markers intended for later studies. The final selection for the current study was *ernerm* and *yerwer* (similar to *I mean* and *really*). Tokens of each nonword were recorded by three female and three male New Zealand English speakers in their thirties. The tokens were time- and amplitude-normalised, and then pitch-normalised using a purpose-built Matlab script to mean values of 195 Hz for females and 110 Hz for males (values based on means of word tokens in the New Zealand Spoken English Database [26]), with a declination slope of 0.7 ERB. High- and low-pitched versions were then created with all pitch values raised by 1.7 ERB or lowered by 0.7 ERB (values based on pre-testing by the authors).

All three sets of stimuli were chosen from larger sets after norming studies. The first (N=32) showed female names were rated 1.13 and male 6.78 on a scale 1 (definitely female) to 7 (definitely male). Participants' accuracy at categorising pitch stimuli (as high or low) was 97.9%. The second study (N=32) showed that small animal pictures were rated 1.41 and large 6.48 on a scale 1 (definitely small) to 7 (definitely large).

A questionnaire targeted participants' basic demographic information and language background. To gauge participants' gender attitudes and beliefs, questions were selected from established gender surveys. Participants had to indicate their level of agreement with statements using a Likert scale from 1 (strongly disagree) to 7 (strongly agree). We constructed five measures of gender attitudes: we selected the subsets of the Ambivalent Sexism Inventory [27] used in the New Zealand Attitudes and Values Survey (NZAVS) [28], i.e. five questions relating to Benevolent Sexism, e.g. "Women, compared to men, tend to have greater moral sensibility", and five relating to Hostile Sexism, e.g. "Women seek to gain power by getting control over men". Male Norms were measured with the five-item Male Norms Inventory [29], e.g. "Boys should prefer to play with trucks rather than dolls". Transgender/Non-Binary attitudes were measured via two statements adapted from the New Zealand Gender Attitudes Survey [30], e.g. "I would be comfortable with a transgender or non-binary person as a colleague", and finally, Social Dominance Orientation was measured using a six-item subset from the NZAVS inventory [31], e.g. "Inferior groups should stay in their place". This measure was included as voice pitch has been shown to relate to perceptions of social as well as physical dominance [10].

Table 1: Example sequence of blocks for IAT Experiments. Shows Consistent-First order

Block	No. of trials	Type	Items on left-key response	Items on right-key response
1	24	Practice	Male names	Female names
2	24	Practice	Low pitch	High pitch
3	4	Practice	Male + Low	Female + High
4	48	Consistent	Male + Low	Female + High
5	36	Practice	Female names	Male names
6	4	Practice	Female + Low	Male + High
7	48	Inconsistent	Female + Low	Male + High

2.3. Design and procedure

The IAT experiments were constructed and run in PsyToolkit, version 3.4 [32, 33], following a standard IAT design, as in [7, 34], see Table 1. In the practice blocks 1-2, participants learn to classify stimuli as being from each pair of a concept and attribute, e.g. male/female and low/high pitch, linked to the left ('E') or right ('I') response key. In the third practice block and the following test block, concept and attribute stimuli are combined. In this example, these blocks are 'consistent', i.e. the expected pairing of concept and attribute, e.g. male and low pitch, are on the same response key. The response key for the concept is then reversed in Block 5, and then the reverse, 'inconsistent', combination of concept and attribute is tested (Blocks 6 and 7). If a participant has the expected implicit association between concept and attribute, they should be faster and more accurate at classifying stimuli in 'consistent' (4) than 'inconsistent' (7) blocks. Two concepts were included in each experiment version. The second concept followed the same design, except that the attribute practice block (2) is omitted as this has not changed. In blocks 1, 2, 4 and 7, each stimulus was repeated twice. In block 5 (reverse concept practice block), stimuli were repeated three times as [35] found this reduces order effects (see below). Four practice items were used for the combined blocks (3, 6). In all practice blocks stimulus order was randomised by participant. In test blocks, participants received one of nine pseudo-randomly ordered lists with no more than three responses with the same key response, or three concept or attribute stimuli, in a row.

Twenty-four versions of the IAT were constructed, involving three concepts, Gender, Size, and Effort (not reported here). All six possible pairings of the three concepts were used. Each version used only one voice gender: either the three male or the three female voices. Finally, both possible orderings of the consistent-inconsistent blocks were used, as it is well-established that the implicit association effect may be substantially smaller in inconsistent-consistent order (e.g. see [34, 35]). Table 1 shows consistent-inconsistent order, in inconsistent-consistent order Blocks 1, 3, 4 were reversed with 5-7.

Participants completed the experiment online using PsyToolkit, using a desktop or laptop in a quiet room with headphones. They completed the IAT first. The instructions and layout for the IAT closely followed those on the Project Implicit website (implicit.harvard.edu/implicit/takeatest.html). Participants first received instructions on the task and saw/heard all stimuli. Instructions were repeated before each block, although there was no break between the combined practice and test blocks. Each response had a timeout of 3s. If a participant responded incorrectly, they saw a red cross on the screen and then had to press the correct response. Participants then completed the demographic questions, followed by the gender attitude questions, presented in pseudo-random order. The experiment took approximately 25 minutes.

2.4. Analysis

The size of implicit associations for each participant for each concept-attribute pairing was gauged by D-scores, calculated with a script modified from [36] in R, based on [34]. Responses below 400ms were removed, and RTs for incorrect responses replaced with the participant's mean RT plus a 600ms 'penalty'. A D-score is the difference between a participant's mean response time in the inconsistent and consistent blocks (i.e. 7-4 in Table 1), divided by their SD in these blocks.

Linear regression models were built in R, with D-score as the dependent variable. Initial models included a five-way interaction of Concept (Gender or Size), Voice Gender (Male or Female voices in the experiment), Order (Consistent first or Inconsistent first), Half (first or second half of the experiment) and participant Gender (see further below). Elimination of non-significant effects used the *stepAIC* function in the MASS package with further manual elimination using *anova*, starting with higher-order interactions, until remaining factors and interactions were significant. Model estimates were extracted using *effects* and plotted with *ggplot2*.

Participant gender correlated with each of the gender attitude measures, which were also correlated with each other. Since correlated predictors can make regression models unreliable, we built separate models for participant gender and each of the gender attitude measures. Scores for each measure (Benevolent Sexism, Hostile Sexism, Male Norms, Transgender/Non-Binary Attitudes, Social Dominance Orientation) were derived by averaging ratings for the statements in each set, after reversing ratings for statements where a high rating indicated a more egalitarian attitude, so that higher indicated less egalitarian views for all statements. Median scores were: Benevolent Sexism 3.6, Hostile Sexism 2.6, Male Norms 2.2, Transgender/Non-Binary 1, Social Dominance 1.8.

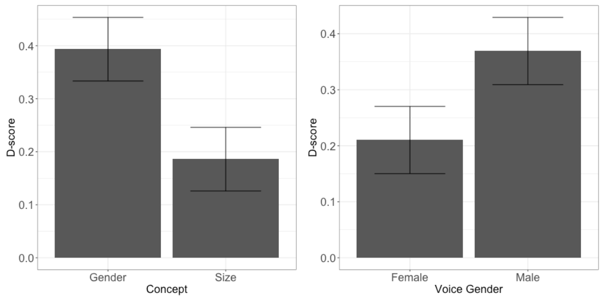

Figure 1: *Fitted D-scores by Concept (left) and Voice Gender (right). Error bars show standard error of the means.*

3. Results

3.1. Participant gender

The final regression model including Participant Gender showed simple effects of Concept (F(1, 310) = 23.15, p<0.001) and Voice Gender (F(1, 310) = 12.25, p<0.001), and interactions of Participant Gender*Order (F(1, 310) = 9.15, p=0.003) and Half*Order (F(1, 310) = 5.71, p=0.017). As can be seen in Figure 1, D-scores were higher, indicating a stronger effect, for the Gender than Size concept (t=-4.81, p<0.001), and for Male than Female Voices (t=6.35, p<0.001).

The Participant Gender*Order interaction can be seen in Figure 2 (left). Comparisons using *emmeans* (fdr method) showed higher D-scores in Consistent-first order than Inconsistent-first for Male (t=5.96, p<0.001) but not Non-male participants (p=0.11). Further, Male participants had higher D-scores than Non-male in Consistent-first order (t=-3.81, p<0.001) but not Inconsistent-first order (p=0.69). As noted above, D-scores are often found to be higher in Consistent-first than Inconsistent-first order, although the interaction with gender was not predicted (see further below). For the Half*Order interaction, comparisons showed D-scores were higher in Consistent-first order than Inconsistent-first in both the first (t=5.5, p<0.001) and second half (t=2.14, p=0.039) of the experiment. D-scores were also higher in the first half than second in Consistent-first order (t=2.43, p=0.023) but not Inconsistent-first (p=0.36). This is consistent with a learning effect, where D-scores were smaller in the second half of the experiment as participants got used to the task, although this could only be seen in Consistent-first order.

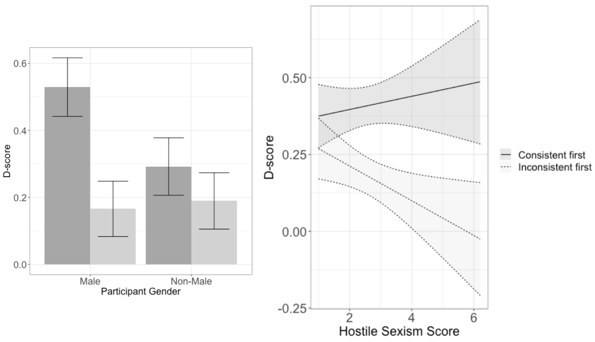

Figure 2: *Fitted D-scores by Participant Gender and Order (left) and Hostile Sexism score and Order (right). Error bars show standard error of the means.*

3.2. Gender attitudes measures

As explained in section 2.4, separate models were built for each of the five gender attitude measures. The models for Benevolent Sexism, Hostile Sexism, Male Norms and Social Dominance Orientation returned simple effects of Concept, Voice Gender and an interaction of Half*Order in line with the Participant Gender model above (details not reported for space reasons). The Hostile Sexism model also included an interaction of Hostile Sexism and Order (F(1, 310) = 4.59, p=0.033). As seen in Figure 2 (right), participants with low Hostile Sexism scores show a small difference between D-scores in Consistent-first and Inconsistent-first orders. As Hostile Sexism scores increase, D-scores in Consistent-first order rise while those in Inconsistent-first order fall. This was not predicted, but is in line with the Participant Gender*Order interaction above.

For Social Dominance Orientation, the interaction with Order was marginal (F(1, 310) = 3.6, p=0.059) and similar to that for Hostile Sexism. The Benevolent Sexism and Male Norms models returned only simple effects of the measure (F(1, 311) = 12.89, p<0.001 and F(1, 311) = 7.79, p=0.006, respectively). As measure scores rise, D-scores fall. The lack of interactions with Order is a likely result of loss of power, since each gender measure was modelled separately. As seen in Figure 2, the reduction of D-scores with increasing Hostile Sexism scores in Inconsistent-first order was larger than the rise in Consistent-first order, so this amounts to an overall reduction. The Transgender/Non-Binary Attitudes model returned a Half*Order interaction and a four-way interaction of Concept*Voice Gender*Transgender/Non-Binary Attitudes*Order (F(1, 300) = 4.01, p=0.046). While there is not space to explore the four-way interaction, broadly the same pattern of Gender Measure*Order was found as in Figure 2, but not for all Concept*Voice Gender combinations.

4. Discussion

This study used IATs to explore implicit associations of voice pitch predicted by the Frequency Code. As predicted, low pitch is implicitly associated with males and large body size, and high pitch with females and small body size. Further findings were generally consistent with our proposal that these associations would be affected by participants' experiences and beliefs, particularly in relation to gender. The implicit association with pitch was stronger for gender than body size, and for male voices than female. Arguably, cultural stereotypes informing these associations are stronger for gender and male voices than size and female voices. For participant gender and most gender attitude measures, we found an unexpected interaction with IAT block order. When the blocks with the 'consistent' associations (e.g. low pitch and male) were presented first, we found the predicted effect: implicit associations were stronger for males and those with less egalitarian gender attitudes. However, in inconsistent-first order, a different pattern emerged: non-males and those with more egalitarian gender attitudes showed no or a very small difference to consistent-first order, while males and those with less egalitarian attitudes showed a much smaller effect. We discuss each of these findings further below.

Implicit associations of body size and gender with voice pitch are based on physical associations. However the association with body size is arguably primary, existing in animal communication systems even when species do not show sexual dimorphism [1]. Nonetheless, we found stronger implicit associations for gender than body size. We submit this fits with our general proposal: cultural stereotypes relating to gender are stronger and more visible than those involving body size (in connection with voice pitch). Similarly, implicit associations for male voices were stronger than female. We suggest this is because, as the historically dominant and privileged gender, ideologies relating to the Frequency Code are more entrenched for males, so the physical associations on which they are based are more salient. For females, given the impact of feminism, these are more challenged, and the iconic links are disrupted and weaker. This matches previous findings that Frequency Code associations do not always hold for female speakers [9, 21, 18].

As predicted, we found stronger implicit associations between voice pitch and body size/gender for males and those with less egalitarian gender beliefs. We suggest this is because the iconic pitch associations based on the Frequency Code largely align with normative gender ideology, e.g. associations between masculinity and dominance. Therefore physical associations on which the iconic associations rest will be stronger for individuals with normative beliefs about gender. However, unexpectedly, we found this effect only when the IAT task was presented in 'consistent-first' order, i.e. with blocks showing the expected voice pitch associations first. We believe that this result does make sense given our proposal. For listeners with likely stronger a priori biases, i.e., males and those with less egalitarian beliefs, experiencing the consistent associations first (male names with low pitch, female with high) reinforces or at least does not contradict these connections. If, however, they experience the inconsistent associations in the first block, this contradicts their a priori bias and leads to a reduction in its effect on the task. These effects persists through the experiment.

Our results show IATs, combined with measures of individual differences, are a promising way to investigate how iconic associations of voice pitch may form part of language cognition and linguistic meaning. In future work, we plan to investigate implicit associations of linguistic features involving pitch, such as uptalk and creaky voice, which have been argued to have Frequency Code-related meanings (see [37]). We will also explore how to more effectively quantify the effect of gender and gender attitude measures in our modelling, e.g. principal components analysis, rather than treating these separately.

We believe that our proposal has the potential to contribute significantly to the understanding of intonational meaning and how this is affected by biological, social, and individual (experiential) factors, as well as to longstanding debates on the (non-)arbitrariness of linguistic meaning. By quantifying how voice pitch can be simultaneously 'natural' and 'social', we offer a new approach to investigating iconicity in speech.

5. Acknowledgements

This work was funded by a Faculty Strategic Research Grant from THW-VUW. Thanks to Jen Hay, Abby Walker, Joe Bulbulia and Gina Grimshaw for advice and to the participants.

6. References

[1] E. S. Morton, "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *The American Naturalist*, vol. 111, no. 981, pp. 855–869, 1977.

[2] J. J. Ohala, "The frequency code underlies the sound-symbolic use of voice pitch," in *Sound Symbolism*, L. Hinton, J. Nichols, and J. J. Ohala, Eds. Cambridge, UK: Cambridge University Press, 1994, pp. 325–347.

[3] C. Gussenhoven, "Paralinguistics: Three biological codes," in *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004, pp. 71–96.

[4] P. Eckert, "The limits of meaning: Social indexicality, variation, and the cline of interiority," *Language*, vol. 95, no. 4, pp. 751–776, 2019.

[5] A. D'Onofrio and P. Eckert, "Affect and iconicity in phonological variation," *Language in Society*, vol. 50, no. 1, pp. 29–51, 2021.

[6] B. Winter, G. E. Oh, I. Hbscher, K. Idemaru, L. Brown, P. Prieto, and S. Grawunder, "Rethinking the frequency code: a meta-analytic review of the role of acoustic body size in communicative phenomena," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1840, p. 20200400, 2021.

[7] A. G. Greenwald, D. E. McGhee, and J. L. K. Schwartz, "Measuring individual differences in implicit cognition: The Implicit Association Test," *Journal of Personality and Social Psychology*, vol. 74, no. 6, pp. 1464–1480, 1998.

[8] D. Rendall, S. Kollias, C. Ney, and P. Lloyd, "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," *The Journal of the Acoustical Society of America*, vol. 117, no. 2, pp. 944–955, 2005.

[9] W. A. van Dommelen and B. H. Moxness, "Acoustic parameters in speaker height and weight identification: Sex-specific behaviour," *Language and Speech*, vol. 38, no. 3, pp. 267–287, Jul. 1995.

[10] M. Armstrong, A. Lee, and D. Feinberg, "A house of cards: bias in perception of body size mediates the relationship between voice pitch and perceptions of dominance," *Animal Behaviour*, vol. 179, pp. 43–51, 2019.

[11] S. B. Most, A. V. Sorber, and J. G. Cunningham, "Auditory Stroop reveals implicit gender associations in adults and children," *Journal of Experimental Social Psychology*, vol. 43, no. 2, pp. 287–294, 2007.

[12] L. Nagels, E. Gaudrain, D. Vickers, P. Hendriks, and D. Bakent, "Development of voice perception is dissociated across gender cues in school-age children," *Scientific Reports*, vol. 10, no. 1, p. 5074, 2020.

[13] B. Borkowska and B. Pawlowski, "Female voice frequency in the context of dominance and attractiveness perception," *Animal Behaviour*, vol. 82, no. 1, pp. 55–59, 2011.

[14] W. Gu, P. Tang, K. Hirose, and V. Auberg, "Crosslinguistic comparison on the perception of Mandarin attitudinal speech," in *Interspeech 2015*. ISCA, 2015, pp. 1334–1338.

[15] C. Gussenhoven and A. Chen, "Universal and language-specific effects in the perception of question intonation," in *Proc. of ICSLP 2000*, Beijing, 2000, pp. 91–94.

[16] R. van Bezooijen, "Sociocultural aspects of pitch differences between Japanese and Dutch women," *Language and Speech*, vol. 38, pp. 253–265, 1995.

[17] K. Pisanski and G. Bryant, "The evolution of voice perception," in *The Oxford Handbook of Voice Studies*, N. S. Eidsheim and K. Meizel, Eds. Oxford: Oxford University Press, 2019.

[18] M. S. Tsantani, P. Belin, H. M. Paterson, and P. McAleer, "Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices," *Perception*, vol. 45, no. 8, pp. 946–963, 2016.

[19] M. Dingemanse, D. E. Blasi, G. Lupyan, M. H. Christiansen, and P. Monaghan, "Arbitrariness, iconicity, and systematicity in language," *Trends in Cognitive Sciences*, vol. 19, no. 10, pp. 603–615, 2015.

[20] A. Walker, J. J. Holliday, M. Jung, and E. Cho, "A closer look at the sound of politeness in Korean," in *New Ways of Analyzing Variation Asia Pacific (NWAV-AP) 6*, Singapore, 2021.

[21] P. McAleer, A. Todorov, and P. Belin, "How do you say 'Hello'? Personality impressions from brief novel voices," *PLOS ONE*, vol. 9, no. 3, p. e90779, 2014.

[22] B. Kurdi, K. A. Ratliff, and W. A. Cunningham, "Can the Implicit Association Test serve as a valid measure of automatic cognition? A response to Schimmack (2021)," *Perspectives on Psychological Science*, vol. 16, no. 2, pp. 422–434, Mar. 2021.

[23] K. Campbell-Kibler, "The Implicit Association Test and sociolinguistic meaning," *Lingua*, vol. 122, no. 7, pp. 753–763, 2012.

[24] P. lvarez Mosquera, "The use of the Implicit Association Test (IAT) for sociolinguistic purposes in South Africa," *Language Matters*, vol. 48, no. 2, pp. 69–90, 2017.

[25] C. Possidnio, J. Graa, J. Piazza, and M. Prada, "Animal Images Database: Validation of 120 images for human-animal studies," *Animals*, vol. 9, no. 8, p. 475, 2019.

[26] P. Warren, "NZSED: building and using a speech database for New Zealand English," *New Zealand English Journal*, vol. 16, pp. 53–58, 2002.

[27] P. Glick and S. T. Fiske, "The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism," *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 491–512, 1996.

[28] C. G. Sibley, "Sampling procedure and sample details for the New Zealand Attitudes and Values Study," New Zealand Attitudes and Values Study, Tech. Rep., 2021. [Online]. Available: http://nzavs.auckland.ac.nz

[29] R. C. McDermott, R. F. Levant, J. H. Hammer, N. C. Borgogna, and D. K. McKelvey, "Development and validation of a five-item Male Role Norms Inventory using bifactor modeling," *Psychology of Men & Masculinities*, vol. 20, no. 4, pp. 467–477, 2019.

[30] National Council of Women of NZ, "New Zealand Gender Attitudes Survey," 2019. [Online]. Available: https://genderequal.nz/ga-survey/

[31] J. Sidanius and F. Pratto, *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge: Cambridge University Press, 1999.

[32] G. Stoet, "PsyToolkit: A software package for programming psychological experiments using Linux," *Behavior Research Methods*, vol. 42, no. 4, pp. 1096–104, Nov. 2010.

[33] ——, "PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[34] A. G. Greenwald, B. A. Nosek, and M. R. Banaji, "Understanding and using the Implicit Association Test: I. An improved scoring algorithm." *Journal of Personality and Social Psychology*, vol. 85, no. 2, pp. 197–216, 2003.

[35] B. A. Nosek, A. G. Greenwald, and M. R. Banaji, "Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity," *Personality and Social Psychology Bulletin*, vol. 31, no. 2, pp. 166–180, 2005.

[36] J. Röhner and P. J. Thoss, "A tutorial on how to compute traditional IAT effects with {R}," *The Quantitative Methods for Psychology*, vol. 15, no. 2, pp. 134–147, 2019.

[37] H. White and S. Calhoun, "Mediated iconicity: Effect of age on affective associations of uptalk and creak," in *18th Conference on Laboratory Phonology*, 2022.