Knowledge of a Talker's f0 Affects Subsequent Perception of Voiceless Fricatives

Orhun Uluşahin [1], *Hans Rutger Bosker* [2,1], *James M. McQueen* [2,1], *Antje S. Meyer* [1,2]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
orhun.ulusahin@mpi.nl

Abstract

The human brain deals with the infinite variability of speech through multiple mechanisms. Some of them rely solely on information in the speech input (i.e., signal-driven) whereas some rely on linguistic or real-world knowledge (i.e., knowledge-driven). Many signal-driven perceptual processes rely on the enhancement of acoustic differences between incoming speech sounds, producing contrastive adjustments. For instance, when an ambiguous voiceless fricative is preceded by a high fundamental frequency (f0) sentence, the fricative is perceived as having lower a spectral center of gravity (CoG). However, it is not clear whether knowledge of a talker's typical f0 can lead to similar contrastive effects. This study investigated a possible talker f0 effect on fricative CoG perception. In the exposure phase, two groups of participants (N=16 each) heard the same talker at high or low f0 for 20 minutes. Later, in the test phase, participants rated fixed-f0 /ʔɔk/ tokens as being /sɔk/ (i.e., high CoG) or /ʃɔk/ (i.e., low CoG), where /ʔ/ represents a fricative from a 5-step /s/-/ʃ/ continuum. Surprisingly, the data revealed the opposite of our contrastive hypothesis, whereby hearing high f0 instead biased perception towards high CoG. Thus, we demonstrated that talker f0 information affects fricative CoG perception.

Index Terms: Talker familiarity, fricative perception, spectral contrast effects, speech perception.

1. Introduction

Human speech consists of complex and variable arrangements of sounds, making each utterance unique. Nevertheless, this seemingly infinite variability does not seem to pose a major challenge to effortless speech comprehension in daily life. One of the ways listeners deal with this variability is through rapid perceptual adjustments. Various signal-driven processes constantly re-tune the perceptual system to better distinguish between speech sounds.

Among these, spectral context effects (SCEs) typically rely on the perceptual enhancement of acoustic contrasts to normalize incoming speech [1]. For instance, listeners adjust their perception of a voiceless fricative's spectral center of gravity (CoG) based on the local f0 context. That is, hearing a higher-f0 lead-in sentence biases perception of an ambiguous fricative between /s/ and /ʃ/ towards /ʃ/ (i.e., lower CoG), while a preceding lower-f0 sentence biases perception towards /s/ (i.e., higher CoG) [2]. These contrastive effects (e.g., high-f0 context biasing towards low-CoG perception) typically operate early in perception, normalizing input across linguistic and non-linguistic signals alike [3], [4].

The listener's linguistic and real-world knowledge also plays a role in speech perception. Specifically, knowledge of a particular talker's speech guides the perception of speech from that talker, while also providing overall intelligibility and disambiguation benefits [5], [6], [7]. For instance, when talker identity is reliably cued using different patterns of f0 contours, listeners' perception of vowel f0 is mediated by perceived speaker identity [8]. In fact, similar knowledge-driven effects have been observed in the domain of fricative perception on account of talker features as abstract as perceived sexual orientation of a talker [9].

While a multidisciplinary line of research describes a long list of SCEs among acoustic context effects [1], the role of prior talker-specific knowledge in inducing SCEs is largely unknown, with some reports even indicating that having to rapidly adjust to talker variability may diminish spectral context effects [10]. Evidence from studies investigating talker effects in the temporal domain suggest that talker information may indeed lead to contrastive perceptual effects, but only if the immediate acoustic context is limited in its capacity to reliably normalize incoming speech. For instance, the perception of vowels with ambiguous durations can be contrastively influenced by a listener's knowledge of a talker's typical speech rate (i.e., faster talker = longer perceived vowel). However, these knowledge-driven effects only seem to manifest in certain conditions, typically when local contextual information is minimized (e.g., categorization involves isolated phonemes from a known talker, [11]), when information acquired from a talker is highly consistent [12], or potentially only when the target contrast is present in the exposure phase of a study [13].

These reports of contrastive talker effects in the temporal domain raise the question whether similar contrastive talker effects can be observed in the spectral domain. For instance, fundamental frequency (f0) can be extracted from all voiced segments in an utterance, is consistently used as a cue for differentiating between speakers [14], [15], and is stable within talkers [8] with most talkers staying roughly within three semitones of their mean f0 in speech [16]. This is in sharp contrast to speech rate, which varies extensively within talkers [17]. Furthermore, research has shown that two-alternative forced-choice (2AFC) tasks involving categorical voiceless fricative perception represent a reliable proxy measure of perceptual adjustments to f0 [2]. Therefore, listeners may use highly consistent knowledge about a talker's typical f0 to disambiguate ambiguous speech input from that talker.

In this study, we hypothesized that f0 information acquired through prior talker exposure would lead to a contrastive effect on subsequent fricative CoG perception. Thus, listeners exposed to a high f0 talker in an exposure phase should perceive ambiguous fricatives as having lower CoG in a subsequent phonetic categorization test phase. Conversely, listeners exposed to a low-f0 talker should perceive the same ambiguous fricatives as having higher CoG. To test these hypotheses, we ran an experiment to investigate the role of talker f0 information on the categorization of an artificial

fricative continuum as /s/ or /ʃ/. In an exposure phase, two groups of participants heard 20 minutes of speech from a native speaker of Dutch whose voice was pitch-shifted higher or lower per group. Later, in a 2AFC test phase, they categorized tokens from an artificial fricative continuum between /s/ and /ʃ/. Moreover, the experiment was run both online (Experiment 1) and in the lab (Experiment 2) in order to assess the replicability of a potential talker effect as well as the reliability of the online testing paradigm.

2. Experiment 1 (Online)

2.1. Method

2.1.1. Participants

We recruited 32 female native Dutch speakers for the experiment (age = 19-38, *M* = 25.55 (*SD* = 5.49)). We restricted our sample to female participants to minimize the differences in productive f0 ranges across our speaker and participants. All participants self-reported no auditory or visual (unless corrected) impairments, and no language-related disorders. Participants were recruited online through Prolific (www.prolific.com). The two halves of the sample were assigned at random to high and low f0 groups.

2.1.2. Materials

For the exposure task, we recorded a female native speaker of Dutch who read ten two-minute snippets of text gathered from a variety of media, with a selection of everyday topics (e.g., news reports). We also generated one true-false question per snippet (See supplementary material for snippets and questions) to help assess participants' engagement in the task. The speaker's mean f0 in these recordings was 232 Hz (*SD* = 57 Hz), which aligned with reports of typical female Northern Standard Dutch speakers [18]. These recordings were then pitch-shifted by ±4 semitones using PSOLA in Praat [19] to create the materials for the high and low f0 exposure groups. During this process, the original mean intensity, intensity contour, and f0 contour of all sentences were all retained. To calculate semitones, we used a formula that treats semitones as logarithmic derivations of Hertz units whereby the difference between two frequencies can be obtained by the following:

$$D = 12 * log_2(f_1/f_2) = 12 / log_{10}(2) * log_{10}(f_1/f_2)$$

For the 2AFC test phase, we used a Praat script (See [20] for the original script; See supplementary material for the modified version used in the present study) to synthesize an artificial 8-step fricative continuum between /s/ and /ʃ/, with endpoints modeled on fricatives extracted from the same female native Dutch speaker's pronunciations of the words *sok* /sɔk/ "sock" and *sjok* /ʃɔk/ "to trudge". Two endpoints were synthesized by shaping white noise around five peaks, with four smaller peaks defined in relation to the highest in amplitude and bandwidth. The resulting sounds (i.e., the continuum endpoints) had five peaks and spectral distributions that closely resembled the speaker's fricatives. These endpoints had identical duration, and the rest of the continuum was linearly interpolated from these endpoints. The fricative in the original *sok* token was then replaced with the synthesized continuum, resulting in 8 sound files, one per fricative. Based on pretesting, we used only a 5-step subset of the original continuum (steps 2-6). Thus, no endpoint tokens of either phoneme were present. The resulting *sok-sjok* continuum was zero-shifted (i.e., shifted by 0 Hz) to emulate any artifacts that might have resulted from the pitch shifting process of the exposure material (mean f0 = 230 Hz). It is worth noting that across both experiments, in the post experiment questionnaire, only two out of 62 participants specifically identified pitch-shifting or "auto-tune" artifacts.

2.1.3. Procedure

The experiment was carried out using the Gorilla online testing platform [21]. Participants were first forwarded to the consent form. After providing informed consent, participants proceeded to the headphone check. This headphone check utilized the Huggins pitch phenomenon, which demands pitch detection in noise that is only possible with binaural interaction [22]. The check involved 12 trials in which participants had to pick one of three noise signals as containing a faint tone [23]. All 32 participants passed at least 8 out of these 12 trials.

In the exposure task, participants heard ten 2-minute snippets at either high f0 (i.e., shifted up by 4 semitones) or low f0 (i.e., shifted down by 4 semitones). After each snippet, participants answered a short true-false question to assess their attentiveness. They were informed that they were "expected to answer eight out of ten questions correctly".

Participants then proceeded to the practice trials for the 2AFC categorization task. The practice trials only included four trials: Two trials featuring the first (i.e., the most /s/-like) and the last (i.e., the most /ʃ/-like) steps of the original 8-step *sok-sjok* continuum. These endpoints were used to ensure that participants could understand the task without interference from the fricative ambiguity manipulation. On each trial, participants heard one word while a fixation cross was displayed. After playback, the two words and corresponding answer keys were displayed. Participants had to use a keyboard press (i.e., keys 'A' and 'L') to indicate which of the two words (i.e., *sok* or *sjok*) they heard. The order of the answer options on screen was counter-balanced across participants. The response period lasted three seconds after word offset, and trials were terminated at key press or upon timeout. Participants then received feedback on their response (i.e., correct/incorrect).

Following the practice task, participants moved on to the main 2AFC task. Here, the trial structure was identical to the practice trials, with the exception that participants did not receive feedback on their responses. Every participant encountered each of the five continuum tokens in random order in mini-blocks of five trials. Thus, the risk of running into identical fricatives in consecutive trials was minimized. The task consisted of 160 trials (i.e., 32 repetitions per step), and included three self-paced breaks after every 40 trials.

After the categorization task, participants began the post-experiment questionnaire (See supplementary material for responses). The completion of the questionnaire concluded the experiment and redirected participants back to Prolific.

3. Experiment 2 (Lab-based)

3.1. Method

3.1.1. Participants

Participant recruitment for Experiment 2 was subject to the same constraints as in Experiment 1. 32 participants were invited for participation. However, due to cancellation and non-compliance with task instructions, only data from 30 participants (age = 18-29, *M* = 22.03 (*SD* = 2.47)) were used in

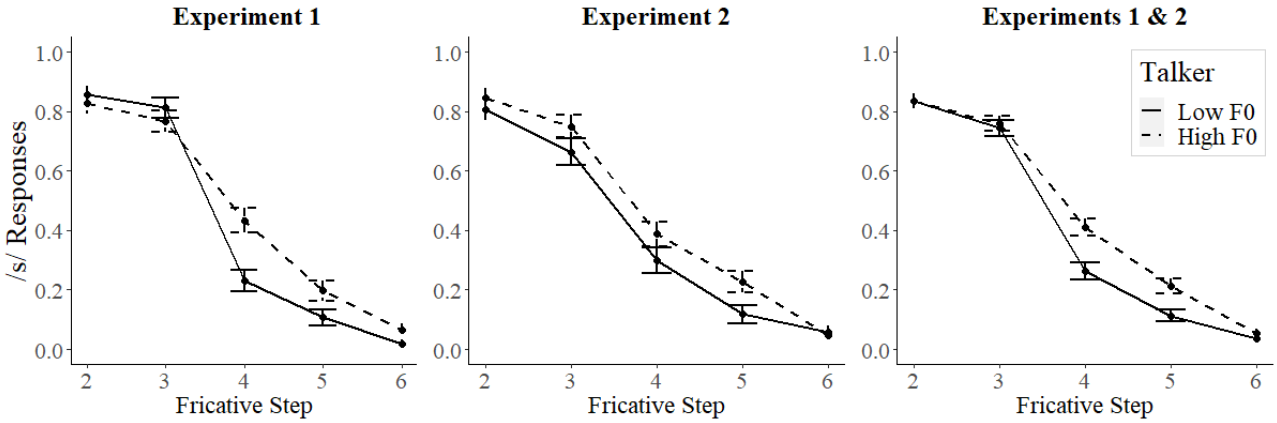

Figure 1: Each panel displays the proportion of /sɔk/ responses by fricative step (i.e., steps 2-6) and talker f0 group (solid = low f0, dashed = high f0). Error bars represent 5% CIs. The difference in response proportions across the two talker f0 groups is visible in all three plots, particularly in the more ambiguous fricative steps (i.e., steps 4-5). While the larger talker effect size at ambiguous fricatives aligned with our predictions, the direction of the effect was unexpected.

analyses. Participants were recruited through the Max Planck Institute for Psycholinguistics participant database. The two halves of the sample were once again randomly assigned to high and low f0 groups.

3.1.2. Materials

Experiment materials were identical to those of Experiment 1.

3.1.3. Procedure

The experiment was programmed and performed using Presentation® software (Version 23.0, Neurobehavioral Systems, Inc., Berkeley, CA). After providing informed consent, participants were seated in a sound-attenuating booth equipped with stereo headphones. They then performed the exposure task, the practice task, the 2AFC categorization task, and the post-experiment questionnaire, all of which were identical to those in Experiment 1.

4. Results

Given that recruitment criteria, task structures, trial numbers, and stimuli were all identical across the two experiments, their results were analyzed together. All data, analysis scripts, and stimuli are publicly available on OSF (https://osf.io/wfp9y).

4.1. Main Analyses

For the analysis of the 2AFC task, we used a generalized linear mixed-effect model (GLMM) with a logistic link function [24] using the "lmerTest" package [25] in R [26].

Out of a total of 245 non-timed-out responses to the practice trials across both experiments, there were 21 errors, 20 of which involved the miscategorization of the /s/ endpoint (i.e., step 1) as the /ʃ/ endpoint (step 8). This trend was consistent with the global /ʃ/ bias observed across both experiments (i.e., ~57% of responses were /ʃ/).

In Experiment 1, seven trials timed out, and 5113 trials were available for analyses. In Experiment 2, due to a technical error, one participant's data for the 2AFC task were incomplete (i.e., 123 trials out of 160) but were still included in the analyses. Given these missing trials and 10 timeouts, 4753 responses were available for analysis. Thus, we analyzed 9866 trials from 62 participants.

Response was coded as a binary variable where 1 corresponded to an /s/ response and 0 corresponded to an /ʃ/ response. Talker f0 was contrast coded such that low f0 corresponded to -0.5 and high f0 corresponded to 0.5. The model also included a fixed effect of Step (z-scored using the scale() function in R), and its interaction with Talker. In addition to the fixed effects of Talker and Step, the model also included a random intercept for Participants and by-participant random slopes for Talker and Step.

Within this model, Step was the strongest predictor of response proportions (β = -2.35, SE = 0.16, z = -14.90, p < 0.001), suggesting that the 5-step fricative continuum was generally performing as intended. Crucially, we found a significant effect of Talker on response proportions (β = 0.61, SE = 0.22, z = 2.80, p = 0.005), indicating that participants in the high f0 group gave more /s/ responses. This constitutes evidence for an *assimilatory* effect of talker f0, as opposed to the hypothesized contrastive effect. Finally, we found a significant interaction between Step and Talker (β = 0.62, SE = 0.31, z = 1.99, p = 0.04), indicating that the differences between the response proportions of the two talker f0 groups varied along the fricative continuum.

4.2. Secondary Analyses

4.2.1. Response homogeneity across experiments

We expanded the main analysis with an Experiment predictor with an interaction term[1] (contrast coded as -0.5 for Experiment 1 and 0.5 for Experiment 2). This did not improve model fit ($\Delta\chi^2$ (9) = 4.37, p = 0.36), nor did it find significant effects of or interactions with Experiment (See supplementary material for data and models). Thus, there was no evidence for different performance across online and lab settings.

4.2.2. No order effects

Extending the model with Miniblock (z-scored using scale()) and all possible interactions[2] did not reveal a significant effect

[1] resp ~ (step * talker * exp) + (1 + step | ppid)

[2] resp ~ (step * talker * mb) + (1 + step * mb | ppid)

of or interaction with Miniblock (See supplementary material for data). Thus, participants' responses did not significantly change as a function of trial order. It is worth noting that the addition of Miniblock improved model fit ($\Delta\chi^2$ (9) = 255, $p < 0.001$), but we did not report it as the primary model on account of an absence of significant effects associated with Miniblock.

5. Discussion

In the present study, we found an effect of talker f0 whereby response proportions in ambiguous fricative categorization were congruent with the f0 talker group. That is, participants who were exposed to a low f0 talker were more likely to categorize ambiguous fricatives as having lower CoG (i.e., /ʃ/-like), and participants who were exposed to a high f0 talker were more likely to categorize the same continuum as having higher CoG (i.e., /s/-like). These results constitute novel evidence for knowledge-driven influences of talker f0 on voiceless fricative perception. It should be noted that our results were obtained from two independent experiments with identical designs. These experiments also featured relatively long exposure tasks (i.e., a number of perceptual benefits of talker familiarity are observable after as little as 10 minutes [6], while our exposure phases lasted 20 minutes). It is therefore unlikely that the exposure phase of the experiment was too weak to induce a talker familiarity effect. Furthermore, the talker effect we report was not restricted to a single fricative from the continuum.

These results oppose existing reports of a contrastive link between f0 and fricative CoG perception. However, it must be emphasized that these reports rely on experiments where f0 information is provided through immediately preceding sentence context, e.g., [2], not through previously acquired talker-specific information. A number of diverse factors may be responsible for the unexpected direction of our talker effect. Given that a multitude of higher-level information such as a talker's (implied) gender [27] or even their perceived sexual orientation [9] can alter the perception of their speech, there may have been different higher-level mechanisms at work when talker information was involved [10]. Participants may have simply imagined the speaker substantially differently (e.g., physical appearance, gender/sexual identity, personality), which might have affected their perception of her speech in ways we could not control. Furthermore, the relatively large size of our pitch-shifting manipulation (i.e., 4 semitones) might have caused some participants to consider the speech in the exposure and test phases as coming from different talkers, preventing any talker-specific perceptual adjustment. Nevertheless, while the latter two factors might have affected between-participant variation, they do not explain the direction of the effect found here.

Additionally, although contrast effects seem to be much more common, there are reports of assimilatory spectral context effects in the literature, sometimes even with the same stimuli that resulted in contrastive effects before (e.g., [28]; contrastive and assimilatory effects of syllabic closure distances on voiced stop categorization). One study [29] specifically asserts that acoustic information can be continually used to retroactively weigh different possible perceptual interpretations of incoming speech, and that backward effects of acoustic information can be assimilatory in nature through this mechanism. Indeed, if there had been acoustic variation (particularly in f0) across 2AFC trials, re-weighing or training across trials or steps could have been possible. However, given that the post-fricative context in our experiments was static and trial structure/organization was highly predictable, constant reweighing of potential parses represents an unlikely mechanism for explaining our pattern of results as participants could quickly learn that there is only one post-fricative parse.

However, through another mechanism, the fact that the post-fricative context was identical in all trials can offer a possible explanation for the surprising direction of the talker effect. Specifically, within our design, it is not possible to determine whether the observed effect is a direct effect of talker information, or a modulation of a local context effect through talker information. Specifically, 20 minutes of exposure to high or low f0 may have affected the perception of the f0 information in /-ɔ-/ in the post-fricative portion of our stimuli in a contrastive manner. For instance, exposure to high talker f0 may have caused some participants to perceive the vowel as having a lower f0. The perception of a *local* low f0 vowel, in turn, may have affected the perception of the fricative's CoG contrastively, increasing the proportion of /s/ responses. A follow-up study in which participants perform a 2AFC task with isolated fricatives (i.e., no word or sentence context) to restrict all f0 information to talker exposure can better assess the directness of the effect observed here.

Finally, our design was highly constrained, affording a high degree of experimental control at the cost of ecological validity and task engagement. During the 2AFC task, participants were asked to make the same decision 240 times with a single minimal pair of monosyllabic Dutch words. Future studies may use a similar paradigm with a language that allows for a higher number of minimal pairs or a fricative continuum with a fewer number of steps in order to reduce the repetitiveness of the task.

It is worth emphasizing that different methodological configurations of the present paradigm have yielded different results in ongoing studies in our lab, some of which seem to point towards contrastive effects of the f0 in an immediately preceding context sentence as well as contrastive effects of talker f0. These new data will enable more precise methodological comparisons across experiments, ultimately leading to a clearer understanding of the factors that influence the direction and size of talker f0 effects.

6. Conclusions

There is a growing literature on the link between f0 and fricative CoG perception. In this study, we successfully identified a robust influence of a talker's typical f0 on subsequent voiceless fricative perception. Surprisingly, our results provide evidence for an assimilatory direction for this knowledge-driven effect, contrary to existing reports of contrastive signal-driven effects. Specifically, exposure to a high-f0 talker led to higher-CoG fricative perception and exposure to a low-f0 talker led to lower-CoG fricative perception. Future empirical work with additional methodological adjustments will be required to establish the direction of talker f0 effects more clearly and to shed light on the mechanisms that underlie them.

7. Acknowledgements

Many thanks to Sophie Slaats for lending her voice to the study across multiple recording sessions as well as allowing the publication of the stimuli for scientific purposes. Thanks to Maaike Nieuwenhuizen for helping with the streamlining of post-experiment questionnaire responses.

8. References

[1] C. Stilp, "Acoustic context effects in speech perception," *WIREs Cognitive Science*, vol. 11, no. 1, p. e1517, 2020, doi: 10.1002/wcs.1517.

[2] O. Niebuhr, "On the perception of 'segmental intonation': F0 context effects on sibilant identification in German," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2017, no. 1, p. 19, Aug. 2017, doi: 10.1186/s13636-017-0115-3.

[3] T. Wade and L. L. Holt, "Effects of later-occurring nonlinguistic sounds on speech categorization," *The Journal of the Acoustical Society of America*, vol. 118, no. 3, pp. 1701–1710, Sep. 2005, doi: 10.1121/1.1984839.

[4] L. L. Holt, "Temporally Nonadjacent Nonlinguistic Sounds Affect Speech Categorization," *Psychol Sci*, vol. 16, no. 4, pp. 305–312, Apr. 2005, doi: 10.1111/j.0956-7976.2005.01532.x.

[5] E. Holmes, Y. Domingo, and I. S. Johnsrude, "Familiar Voices Are More Intelligible, Even if They Are Not Recognized as Familiar," *Psychol Sci*, vol. 29, no. 10, pp. 1575–1583, Oct. 2018, doi: 10.1177/0956797618779083.

[6] E. Holmes, G. To, and I. S. Johnsrude, "How Long Does It Take for a Voice to Become Familiar? Speech Intelligibility and Voice Recognition Are Differentially Sensitive to Voice Training," *Psychol Sci*, vol. 32, no. 6, pp. 903–915, Jun. 2021, doi: 10.1177/0956797621991137.

[7] L. C. Nygaard and D. B. Pisoni, "Talker-specific learning in speech perception," *Perception & Psychophysics*, vol. 60, no. 3, pp. 355–376, Jan. 1998, doi: 10.3758/BF03206860.

[8] K. Johnson, "The role of perceived speaker identity in F0 normalization of vowels," *The Journal of the Acoustical Society of America*, vol. 88, no. 2, pp. 642–654, Aug. 1990, doi: 10.1121/1.399767.

[9] B. Munson, S. V. Jefferson, and E. C. McDonald, "The influence of perceived sexual orientation on fricative identification," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2427–2437, Apr. 2006, doi: 10.1121/1.2173521.

[10] A. A. Assgari and C. E. Stilp, "Talker information influences spectral contrast effects in speech categorization," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3023–3032, Nov. 2015, doi: 10.1121/1.4934559.

[11] E. Reinisch, "Speaker-specific processing and local context information: The case of speaking rate," *Applied Psycholinguistics*, vol. 37, no. 6, pp. 1397–1415, Nov. 2016, doi: 10.1017/S0142716415000612.

[12] M. Maslowski, A. S. Meyer, and H. R. Bosker, "How the tracking of habitual rate influences speech perception.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 45, no. 1, pp. 128–138, Jan. 2019, doi: 10.1037/xlm0000579.

[13] C. Ting and Y. Kang, "The Effect of Habitual Speech Rate on Speaker-Specific Processing in English Stop Voicing Perception," *Lang Speech*, p. 00238309231188078, Aug. 2023, doi: 10.1177/00238309231188078.

[14] J. Kreiman, B. R. Gerratt, K. Precoda, and G. S. Berke, "Individual differences in voice quality perception," *J Speech Hear Res*, vol. 35, no. 3, pp. 512–520, Jun. 1992, doi: 10.1044/jshr.3503.512.

[15] O. Baumann and P. Belin, "Perceptual scaling of voice identity: common dimensions for different vowels and speakers," *Psychological Research*, vol. 74, no. 1, pp. 110–120, Jan. 2010, doi: 10.1007/s00426-008-0185-z.

[16] E. Pépiot, "Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers," in *Speech Prosody 2014*, ISCA, May 2014, pp. 305–309. doi: 10.21437/SpeechProsody.2014-49.

[17] H. Quené, "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," *The Journal of the Acoustical Society of America*, vol. 123, no. 2, pp. 1104–1113, Feb. 2008, doi: 10.1121/1.2821762.

[18] P. Adank, R. van Hout, and R. Smits, "An acoustic description of the vowels of Northern and Southern Standard Dutch," *The Journal of the Acoustical Society of America*, vol. 116, no. 3, pp. 1729–1738, Sep. 2004, doi: 10.1121/1.1779271.

[19] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." Nov. 15, 2022. [Online]. Available: http://www.praat.org/

[20] M. Winn, A. Rhone, M. Chatterjee, and W. Idsardi, "The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants," *Frontiers in Psychology*, vol. 4, 2013, Accessed: Nov. 27, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00824

[21] A. L. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed, "Gorilla in our midst: An online behavioral experiment builder," *Behav Res*, vol. 52, no. 1, pp. 388–407, Feb. 2020, doi: 10.3758/s13428-019-01237-x.

[22] E. M. Cramer and W. H. Huggins, "Creation of Pitch through Binaural Interaction," *The Journal of the Acoustical Society of America*, vol. 30, no. 5, pp. 413–417, May 1958, doi: 10.1121/1.1909628.

[23] A. E. Milne *et al.*, "An online headphone screening test based on dichotic pitch," *Behav Res*, vol. 53, no. 4, pp. 1551–1562, Aug. 2021, doi: 10.3758/s13428-020-01514-0.

[24] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, Nov. 2008, doi: 10.1016/j.jml.2007.12.005.

[25] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82, pp. 1–26, Dec. 2017, doi: 10.18637/jss.v082.i13.

[26] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022. [Online]. Available: https://www.R-project.org/

[27] K. Johnson, E. A. Strand, and M. D'Imperio, "Auditory–visual integration of talker gender in vowel perception," *Journal of Phonetics*, vol. 27, no. 4, pp. 359–384, Oct. 1999, doi: 10.1006/jpho.1999.0100.

[28] B. H. Repp, "Bidirectional contrast effects in the perception of VC-CV sequences," *Perception & Psychophysics*, vol. 33, no. 2, pp. 147–155, Mar. 1983, doi: 10.3758/BF03202832.

[29] A. Rysling, A. Jesse, and J. Kingston, "Regressive spectral assimilation bias in speech perception," *Atten Percept Psychophys*, vol. 81, no. 4, pp. 1127–1146, May 2019, doi: 10.3758/s13414-019-01720-9.