



Ethnolinguistic Identification of Vietnamese-German Heritage Speech

Thanh Lan Truong¹, Andrea Weber¹

¹English Department, University of Tübingen, Germany

thanh-lan.truong@uni-tuebingen.de, andrea.weber@uni-tuebingen.de

Abstract

We examined whether listeners can identify the heritage of speakers based on their voices. In Experiment 1, native German listeners with and without an Asian heritage listened to sentences that had either been produced by native German speakers with a Vietnamese heritage or by Germans without any Asian heritage. Asian heritage listeners performed more accurately when speakers matched their heritage background, but they could identify German speakers with no Asian heritage with above chance accuracy. Listeners with no Asian heritage performed more accurately when the speakers had no Vietnamese background than when they did, but they could not identify speakers with a Vietnamese heritage with above chance accuracy. In Experiment 2, the pattern for German listeners with no Asian heritage was replicated with monotonized stimuli. Thus, matching heritage background facilitates ethnolinguistic identification for Vietnamese-German heritage speech, which persists even when pitch is absent.

Index Terms: Heritage speech, ethnolinguistic variation, Vietnamese German, speech perception

1. Introduction

Starting with the seminal study on New York City English by [1], small differences in segmental and suprasegmental features have been found to facilitate the identification of a speaker's ethnolinguistic background [2]. Thus, a speaker's voice can be an important medium, as it not only conveys the meaning of a message (i.e., linguistic cues) but also the identity of the messenger (i.e., indexical cues) [3]. The voice particularly can provide crucial information about the speaker's age, gender, emotions, and heritage of the speaker's identity [4, 5].

One extensively studied topic in this context concerns indeed the discriminability of African American and European American voices, with earlier seminal studies dating back to the 1950s [6]. Using a variety of methodologies, these studies set out to determine which acoustic cues listeners rely on for ethnolinguistic identification in an American context, and suggested cues ranging from vowel quality [6] and phonation type [5] to intonation [7] and voice quality cues such as jitter and shimmer [8]. Overall, researchers concluded that listeners access a variety of cues given certain experimental conditions, and they drew a complex picture of heritage identification, with listeners not only relying on a single cue for identification but rather on multiple cues, depending on which cues are available [5]. Nonetheless, the studies documented consistent results in that speech from European Americans could be reliably distinguished from African American speech based on the acoustic properties of the voice [6]. However, this ability seems to vary somewhat between ethnolects and ethnoracial groups [9].

For example, heritage identification for Asian heritage speakers in an American context is less clear. While listeners from various backgrounds showed high accuracy rates at above chance with White Americans and Black Americans, they only achieved below chance when identifying Asian heritage speakers [2]. A growing but still limited number of studies have so far explored ethnolinguistic identification of Asian American speakers, raising the question of whether there is a perceptible distinction between Asian American English and White American English [10, 11, 12, 13]. While no reliable perception of Korean American identity was found [14], other studies found that listeners can perceive the heritage of Asian American speakers at above chance rates [10, 11, 12]. The pioneering work by [10] found in a forced-choice task that Asian American and White American listeners can distinguish Asian Americans from White Americans without visual cues and regardless of the listeners' own ethnicity. Following up on that, [11] found that Asian American listeners, raised in New York City, could be even more successful than non-Asian listeners when identifying Asian American speakers [11]. Similarly, Asian American speakers were judged as significantly more likely to be of Asian heritage in a rating task with Asian American listeners [12]. This pattern of results suggests that a shared heritage can possibly facilitate accurate identification. However, there is to date no consensus regarding the specific acoustic cues that listeners rely on for ethnolinguistic identification [11, 13, 12]. This might be linked to the fact that there is no clear definition of 'Asian American English' [15, Chapter 1] or other ethnically-specific varieties of 'Asian sounding' [12]. [11] analyzed suprasegmental features including jitter, shimmer, phonation type, and rhythm. Although they observed some differences in phonation type between the speaker groups, they did not find that these acoustic features correlated strongly with identification rates. [12] further suggested tone and prosody as potential factors for heritage identification in their metalinguistic commentary. While most studies looked at specific groups of Asians (e.g., Chinese or Korean), one exploratory study examined a small speech sample by multiple Asian American ethnicities (Chinese, Filipino, Korea, and Vietnamese) [9]. The analyses have shown both similarities and significant differences within this broader Asian community, which makes it the more challenging to define 'Asian sounding' voices. Besides the missing consensus on which acoustic features drive Asian heritage identification, there are also to date no studies that investigated Asian heritage identification outside of an American context.

In the present study, we contribute to this line of research by examining ethnolinguistic identification of an Asian heritage background in a German language context. Similar to the unique position that Asian Americans hold within the American society - one of 'forever foreigner' [15, Chapter 1], in Germany,

Asian-German heritage speakers are also met with contradictory views [16]. On the one hand, they are clearly portrayed as outsiders. On the other hand, they are often depicted as a ‘model minority’ or ‘elite migration’ [17]. This complexity illustrates a blend of unity and diversity for Asian-Germans in the German society. We asked in our study if Vietnamese-heritage speakers of German sound Asian when they speak German. Native German listeners, who either had a bicultural East/Southeast Asian heritage background (= Asian heritage Germans) or a monocultural German background (= non-Asian heritage Germans), had to identify speakers who were either bicultural Vietnamese heritage Germans or monocultural non-Vietnamese heritage Germans. Based on previous findings, we predicted that Asian heritage Germans would outperform those without such a heritage background. In addition, it was expected that Asian heritage listener would be particularly good at identifying other Asian heritage speakers.

2. Experiment 1

2.1. Material

Sixteen speakers (8 non-Vietnamese Germans, mean age = 25.3: 4 male, 4 female; 8 Vietnamese-Germans, mean age = 26.5: 4 male, 4 female) were recruited for recording. All Vietnamese heritage German speakers were second generation, first born and raised in Germany, whose parents immigrated to Germany later in life. All of them reported using Vietnamese on a daily basis but considered their proficiency in German to be higher than their heritage language. The speakers recited one sentence from orthographic transcription: *Flöhe können um das Hundertfache ihrer eigenen Körperlänge in die Höhe springen* (“flea can jump a hundredfold time their body length”). We opted for this sentence as its length would provide listeners with enough information for speaker heritage identification. Recordings were made in a sound attenuated room with a high-quality microphone and a sampling rate of 44 kHz with Praat (version: 6.3.09) [18]. The recorded sentences from the Vietnamese-German speakers were not significantly longer on average than the recorded sentences from the non-Vietnamese German speakers (Vietnamese-Germans: 4658.5 ms, non-Vietnamese Germans: 4820.5 ms) ($t = -0.9$, $df = 14$, $p = 0.4$). There was also no significant difference in F0 of the complete sentence between those groups (mean pitch of bicultural speakers = 155.1 Hz; mean pitch of monocultural speakers = 172.7 Hz) ($t = -0.7$, $df = 14$, $p = 0.5$).

2.2. Participants

Twenty-five Asian heritage Germans¹ and 25 non-Asian Germans (who grew up with only German) between the ages of 18 and 35 years participated in the study as listeners (mean age: 25.1; 29 females, 2 undisclosed). All participants were recruited and compensated through Prolific (www.prolific.com) [19]. None of the participants reported having hearing or visual impairments. Three Asian heritage Germans and on non-Asian German had to be excluded since they did not complete the experiment according to the instructions. While two non-Asian German participants reported having some knowledge in Japanese but had little to no contact with Japanese speakers, none of them reported having prior knowledge of any other Asian language.

¹Heritage background comprised 13 Vietnamese, 6 Chinese, 2 Japanese, 2, Korean, 1 Mongolian, 1 Bengali.

2.3. Procedure

The experiment² was conducted online and run on Gorilla, an experiment builder software (www.gorilla.sc) [20]. Participants gave electronically informed consent, followed by written instructions, asking them to do the experiment in a quiet room. Participation was only permitted using a laptop or a computer with Google Chrome. Participants were automatically disqualified if the technical requirements were not met. To ensure that participants used headphones during the experiment, we conducted a headphone-screening test beforehand. This task comprised six trials, of which participants were required to pass five of them to proceed with the experiment [21]. In each trial, three intervals of randomly ordered white noise with equal frequency and duration were presented, but one interval contained a Huggins Pitch tone, which they needed to identify. All experimental material was loaded prior to the start of the experimental session to ensure no loading issues. The experiment started with two practice trials, followed by the 16 recorded sentences, presented in random order for each participant. After each recording, participants were asked to indicate if the speaker was either Vietnamese heritage German (bicultural) or non-Vietnamese heritage German (monocultural). After each response, participants to indicate on a 5-point Likert scale how confident they were with their response. The scale ranged from ‘just a guess’ to ‘sure’. After the experiment, participants filled out a language background questionnaire, asking about their prior experience with East-Asian languages. Finally, they were also asked to indicate what they believed aided them in identifying speakers’ heritage. Options to choose from included speech rhythm, speech rate, intonation, voice quality, and a free comment text box.

2.4. Results

Overall, both Asian heritage and non-Asian heritage listeners could distinguish the heritage of the speakers at above chance (mean of the raw data for bicultural speakers = 0.64; mean of the raw data for monocultural speakers = 0.77). Most intriguingly, while Asian heritage Germans performed above chance for both speaker groups, non-Asian heritage listeners performed above chance for only the monocultural speakers (see Figure 1).³ We used a generalized linear mixed effects model (GLMM) with the lme4 package in R (version: 4.2.3) [22] for data analyses. The model included, *correctness* as the dependent variable and *speaker group* and *listener group* as the independent variables in an interaction term. ‘Bicultural German’ and ‘monocultural German’ for *speaker group* (bicultural vs. monocultural) and *listener group* (Asian heritage German vs. non-Asian heritage German) were sum coded as 0.5 and -0.5, respectively. The random effect structure included *by-participant* varying intercepts and varying slopes for *speaker group*. A significant main effect of *speaker group* ($b = -0.66$, $SE = 0.22$, $z = -3.03$, $p = .002$) was found. Furthermore, an interaction was observed between *speaker group* and *listener group* ($b = -1.52$, $SE = 0.42$, $t = -3.63$, $p = .0002$), indicating that when speaker and lis-

²The procedure for the present study was approved by the DGFS (Deutsche Gesellschaft für Sprachwissenschaft) ethics committee for the Chair of Psycholinguistics and Applied Linguistics at the University of Tübingen.

³The mean values for the conditions were: Asian heritage German listeners - bicultural speakers = 0.73, Asian heritage German listeners - monocultural speakers = 0.71; non-Asian heritage German listeners - bicultural speakers = 0.56, non-Asian heritage listeners - monocultural speakers = 0.83.

tener matched in heritage, the correctness rate increased. The findings suggest that listeners can identify the heritage of the speaker quite well and that they experience a boost in identification when listeners match with the heritage of the speakers. Taken together, Asian heritage German listeners were better at identifying the heritage background of bicultural German speakers than that of monocultural German speakers; non-Asian heritage German listeners were better at identifying the heritage background of monocultural German speakers than that of bicultural German speakers.

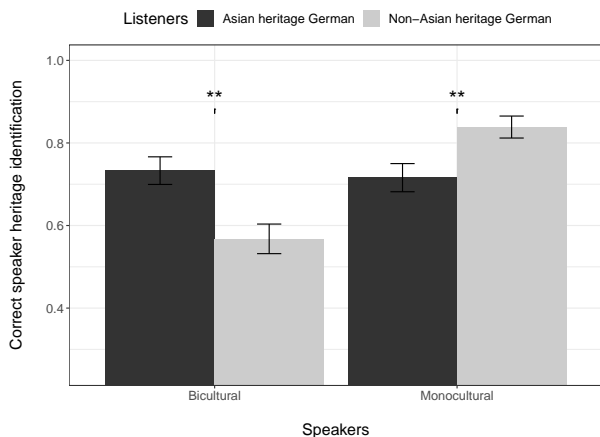


Figure 1: Average correct identification scores in the heritage identification test of the raw data. Vertical bars represent standard errors. Asterisks (**) indicate significance ($p < .01$).

We conducted a subsequent analysis to investigate whether a heritage match is also reflected in how confident participants were in their judgment of correct identification responses. Their confidence was measured by converting the Likert-scale responses to a numeric response, with ‘just a guess’ corresponding to ‘1’ to ‘certain’ corresponding to ‘5’. A linear mixed-effects model (LMM) was used to investigate participants’ confidence in correctly identifying speakers’ ethnicity. The model included *ratings* as the dependent variable and the interaction between *speaker group* and *listener group* as independent variable (bicultural speakers and monocultural speakers, sum coded as 0.5 and -0.5, respectively). In addition, a random intercept for participants and random slope for *speaker group* was also included. The analysis showed also an interaction between *speaker group* and *listener group* ($b = -0.7$, $SE = 0.24$, $t = -2.9$, $p = .005$), suggesting that when speakers and listeners match in their heritage, listeners were more confident in correctly identifying the speakers’ ethnicity (see Figure 2).

Overall, participants were able to identify the heritage of the speakers quite well and performed even better when there was a heritage match between speaker and listener, in line with [11]. However, little is known about which acoustic cues the listeners possibly utilized for heritage identification. All of our bicultural and monocultural speakers were born and raised in Germany. Their pronunciation in German was clearly native, and the bicultural speakers did not differ from the monocultural in any obvious ways. Nevertheless, some more subtle cues must have given away their heritage background [2]. We then examined participants’ comments, which indicated the acoustic signals they believed helped them identify the speakers’ heritage. Out of 46 participants, 40 participants selected intonation as the primary acoustic cue, followed by speech rhythm (30 re-

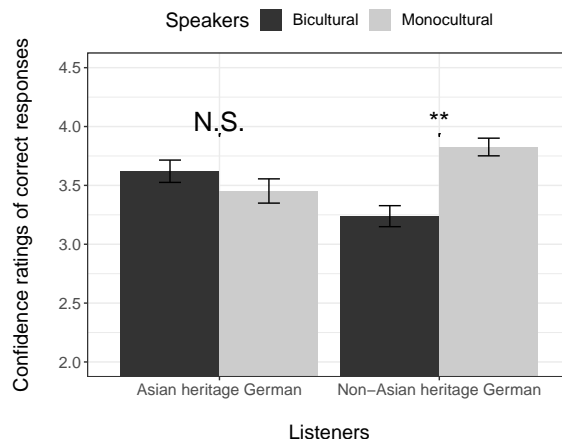


Figure 2: Participants confidence ratings of correct responses. Higher numbers indicate increased confidence. Vertical bars represent standard errors. Asterisks (**) indicate significance ($p < .01$). Non-significance is denoted by *N.S.*

sponses), voice quality (12 responses), and speech rate (10 responses). This aligns with the metalinguistic commentary on intonation by [12], as Korean American participants tended to mention suprasegmental attributes like intonation as the major influence in their heritage identification. Prompted by this, we wanted to extend the findings of Experiment 1, and eliminated F0 in our stimuli in Experiment 2. This time, we only tested non-Asian German listeners and their ability to distinguish between monocultural and bicultural speakers. If listeners can shift to acoustic cues other than intonation, then they should still be able to successfully identify the heritage of the speakers with flattened stimuli. However, if F0 is a crucial cue for listeners, then identification rates should generally drop and the boost from a heritage match between listener and speaker may disappear.

3. Experiment 2

3.1. Material

For Experiment 2, we used the sentence recordings from Experiment 1, but we monotonized all stimuli in Praat [18] and flattened the pitch contour, using the Praat vocal toolkit (<https://www.praatvocaltoolkit.com/index.html>).

3.2. Participants

Twenty-five non-Asian heritage Germans, who all grew up monolingually with German, participated. They were between the ages of 19 and 40 years (mean age = 26.7, $SD = 5.9$, 15 female). All participants were recruited and compensated through Prolific. None of the participants reported having hearing or visual impairments. Two participants had to be excluded from further analyses since they did not follow the instructions. Only one participant reported having some knowledge in Korean.⁴

3.3. Procedure

The procedure was identical to Experiment 1.

⁴We re-run the analysis after excluding this participant, and it did not change the results ($b = -0.65$, $SE = 0.21$, $z = -3.02$, $p = 0.002$).

3.4. Results

Although the identification rate of non-Asian heritage German listeners was slightly lower than in Experiment 1, the performance pattern was similar. Specifically, non-Asian heritage German listeners identified monocultural speakers' identity above chance (mean of the raw data = 0.69) but not for the bicultural speakers (mean of the raw data for bicultural speakers = 0.54) (see Figure 3). As in Experiment 1, we conducted a GLMM analysis. The model included *correctness* as the dependent variable and *speaker group* as an independent variable. The random effect structure was identical to that of Experiment 1. The results showed a significant main effect of *speaker group* ($b = -0.62$, $SE = 0.21$, $z = -2.07$, $p = .002$), indicating that heritage identification was again better when listener and speaker matched in heritage background, even when no pitch information was available; that is, non-Asian German listeners identified monocultural German speakers better than bicultural German speakers.

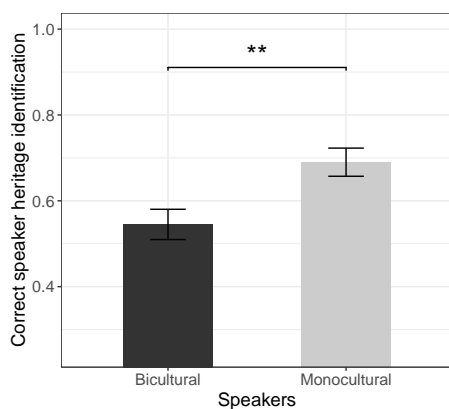


Figure 3: Average correct identification scores in the heritage identification test of non-Asian German heritage listeners of the raw data. Vertical bars represent standard errors. Asterisks(**) indicate significance ($p < .01$).

The analysis for participants' confidence ratings was similar to Experiment 1 ($b = -0.4$, $SE = 0.16$, $t = -2.65$, $p = .013$) (see Figure 4). Non-Asian heritage German listeners were more confident in correctly identifying the ethnicity of the monocultural speakers, in line with the findings of Experiment 1.

4. Conclusions

We investigated whether German listeners can distinguish the ethnolinguistic background of Vietnamese-German heritage speakers from non-Asian German speakers based on the speakers' voices, using the forced-choice task. Experiment 1 showed higher accuracy rates when the heritage of speakers and listeners matched, which is in line with [11]. That is, Asian heritage German listeners excelled at identifying bicultural speakers, while non-Asian heritage German listeners performed better with monocultural speakers. Thus, both listener groups performed more accurately when the speakers matched their heritage background. However, Asian heritage listeners could also identify heritage-mismatching German speakers with above chance accuracy. Conversely, non-Asian German listeners failed to do the same with Asian German heritage speakers, which suggests that Asian heritage listeners are exposed to both speech communities, enabling them to identify matching and

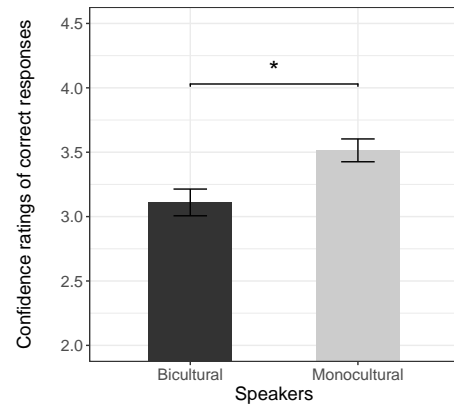


Figure 4: Participants confidence ratings of correct responses of non-Asian German heritage listeners. Higher numbers indicate higher confidence. Vertical bars represent standard errors. The asterisk (*) indicates significance ($p < .05$).

mismatching heritage speakers better than their non-Asian German counterparts. This pattern was also reflected in their confidence ratings. While no significant difference was observed for Asian heritage German listeners, non-Asian heritage German listeners exhibited higher confidence when identifying speakers from their own speaker group. As a whole, these results showed that Vietnamese heritage German speakers can indeed sound 'Asian' when speaking German and therefore successfully extended previous findings in the US context to the German context. It is as of yet not established which acoustic cues listeners utilize for heritage identification. According to [5], asking which cues listeners use for heritage identification may not be the most appropriate approach, as this can be examined in numerous ways, posing significant challenges in experimental design, ranging from stimuli selection to controlled listening environment [23]. Motivated by our participants' responses that intonation may have helped them when identifying the heritage of the speakers, which is also in line with [12] metalinguistic commentary, we monotonized all stimuli in Experiment 2. Findings of Experiment 2 showed that non-Asian listeners could identify the heritage of the speakers, and the heritage-match advantage still positively affected correct identification rates. This finding suggests that listeners were capable of using cues other than pitch. However, this does not mean that pitch information was not used in Experiment 1. Given that pitch information plays a crucial role in speech processing [24] and the indexicality of the voice is processed as early as 200-300ms after word onset [25], it seems rather unlikely that listeners would disregard pitch information when available. Although the scale of research needed to characterize 'Asian sounding' in German is larger than could be achieved here, this study took a preliminary step by providing an important finding that 'Asian sounding' is not exclusive to the American context. That said, this study took an important first step in research on the Vietnamese-German community as a whole, and there remains a need for larger, more in-depth studies that analyze Vietnamese-German heritage speech styles and regions. Following from this study, our hope is that the present findings provide a foundation of initial data on Vietnamese-German heritage speech and that it will encourage further investigation and understanding of linguistic diversity in Germany.

5. References

- [1] W. Labov, *The Social Stratification of English in New York City*, 2nd ed. Cambridge: Cambridge University Press, 1966.
- [2] E. Kushins, "Sounding like your race in the employment process: An experiment on speaker voice, race identification, and stereotyping," *Race and Social Problems*, vol. 6, pp. 237–248, 2014.
- [3] D. Abercrombie, *Elements of General Phonetics*. Edinburgh University Press, 1967.
- [4] S. C. Creel and M. R. Bregman, "How talker identity relates to language processing," *Language and Linguistics Compass*, vol. 5, no. 5, pp. 190–204, 2011.
- [5] E. R. Thomas, N. E. Lass, and J. Carpenter, *Chapter 12. Identification of African American Speech*. Berlin, New York: De Gruyter Mouton, 2010, pp. 265–288.
- [6] E. R. Thomas and J. Reaser, "Delimiting perceptual cues used for the ethnic labeling of african american and european american voices," *Journal of Sociolinguistics*, vol. 8, pp. 54–87, 2004.
- [7] C. G. Foreman, "Dialect identification for prosodic cues," in *14th International Congress of Phonetic Sciences*, San Francisco, CA, USA, Aug. 1999, pp. 1237–1240.
- [8] T. Purnell, W. Idsardi, and J. Baugh, "Perceptual and phonetic experiments on american english dialect identification," *Journal of Language and Social Psychology*, vol. 18, no. 1, pp. 10–30, 1999.
- [9] C. B. Chang and D. Dionne, "Unity and diversity in Asian American language variation: Data from Chinese, Filipino, Korean, and Vietnamese Americans," in *Proceedings of Meetings on Acoustics*, vol. 46, no. 1, 2022, p. 060002.
- [10] D. B. Hanna, "Do I sound "Asian" to you?: Linguistic markers of Asian American identity," *University of Pennsylvania Working Papers in Linguistics*, vol. 4, no. 2.
- [11] M. Newman and A. Wu, "Do you sound Asian when you speak English? Racial identification and voice in Chinese and Korean Americans' English," vol. 86, no. 2, pp. 152–178.
- [12] A. Cheng and S. Cho, "The effect of ethnicity on identification of Korean American speech," *Languages*, vol. 6, no. 4, p. 186.
- [13] P. Wong and M. Babel, "Perceptual identification of talker ethnicity in Vancouver English," *Journal of Sociolinguistics*, vol. 21, no. 5, pp. 603–628.
- [14] E. H. Lee, *Speech production and perception of heritage speakers of Korean*. Los Angeles: UCLA MA thesis, 2014.
- [15] A. Reyes and A. Lo, "Introduction: On Yellow English and other perilous terms," in *Beyond Yellow English: Toward a Linguistic Anthropology of Asian Pacific America*, A. Reyes and A. Lo, Eds. Oxford University Press, 2009, pp. 3–17.
- [16] N. D. Hartlep and P. Bui, "Vietnamese students and the emerging model minority myth in Germany," *Sage Publications Ltd*, pp. 518–536, 2020.
- [17] K. König, *Spracheinstellungen Und Identitätskonstruktion. Eine Gesprächsanalytische Untersuchung Sprachbiographischer Interviews mit Deutsch-Vietnamesen*. München: De Gruyter (A), 2014.
- [18] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer [computer program], version 6.3.09," 2023. [Online]. Available: <http://www.praat.org/>
- [19] (2023) Prolific. Accessed December 4, 2023. [Online]. Available: <https://www.prolific.co/>
- [20] A. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. Evershed, "Gorilla in our midst: An online behavioral experiment builder," *Behavior Research Methods*, vol. 52, no. 52, pp. 388–407, 2020.
- [21] A. E. Milne, R. Bianco, K. C. Poole, S. Zhao, A. J. Oxenham, A. J. Billig, and M. Chait, "An online headphone screening test based on dichotic pitch," *Behavioral Research*, vol. 53, pp. 1551–1562, 2021.
- [22] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Available: <https://www.R-project.org/>
- [23] E. R. Thomas, "Sociophonetic applications of speech perception experiments," vol. 77, no. 2, pp. 115–147.
- [24] J. Kuang and M. Liberman, "Integrating voice quality cues in the pitch perception of speech and non-speech utterances," *Frontiers in Psychology*, vol. 9, 2018.
- [25] J. J. A. Van Berkum, D. Van den Brink, C. M. J. Y. Tesink, M. Kos, and P. Hagoort, "The neural integration of speaker and message," *Journal of Cognitive Neuroscience*, vol. 20, no. 4, pp. 580–91, 2008.