The following publication Shao, J., Bakhtiar, M., & Zhang, C. (2022). Impaired Categorical Perception of Speech Sounds Under the Backward Masking Condition in Adults Who Stutter. Journal of Speech, Language, and Hearing Research, 65(7), 2554-2570 is available at https://doi.org/10.1044/2022_JSLHR-21-00276

Impaired categorical perception of speech sounds under the backward masking condition in adults who stutter

Jing Shao[1], Mehdi Bakhtiar[2*], Caicai Zhang[3]

1. Department of English Language and Literature, Hong Kong Baptist University, Hong Kong Special Administrative Region.
2. Unit of Human Communication, Development, and Information Sciences, University of Hong Kong, Hong Kong Special Administrative Region.
3. Department of Chinese and Bilingual studies, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region.

*Corresponding author:

Mehdi Bakhtiar[*], Unit of Human Communication, Development, and Information Sciences, University of Hong Kong, Hong Kong Special Administrative Region. Email: mbakht@hku.hk

Running head: Categorical perception under backward masking

Disclosure: *The authors have declared that no competing interests existed at the time of publication. No specific funding has been received for this project.*

Impaired categorical perception of speech sounds under the backward masking condition in adults who stutter

Abstract

Purpose: Evidence increasingly indicates that people with developmental stuttering have auditory perception deficits. Our previous research has indicated similar but slower performance in categorical perception of the speech sounds under the quiet condition in children and adults who stutter compared with their typically fluent counterparts. We hypothesised that the quiet condition may not be sufficiently sensitive to reveal subtle perceptual deficiencies in people who stutter. The current study examined this hypothesis by testing the categorical perception of speech and nonspeech sounds under backward masking condition (i.e., a noise was presented immediately after the target stimuli).

Method: Fifteen Cantonese-speaking adults who stutter (AWS) and 15 adults who do not stutter (AWNS) were tested on the categorical perception of four stimulus continua, namely, consonant varying in VOT, vowel, lexical tone, and nonspeech, under the backward masking condition using identification and discrimination tasks.

Results: AWS demonstrated a broader boundary width than AWNS in the identification task. AWS also exhibited a worse performance than AWNS in the discrimination of between-category stimuli but a comparable performance in the discrimination of within-category stimuli, indicating reduced sensitivity to sounds that belonged to different phonemic categories among AWS. Moreover, AWS showed similar patterns of impaired categorical perception across the four stimulus types, although the boundary location on the VOT continuum occurred at an earlier point in AWS than in AWNS.

Conclusion: The findings provide robust evidence that AWS exhibit impaired categorical perception of speech and nonspeech sounds under the backward masking condition. Temporal processing (i.e. VOT manipulation), frequency/spectral/formant processing (i.e., lexical tone or vowel manipulations), and non-linguistic pitch processing were all found to be impaired in AWS. Altogether, the findings support the hypothesis that AWS might be less efficient in accessing the phonemic representations when exposed to a demanding listening condition.

Keywords: stuttering, adults who stutter, categorical perception, backward masking

Introduction

Stuttering is a speech disorder characterised by disruptions in speech fluency due to an intermittent loss of speech motor control (Guitar, 2013). The prevalence of developmental stuttering is around 1% in the adult population and ranges from 5% to 8% among preschool-age children (Bloodstein, 1995; Yairi & Ambrose, 2013). Although stuttering manifests symptomatically through motor speech behaviours (Teesson, Packman, & Onslow, 2003), accumulating evidence points to a role for abnormal auditory perception in this condition.

At the neural level, multiple neuroimaging studies of people who stutter (PWS) have indicated deactivation in the auditory cortex in relation to speech processing (for a review, see Etchell, Civier, Ballard, & Sowman, 2018). For instance, Sato et al. (2011) reported that adults who stutter (AWS) did not show the typical left-hemispheric advantage in the auditory perception of phonemic contrasts compared with prosodic contrasts, suggesting that AWS have atypical functional lateralisation of speech processing. Several event-related potential (ERP) studies have suggested that PWS may exhibit atypical responses in both speech-sound discrimination (Corbera, Corral, Escera, & Idiazábal, 2005; Jansson-Verkasalo et al., 2014; Liotti et al., 2010) and nonspeech-sound processing (Goncalves, Andrade, & Matas, 2015; Hampton & Weber-Fox, 2008; Kaganovich, Wray, & Weber-Fox, 2010). For instance, Hampton and Weber-Fox (2008) used a pure-tone oddball task to show that the early auditory perceptual processes (as revealed by the N100 and P200 ERP components) were disrupted in only a small subset of AWS, whereas later perceptual processes (as revealed by the P300 ERP component) were more disrupted in AWS than in adults who do not stutter (AWNS), implying that the former group experiences weaker updates in working memory in response to target tones. Similarly, Kaganovich et al. (2010) demonstrated that children who stutter (CWS) had normal N1 and P1 components but exhibited reduced P300 amplitudes to rare deviants compared with children who do not stutter (CWNS). In contrast, Corbera et al. (2005) found that AWS and AWNS showed similar mismatch negativity (MMN) responses when listening in the simple tone condition; however, AWS demonstrated significantly enhanced MMN when listening to phonetic contrasts in the speech condition. Furthermore, enhanced MMN was shown to be positively correlated with stuttering severity, suggesting that AWS may have abnormal speech sound representations (Corbera et al., 2005).

At the behavioural level, different tasks are used to examine auditory perception in PWS. Sentence identification is a commonly used task. Hall and Jerger (1978) found that PWS performed worse than people who do not stutter (PWNS) when identifying synthetic sentences, indicating a deficiency in auditory perception. However, other studies have reported that group differences between PWS and PWNS in sentence identification were either conditional (e.g., appeared in the presence of competing talkers) (Toscher & Rupp, 1978) or nonsignificant (Kramer, Green, & Guitar, 1987). Moreover, studies based on nonword repetition tasks have observed less accurate and slower performance in PWS than in their typically fluent counterparts (for a review, see Ofoe, Anderson, & Ntourou, 2018). Whether these tasks can sufficiently tease apart problems in auditory processing, linguistic/phonological processing, and/or articulatory motor programming and execution remains controversial.

More recently, categorical perception has been used to assess the perceptual abilities of PWS. In the auditory modality, categorical perception refers to the phenomenon in which two stimuli from different phonemic categories are more detectable than two stimuli from the same category, even though the stimulus pairs have equivalent acoustic differences (Bent & Holt, 2013; Harnad, 1987; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Massaro &

Simpson, 2014). This difference may reflect higher-level phonological processing abilities (Zhang, Shao, & Huang, 2017). Categorical perception is usually assessed using an identification task and a discrimination task. Typically, normal categorical perception manifests as an abrupt response shift across categorical boundaries (e.g., from /pa/ to /p^{h}a/) in the identification task and higher discrimination accuracy between stimulus pairs that cross categorical boundaries in the discrimination task (Repp, 1984).

Neef et al. (2012) were the first to compare categorical perception between AWS and AWNS using two voice-onset time (VOT) continua (/bə /-/pə/ and /də/-/tə/). The models fitted to the identification of responses were flatter when using data from AWS than data from AWNS. Additionally, AWS exhibited poorer discrimination performance than AWNS. Together, these findings suggest reduced speech perceptual acuity in AWS, which may have been due to ambiguous phoneme representations.

In line with this research, Bakhtiar, Zhang, and Ki (2019) examined categorical perception in Cantonese-speaking CWS. They found that the response times (RTs) for identifying stimuli across two categories were longer in CWS than in CWNS, suggesting a slowing of the processing speed associated with accessing phonemic representations of stimuli that cross phonemic boundaries in CWS. These findings are somewhat incongruent with those of Neef et al. (2012), as Bakhtiar et al. (2019) did not find group differences in either the slope of the response shift or the accuracy of stimulus discrimination across phonemic boundaries.

To further test whether the divergent findings of Bakhtiar et al. (2019) and Neef et al. (2012) were due to differences in the age groups included in the samples (children and adults, respectively), Bakhtiar, Shao, Cheung, and Zhang (2021) examined the categorical perception abilities of AWS and AWNS using a similar paradigm. They observed comparable categorical perception between the two groups and a trend towards slower discrimination of between-category stimuli than of within-category stimuli in the AWS group relative to the AWNS group. These findings confirm the previous findings of Bakhtiar et al. (2019) but are again inconsistent with the results of Neef et al. (2012) in adults. One explanation for this discrepancy might be that the categorical perception tasks used in previous studies (Bakhtiar et al., 2019, 2021) may have not been sufficiently demanding or sensitive to reveal subtle speech perception difficulties in CWS and AWS. For instance, the step distance between the acoustic stimuli in the VOT continuum was larger in our study than in that of Neef et al. (2012) (~6 ms versus ~1 ms). This might explain our previous observation of slower but comparable categorical perception accuracy in AWS and CWS versus their typically fluent counterparts (Bakhtiar et al., 2019, 2021).

Another paradigm that has been used to investigate the perceptual abilities of PWS is backward masking (Basu, Schlauch, & Sasisekaran, 2018; Howell, Davis, & Williams, 2006; Howell & Williams, 2004). Backward masking has long been thought to rely heavily on central auditory processing and has been extensively investigated in both adults and children (Buss, Hall, Grose, & Dev, 1999; Hill, Hartley, Glasberg, Moore, & Moore, 2006; Hill, Hogben, & Bishop, 2006; Porter, Spitzer, Buss, Leibold, & Grose, 2018; Rosen, Adlar, & Van Der Lely, 2009; Rosen & Manganari, 2001). This measure is used to assess a person's ability to detect a sound that is followed by a masker. Backward masking requires segregation of the stimulus from the masking noise and thus demands an increase in cognitive effort, which can disrupt both the sensory and phonological processing involved in speech perception. Studies have shown that both AWS and CWS performed worse in non-speech tone detection tasks under the backward masking condition than the quiet condition, especially when compared with their typically fluent counterparts (Howell et al., 2000, 2004) and children who had recovered from stuttering (Howell et al., 2006). For instance, Howell et al. (2006) reported that the backward-masked thresholds in children with persistent stuttering were 10 dB higher than those in children who had recovered from stuttering. In another study, Basu et al. (2018) examined the auditory perceptual abilities of AWS using the backward masking of tones and speech syllables when the masker was presented after a delay of 0 or 300 ms before the onset of the auditory signal. In the tone detection task, no significant difference was observed between AWS and AWNS in either the quiet or backward masking condition with delays of 0 and 300 ms. However, an additional analysis (subtraction of the thresholds at quiet from the thresholds for backward masking at 0 ms) revealed a significant difference between the groups. Based on these results, the authors concluded that the tone detection task may not be sufficient to distinguish AWS from AWNS. Regarding the speech recognition task, which required verbal repetition of recognised syllables, the AWS group performed significantly more poorly in both quiet and backward masking conditions. Basu et al. (2018) proposed that AWS may not be able to distinguish phonemic boundaries.

As noted above, we speculate that the inconsistent findings reported by Neef et al. (2012) and Bakhtiar et al. (2019, 2021) may be due to the differences in task demands. Therefore, the first aim of the current study was to explicitly test the hypothesis that AWS show speech perception deficits in more demanding tasks. Accordingly, we introduced backward masking into the classical categorical perception paradigm. Therefore, it is expected that the categorical perception under the backward masking may provide a more demanding condition, enabling the detection of subtle auditory perceptual and phonological deficits in PWS.

The second aim of this study is to examine whether the categorical perception deficit identified in AWS by Neef et al. (2012) is specific to temporal processing (e.g., VOT) of speech sounds or whether other acoustic aspects, such as frequency or spectral analysis (i.e., vowels and lexical tones), might also be implicated. Vowels are acoustically characterised by formants in the sound spectrum and thus primarily rely on frequency/spectral analysis (Ladefoged & Johnson, 2015; Peterson & Barney, 1952). The first two formants (F1–F2) are crucial for distinguishing most vowels (Ladefoged & Johnson, 2015; Peterson & Barney, 1952): the F1 frequency is roughly associated with the height of a vowel, and the F2 frequency with the frontness of a vowel. Therefore, the perception of vowels depends crucially on the analysis and detection of the frequency locations of formants in the sound spectrum. Previous neuroimaging studies (Beal et al., 2010; Corbera et al., 2005; Jansson-Verkasalo et al., 2014) have reported abnormal brain responses to vowel stimuli in PWS. Compared with CWNS, CWS demonstrated significantly smaller MMN amplitudes in response to vowel changes (Jansson-Verkasalo et al., 2014). Using the MEG technique, Beal et al. (2011) compared auditory processing and speech-induced suppression during active vowel production versus passive listening to pre-recorded vowel productions and observed a longer left-hemisphere latency for M50 in CWS than in CWNS. Together, these results imply that PWS may also exhibit impaired spectral processing of speech. Based on these neuroimaging findings, we speculate that the speech deficit in PWS may extend beyond the VOT to affect the categorical perception of vowels.

Fundamental frequency (F0), the acoustic correlate of pitch, is used extensively to mark syllabic stress and intonation in non-tonal languages; it also is used systematically to distinguish word meanings in tonal languages (e.g., Cantonese). Studies on syllabic stress have found that stuttering is more likely to occur on stressed syllables than on unstressed syllables (Natke, Grosser, Sandrieser, & Kalveram, 2002; Natke, Sandrieser, van Ark, Pietrowsky, & Kalveram, 2004). However, relevant studies of speakers of tonal languages are rare. One study of Mandarin speakers indicated that stuttering-like disfluencies are more frequently associated with the Mandarin syllables that use tones 3 (low diping tone) and 4 (high falling tone) (Chou, Zebrowski, & Yang, 2015). The results of another study on Cantonese speakers seemed to suggest that the production of lexical tones was not affected by stuttering in AWS (Law et al., 2018). Although the above behavioural findings contradict each other, evidence from neuroimaging studies indicates that AWS and AWNS exhibit differences in the neural processing of rising and falling tones in Mandarin (Howell, Jiang, Peng, & Lu, 2012). The established relationship between syllabic stress and stuttering and the possible association between lexical tone and stuttering-like disfluencies have led to speculation that PWS may also have a detectable deficit in the categorical perception of lexical tones. Accordingly, we examined the categorical perception of vowels and lexical tones in addition to the VOT to test the hypothesis that stuttering influences the frequency/spectral/formant features of speech processing in the categorical perception paradigm.

Furthermore, some behavioural and neural evidence suggests that AWS may exhibit abnormal non-linguistic auditory processing (i.e., pure tones) compared with AWNS (Chang, Kenney, Loucks, & Ludlow, 2009; Hampton & Weber-Fox, 2008; Kaganovich et al., 2010). Therefore, non-linguistic auditory processing under a categorical perception task may also be compromised in AWS. To achieve a full understanding of the categorical perception deficit in PWS, we included pure tone stimuli as a non-linguistic condition in the current study, in addition to the three types of speech sound stimuli (i.e., consonant varying in VOT, vowel, and lexical tone).

Therefore, this study compared the performance of Cantonese-speaking AWS and AWNS in the categorical perception of four types of stimuli (i.e., VOT, vowel, lexical tone, and nonspeech tones) under backward noise masking. In light of previous findings (Bakhtiar et al., 2021; Bakhtiar et al., 2019; Basu et al., 2018; Howell & Williams, 2004), if the results of our previous studies (Bakhtiar et al. 2019, 2021) can be attributed to a lack of task demand, a more demanding listening condition should reveal subtle speech perception deficits. Thus, AWS should perform more poorly than control participants in the current study design (i.e., categorical perception with backward masking). Concerning the effect of stimulus type (i.e., VOT, vowel, lexical tone, and pure tone), VOT relies on temporal processing (i.e., duration), whereas vowel and lexical tone rely on spectral processing (i.e., frequency). If a perception deficit influences the ability of AWS to process frequencies/formants, this population also would be expected to perform more poorly than AWNS under the study conditions. Regarding non-linguistic pure tone perception, the perception of nonspeech sounds was found to be impaired in at least a subset of AWS (Hampton & Weber-Fox, 2008). Therefore, we predict that Cantonese-speaking AWS also exhibit a deficit in the perception of pure tone stimuli (analogues of lexical tones). Alternatively, if this deficit is mainly limited to the duration or temporal processing, we would expect AWS to show impairment in the perception of VOT but not of vowel, lexical tone, and pure tone.

Method

Participants

The study participants were 15 AWS (13 males; age range: 20–37 years; Mean (M) = 25.60; Standard Deviation (SD) = 4.75) and 15 controls (13 males; age range: 20–35 years; M = 25.27; SD = 4.38). The majority of AWS were recruited from the participant pool at the fluency clinic of the Hong Kong Polytechnic University, whereas the remaining AWS and AWNS were recruited using flyers posted at the university campus or through relevant social networks. The two groups had similar sociodemographic characteristics (e.g., age, gender, handedness, education level, and language profile). None of the participants reported having dyslexia or any neurological or psychological disorders. The participants' socio-demographic information is shown in Table 1. The study protocol was approved by the Human Subjects Ethics Committee of the Hong Kong Polytechnic University, and written consent was obtained from all of the participants before study commencement. All of the participants received remuneration for their participation at the end of the study.

<insert Table 1 about here>

Screening tests

Hearing test

Self-reported normal hearing was confirmed by hearing screening tests using a pure tone audiometer at frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz. Screening was initially conducted at an intensity of 25 dB, which was then was reduced by 10-dB steps to reach normal hearing acuity. All of the participants showed normal hearing acuity.

Fluency assessment

To assess stuttering severity, connected samples of the participants' conversational speech, story narration, and reading of a phonetically balanced passage (each containing at least 600 syllables) were video-recorded and analysed independently by two speech therapists. Stuttering severity was estimated by calculating the percentage of syllables stuttered (%SS), the average length of the three longest stuttering events, and the degree of physical concomitants based on the Stuttering Severity Index-3 (Riley, 1994). To determine the inter-rater agreement of the %SS measures, a third of the speech samples were independently evaluated by two other speech therapists at the Hong Kong Polytechnic University, and an intraclass correlation coefficient of 0.995 (95% confidence interval) was calculated, indicating high inter-rater reliability. The results for each stuttering participant are shown in Table 1. Among the 15 AWS, 11 exhibited mild stuttering, three had very mild stuttering, and only one participant exhibited severe stuttering.

Letter decision task

Although PWS have been shown to have a motor timing deficit (Olander, Smith, & Zelaznik, 2010), a more recent study with a larger sample suggested that a bimanual motor timing deficit is not a core characteristic of early developmental stuttering (Hilger, Zelaznik, & Smith, 2016). Nonetheless, to rule out the effect of a general motor deficit on the RT, both groups were tested using a letter decision task. The participants were instructed to press different mouse buttons in response to the different letter stimuli (e.g., the letters 'X' and 'O') shown on the screen. The results were not normally distributed and were subjected to a between-group comparison using the Mann–Whitney U test. No between-group differences in accuracy (U = 75, $p = .105$) and RT (U = 96, $p = .494$) were observed.

Stimuli

Four stimulus continua—consonants (VOT), vowels, lexical tone, and pure tone (nonspeech)—were generated following the methodology of Zhang et al. (2017) (see Figure 1). Three minimal pairs of Cantonese words were selected: /pa55/ (疤 'scar') versus /pha55/ (趴 'to lie down') for the VOT continuum; /fu55/ (膚 'skin') versus /fɔ 55/ (科 'section') for the vowel continuum; and /ji55/ (醫 'to treat/cure') xversus /ji25/ (椅 'chair') for the lexical tone continuum. The pure tone continuum consisted of nonspeech analogues of the lexical tone continuum, meaning that the pure tone stimuli carried the same F0 curves as the lexical tone stimuli. In the minimal word pairs described above, the lexical tone is annotated using Chao's tone letters, which range from 1 (lowest pitch) to 5 (highest pitch) (Chao, 1930). Each tone is described using two letters, which indicate the pitches at the beginning and end of a syllable. A male native Cantonese speaker was recorded while naturally reading the words aloud in isolation. Each word was repeated six times, and a clear token was selected for each pair to generate the stimulus continuum.

For the VOT continuum, the word /p^{h}a55/ was normalised to a mean intensity of 60 dB using Praat software (Boersma & Weenink, 2018) and then segmented into three parts: the burst release (~4.7 ms), aspiration (~36 ms), and vowel /a55/ (~420 ms). Using the overlap-add re-synthesis function in Praat, the aspiration was shortened proportionally to manipulate it between 0 and 36 ms in seven steps (i.e., ΔVOT = 6 ms). Each of the seven aspiration steps was concatenated with the preceding burst release and the following vowel, yielding seven stimuli with a varied VOT between /pa55/ and /p^{h}a55/.

For the vowel continuum, two words (/fu55/ 膚 'skin' and /fo55/ 科 'section') were normalised to a duration of 500 ms and mean intensity of 60 dB using Praat. The consonant /f/ and following vowel (i.e., /u/ or /o/) were then segmented out. For the vowels /u/ and /o/, 11 time points were measured in 10% intervals across the first formant (F1). The smallest F1 value in measurements for /u/ and the largest F1 value in measurements for /o/ were selected as the two end points of the F1 continuum, which was divided equally by Hz into seven steps (ΔF1 ≈ 42 Hz). For F2 to F4, we used the average frequencies of /u/ and /o/. Seven stimuli were generated by setting the frequencies of F1 to F4 in the seven steps to the prescribed values using Praat, with /u55/ as the basis of manipulation. The seven synthesised stimuli then were concatenated with the preceding consonant /f/ to yield seven stimuli varying between /fu55/ and /fo55/.

For the lexical tone continuum, the durations of the two selected words (/ji55/ 醫 'doctor' and /ji25/ 椅 'chair') was first normalised to 500 ms and their mean intensity to 60 dB using Praat. Next, the F0s were measured at 11 time points (in 10% intervals) across the entire duration of /ji55/ and /ji25/. The F0 distance between /ji55/ and /ji25/ at each time point was calculated and evenly divided into seven steps in semitones (e.g., $\Delta F0 \approx 0.74$ semitone at the onset of the stimulus, which decreased towards the end of the stimulus). Finally, the original F0 contour of the syllable /ji55/ was replaced with the seven equally distanced F0 contours using the overlap-add re-synthesis function in Praat, and a continuum of seven equally distanced pitch trajectories between the high-level tone and the high-rising tone was generated.

The pure tone continuum was composed of nonspeech analogues of the lexical tone continuum. A 500-ms pure tone sound was generated with a frequency of 145 Hz, which was close to the mean F0 of /ji55/. Studies have revealed that nonspeech stimuli have a softer sound than speech stimuli (Zhang, Peng, & Wang, 2012; Zhang, Shao, & Chen, 2018). To ensure that the nonspeech and speech stimuli were equally audible, the mean intensity of the nonspeech stimuli was manipulated to 75 dB, 15 dB higher than that of the speech stimuli. The seven equally distanced F0 contours in the lexical tone continuum were then extracted and superimposed onto the pure tone sound, generating a continuum of seven pure tone stimuli.

Finally, the noise stimulus (i.e., masker) was a broadband noise (i.e., 10 kHz) with a duration of 300 ms. To ensure its effectiveness, the intensity of the masker (80 dB) was louder than that of the stimuli.

<insert Figure 1 about here>

Procedure

Each stimulus continuum was presented in an identification task and a discrimination task using E-prime 2 software (Schneider, Eschman, & Zuccolotto, 2002).

Identification task

In the identification task, each stimulus continuum (i.e., consonant [VOT], vowel, lexical tone, and pure tone [nonspeech]) was presented in a separate block. Within each block, each target stimulus was repeated eight times, yielding 56 randomly ordered trials for a total of 224 trials (56 trials × 4 conditions). Within each trial, the target stimulus was presented, and the masker was played immediately after the target offset; thus, the interval between the target stimulus and the masker noise was 0 ms. The participants were instructed to identify the sound they heard by pressing the corresponding button on the Chronos response box: /pa55/ (疤 ‘scar’) or /p^{h}a55/ (趴 ‘to lie down’) in the consonant condition, /ji55/ (醫 ‘to treat/cure’) or /ji25/ (椅 ‘chair’) in the lexical tone condition and the pure tone (nonspeech) condition, and /fu55/ (膚 ‘skin’) or /fɔ55/ (科 ‘subject’) in the vowel condition. The participants were required to respond within 5 s; if no response was detected within 5 s, the task proceeded to the next trial automatically. The stimulus was presented to the participants binaurally through headphones. Each participant was given practice trials to familiarise them with the procedure before the formal test.

Discrimination task

In the discrimination task, each stimulus continuum was presented within a separate block. Seventeen pairs were created for discrimination per continuum, with seven matched pairs (i.e., stimuli pairs 1–1, 2–2, 3–3, 4–4, 5–5, 6–6, and 7–7), five different pairs separated by two steps in forward order (i.e., stimuli pairs 1–3, 2–4, 3–5, 4–6, and 5–7), and five pairs in backward order (i.e., stimuli pairs 3–1, 4–2, 5–3, 6–4, and 7–5). To control the experiment length, two sets of experiments were generated. Half of the participants in each group were assigned to set A, which contained seven same pairs and five different pairs in forward order, and the other half were assigned to set B, which contained seven same pairs and five different pairs in backward order. The same pairs were repeated five times, and the different pairs were repeated seven times, yielding 70 trials per set. Overall, 280 trials were conducted in the four blocks (70 trials × 4 conditions).

In each trial, the two target stimuli were both immediately followed by the masker at an interval of 0 ms. The end of the first masker and the beginning of the second target stimulus were separated by a fixed interval of 500 ms. The participants were instructed to discriminate whether the two target stimuli were the same or different by pressing buttons on the Chronos response box (i.e., the leftmost and second leftmost buttons to indicate decisions of 'same' and 'different, respectively) within 3 s. Each participant was given practice trials to help familiarise them with the task procedure.

In both tasks, the stimuli were presented binaurally through headphones to the participants in a soundproof room. The stimuli were presented at a listening level deemed comfortable to the subjects, and the volume level was kept constant within the experiments and across all subjects. The presentation order of the identification and discrimination tasks was counterbalanced, with half of the participants beginning with the identification task and the other half beginning with the discrimination task. Each task contained four blocks. In each task, the order of the four blocks was randomised. The task and block orders for the AWS participants were identical to those of the corresponding AWNS participants. A 15-minute break was given between the identification and discrimination tasks. Approximately 1.5 hours were needed to complete both tasks, including breaks. The accuracy and RTs were measured from the offset of the stimuli and recorded.

Data analysis

For the identification task, the boundary location and boundary width were first calculated for each participant in each stimulus continuum using probit analysis (Hallé, Chang, & Best, 2004). The boundary location indicates the position of a perceptual shift across two categories and is defined as the 50% crossover point in a continuum. The boundary width represents the steepness of the response shift across categorical boundaries and is defined as the distance in the stimulus step between 25% and 75% of the identification responses as determined by the probit analysis. For example, if 25% of stimulus 2 and 75% of stimulus 5 were identified as /p^ha55/ (趴 'to lie down') in the VOT continuum, the boundary width would be 3 (5 – 2 = 3).

For the discrimination task, the data were analysed using the sensitivity index d'. This index is computed for each participant as the z-score of the hit rate (i.e., 'different' responses to different pairs) minus that of the false alarm rate (i.e., 'different' responses to same pairs) for the pairs in each stimulus continuum. The pairs were then divided into between-category and within-category pairs for each participant based on the boundary location in each stimulus continuum obtained in the identification task. For example, if the boundary location was 3.5, the two-step pairs (i.e., 2-4 and 3-5) were assigned to the between-category group, and the remaining pairs were assigned to the within-category group. The d' score was then averaged from all the pairs that either spanned two categories or fell into one category for each stimulus continuum (Macmillan & Creelman, 2005).

For the analysis of each participant's RT in both the identification and discrimination tasks, trials with RTs below 50 ms were removed first; trials with RTs that were greater or less than 3 SDs of the mean per participant were also removed. The discard rate was 10.5% for the identification task (10.8% in AWS and 10.3% in AWNS) and 8.70% for the discrimination task

(8.23% in AWS and 9.08% in AWNS). The identification RTs were then classified into two groups—between-category and within-category—based on the boundary obtained from the probit analysis for each participant and each stimulus continuum. Similarly, the discrimination RT data were classified into two groups—between-category and within-category—according to each participant's boundary location.

Statistical analysis

Linear mixed-effects (LME) analyses were performed on the R 4.0.3 platform (R Core Team, 2020) using the *lme4* (Bates, Maechler, & Bolker, 2012), *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017), and *emmeans* packages (Lenth & Lenth, 2018). The *anova* function of the R package was used to obtain the *p* values of the main effects and the interactions in the final models by calculating Satterthwaite's approximation to degrees of freedom. The *emmeans* package was used to conduct pairwise comparisons with Tukey's correction.

Concerning the identification task, two LME models were constructed to compare the boundary location and boundary width across the groups (AWS and AWNS), stimulus types (consonant varying in VOT, vowel, lexical tone and nonspeech), and their two-way interaction (group × stimulus type). The model also included the random intercept of subjects.

For the identification RT analysis, LME models were used to compare the identification performance across the groups, stimulus types, and categories (between-category and within-category). The maximal model was first fitted using the above variables and their three-way interaction (group × stimulus type × category) as the fixed factors. The random intercepts of the stimuli and subjects, along with the random slope of groups per stimulus and the random slope of the interaction between categories and stimulus type per subject were treated as the random factors. Then, the random intercepts and slopes were removed one by one to reach a final simpler model using the likelihood ratio test (LRT) in R. For example, the LRT was used to compare a model without the random slope of groups per stimulus with the maximal model to determine whether the excluded factor was a contributing predictor. If the two models were not significantly different, the random slope was removed from the model. Finally, a model that used the random intercept of stimuli and the random slope of stimulus type per subject was selected, as this model did not differ significantly from the maximal model. In terms of the fixed effects, the same method described above was used to remove the least contributing predictors. For example, a model without the three-way interaction of group × stimulus type × category was compared with the maximal model. If this model was not significantly different from the maximal model, the three-way interaction was removed from the maximal model to achieve a simpler model. For the identification RTs, the final constructed model consisted of the random intercept of the stimuli, the random slope of the stimulus type per subjects, and three fixed factors (group, stimulus type, and category) and their two-way interactions (group × stimulus type, group × category, and stimulus type × category).

For the discrimination task analysis, LME analysis was conducted to compare the d' scores across the groups, stimulus types, and categories as defined above. A maximal LME model was constructed using these variables and their three-way interaction (group × stimulus type × category) and the random slope of the stimulus type per subject. A simpler model was constructed using the method described above. The final model consisted of the random intercept of subjects and three fixed factors (groups, stimulus type, and category) and their two-way interactions (group × stimulus type, group × category and stimulus type × category).

The discrimination RT analysis was conducted using the same protocol described above for the identification RT analysis. LME models were constructed to compare the RTs across the groups, stimulus types, and categories. The maximal model was first fitted using the above variables and their three-way interaction (group × stimulus type × category) as fixed factors and the random intercepts of stimuli and subjects, the random slope of groups per stimuli, and the random slope of categories by stimulus type per subject as random factors. The final model was achieved with the random intercept of the stimuli and the random slope of stimulus types per subjects, and three fixed factors including the groups, stimulus types and categories and their two-way interactions (i.e., groups × stimulus types , groups × categories and stimulus types × categories).

Results

Identification task

The identification curves for the four stimulus types in the two groups are shown in Figure 2. The boundary location and width for each group in each stimulus type are shown in Figure 3. The boundary location and width for each group that were averaged across four stimulus types are shown in Figure 4. Descriptive boundary location and width data are summarised in Table 2. Figure 5 displays the identification RT for different categories and stimulus types.

Concerning the boundary location, a main effect of *stimulus type* ($F(3, 84) = 21.961$, $p < .001$) and a significant two-way interaction between *groups* and *stimulus type* ($F(3, 84) = 3.256$, $p = .02$) were observed. Pairwise comparisons were conducted to explore the interaction between *group* and *stimulus type.* Within AWS, the value of the boundary location was significantly greater in the vowel continuum than in the other three conditions (*p*s < 0.001), implying that the response shift in AWS from /u/ to /o/ occurred later in the vowel continuum (i.e., closer to /o/). The differences between the other three stimulus types were not significant. Within AWNS, the values of the boundary locations were significantly higher in the VOT and vowel continua than in the nonspeech and lexical tone continua, indicating that the response shifts from /u/ to /o/ and /pa/ to /p^{h}a/ occurred later than the shifts in the other two stimulus types (ps < 0.05). Regarding the group effect, the AWS group had a significantly lower boundary location value than the AWNS group in the VOT condition (Estimate = -0.777, Std. Error = 0.223, t = -3.485, $p < .001$), indicating that the response shift from /pa/ to /p^{h}a/ in the VOT continuum occurred earlier in the AWS group than in the AWNS group. No other effects were significant.

Regarding the boundary width, significant main effects of *group* ($F(1, 28) = 8.834$, $p = .006$) and *stimulus type* ($F(3, 84) = 8.759$, $p < .001$) were observed. Although the effect of *group* (according to Figure 3) appeared to be more pronounced for the VOT and nonspeech stimuli than for the other stimuli, the overall interaction between *group* and *stimulus type* was not significant ($F(3, 84) = 2.024$, $p = .116$). The boundary width in the AWS group ($M = 1.329$, $SD = 0.882$) was significantly larger than that in the AWNS group ($M = 0.962$, $SD = 0.525$), indicating that the shift from one category to another was less abrupt and the materials were less categorically perceived in the AWS group. Regarding the effect of *stimulus type*, the boundary widths for the VOT and nonspeech continua were significantly larger than those for the vowel and lexical tone continua (ps < 0.01).

Regarding the identification RTs, the model revealed significant main effects of *category* ($F(1, 5890) = 13.258$, $p = .002$) and *stimulus type* ($F(1, 32) = 3.689$, $p = .02$). The interaction between *category* and *stimulus type* ($F(1, 5891) = 2.472$, $p = .05$) was also significant. Pairwise comparisons showed that the RTs for the between-category stimuli were significantly longer than those for the within-category stimuli in the nonspeech and VOT conditions (ps < 0.03). No other effects were significant.

<insert Figures 2 and 3 about here>

Discrimination task

The d’ scores of each stimulus continuum are shown in Figure 6. The interaction plots of the average d’ scores of the between- and within-category pairs in the four stimulus continua are shown in Figure 7. The discrimination RT is displayed in Figure 8. Regarding d’, significant effects of *group* ($F(1, 199) = 5.896$, $p = .02$) and *stimulus type* ($F(1, 199) = 14.450$, $p < .001$) were observed. The two-way interactions between *group* and *category* ($F(1, 199) = 8.826$, $p = .003$) and between *stimulus type* and *category* ($F(1, 199) = 8.640$, $p < .001$) were also significant.

Pairwise comparisons were first conducted to examine the interaction between *group* and *category*. The d’ score for between-category discrimination was significantly lower in the AWS group than in the AWNS group (Estimate = -0.776, Std. Error = 0.213, $t = -3.643$, $p < .001$), whereas no group difference was observed for within-category discrimination (Estimate = -0.097, Std. Error = 0.213, $t = -0.456$, $p = .650$). Both the AWS and AWNS groups showed higher d’ scores for between-boundary discrimination than for within-category discrimination (*p*s $< .001$). These results suggest that although both groups showed categorical patterns, the AWS group exhibited a significantly poorer performance in between-category discrimination than the AWNS group. This finding implied greater ambiguity in this type of discrimination in the AWS group, suggesting an impairment in categorical perception.

Second, pairwise comparisons were conducted to examine the interaction between *stimulus type* and *category*. The results revealed significantly higher d’ scores for between-category discrimination than for within-category discrimination across all four stimulus types (*p*s $< .001$) (see Figure 6). For the between-category pairs, the d’ score of the VOT condition was significantly lower than those of other three types (ps < .05), whereas the d' score of the vowel condition was significantly higher than those of the other three types (ps < .05); for the within-category pairs, the d' scores of the VOT and vowel conditions were significantly lower than those of the lexical tone and pure tone conditions (ps < .01).

In the discrimination RT analysis, a significant interaction was revealed between *stimulus type* and *group* ($F(3, 3787.3) = 2.927$, $p = .03$). However, pairwise comparisons did not reveal any difference between the AWS and AWNS groups across any stimulus type. For the AWS group, the RTs in the nonspeech condition were longer than the VOT condition (Estimate = -101.902, Std. Error = 47.307, $z = -2.154$, $p = .04$), whereas the differences were not significant in AWNS group.

<insert Table 2 and Figures 4 and 5 about here>

Additional analysis

As described above, participants were required to respond to a total of 280 trials in the discrimination task and a total of 224 trials in identification task. About 1.5 hours were required to complete the whole experiment. Previous studies have shown that PWS have weaker attention focus/persistence skills (Anderson & Wagovich, 2010; Ofoe et al., 2018). To test the possibility that attention efforts may have contributed to the degraded performance in our study, additional analysis was conducted on the discrimination sensitivity index d' for AWS group. The discrimination data was split into two subsets: first half and second half. The mean d' scores for the between-category pairs were calculated in each stimulus type for the two subsets of data. T-tests were then conducted to compare whether the d' scores were significantly different between the two subsets. The results revealed that the performance of AWS between the first and second halves of the discrimination task was not significantly different across all four stimulus types (*p*s > .2), suggesting that attention focus/persistence skills may have not affected the performance of AWS in current study. Detailed results for the statistical analysis are reported in the supplementary material.

Discussion

Auditory processing deficits have been hypothesised to play a role in stuttering (Bloodstein, 1995). Several studies have reported associations between stuttering and abnormal speech and non-speech auditory perception (Basu et al., 2018; Beal et al., 2011; Hampton & Weber-Fox, 2008; Howell et al., 2006; Jansson-Verkasalo et al., 2014). We previously reported comparable levels of categorical perception between Cantonese-speaking CWS and AWS and their typically fluent counterparts in terms of the identification and discrimination of phonemic categories when measured under quiet conditions; however, both CWS and AWS showed slower processing speeds in the identification and discrimination of phonemes located across categorical boundaries (Bakhtiar et al., 2019, 2021). With the current study, we increased the task demand by combining the categorical perception paradigm with backward masking. Backward masking increased the demand for mapping the incoming auditory stimuli onto their phonemic representations, a categorical perception deficit is expected to emerge in the AWS under such conditions. We further examined whether AWS would perform more poorly than AWNS in the categorical perception of different phonemic categories of linguistic (i.e., VOT, vowel and lexical tone) and non-linguistic stimuli (pure tone).

The results provide compelling evidence that AWS have impaired categorical perception under the backward masking condition. First, the AWS exhibited a larger boundary width than the AWNS, implying that their access to phonemic representations during identification might have been impaired. The boundary width is a robust index that distinguishes sharp from shallow categorical perception. For instance, speakers of tonal languages usually show a narrower boundary width in the categorical perception of lexical tones than speakers of non-tonal languages, and this difference is due to the former group's lifetime exposure to tones (Peng et al., 2010). During the identification process, stimulus categorisation involves mapping the acoustic signals onto stored phonemic representations (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Wiener, Turkeltaub, & Coslett, 2010). Therefore, our observation of a larger boundary width in the AWS than in the AWNS suggests that the mapping from acoustic signals to phonemic representations was impaired in the former group.

Second, the AWS group exhibited significantly lower discrimination accuracy in the between-category stimulus pairs than the AWNS group, whereas both groups showed comparable accuracy in discriminating the within-category pairs. This result provides further evidence of impaired categorical perception in AWS. The discrimination task involved the comparison of two presented auditory signals, which was possibly mediated by a comparison of phonological features (Rauschecker & Scott, 2009; Turkeltaub & Branch Coslett, 2010). The lower discrimination sensitivity of AWS, especially in the between-category pairs, suggested that these participants found it difficult to distinguish paired sounds that belonged to two different phonemic categories. These findings, combined with the larger boundary width, suggest that AWS might have impaired access to phonemic representations.

Consistent with findings from other studies (Bakhtiar et al., 2021; Bakhtiar et al., 2019; Basu et al., 2018; Lotfi, Dastgerdi, Farazi, Moossavi, & Bakhshi, 2020; Neef et al., 2012), the current results provide empirical evidence that PWS may have speech perception deficits. This finding supports the Directions Into Velocities of Articulators (DIVA) model (Guenther, 2006;

Guenther, Ghosh, & Tourville, 2006) and Gradient Order DIVA (GODIVA) model (Bohland, Bullock, & Guenther, 2010). According to the DIVA model, some speech errors might originate during activation of the speech sound map in the left ventromedial premotor cortex, which sends inputs to the auditory and somatosensory error maps. In turn, the auditory and somatosensory error maps deliver inputs in the form of feedback commands to the articulatory velocity and position maps in the motor cortex. The GODIVA model extends the DIVA model by incorporating more brain regions, such as the posterior inferior frontal sulcus and basal ganglia, to account for the planning, timing, and coordination of multisyllabic speech sequences (Bohland et al., 2010). Drawing on the motor loops described by the DIVA and GODIVA models, we speculate that the speech perception deficit observed in the current study may contribute to the altered speech kinematics and speech disfluency in PWS.

Nonetheless, a robust deficit in categorical perception was only observed under the backward masking condition. Backward masking reflects the central auditory phenomenon (Buss et al., 1999; Elliott, 1962; Howell, Rosen, Hannigan, & Rustin, 2000; Howell et al., 2006). Interruption theory, a relevant model, indicates that backward masking leads to the perception that the leading stimulus is terminated or disrupted by the masker (Breitmeyer & Ogmen, 2000; Turvey, 1973). The masker ‘interrupts’ or erases the memory of the signal by interrupting or terminating the processing of the target stimulus from initial sensory memory storage to short-term memory storage. This interruption leads to degraded performance under the backward masking condition. In this study, we hypothesised that backward masking interrupts the sensory processing of the target stimulus, thus increasing the difficulty of mapping sensory information to stored phonological representations.

Regarding the effect of stimulus type, in addition to the impaired perception of VOT reported by Neef et al. (2012), we found that AWS showed impaired perception of both frequency/spectral-based stimuli (e.g., vowels and lexical tones) and non-linguistic stimuli (e.g., pure tones). Furthermore, although AWS had a smaller boundary location than AWNS in the VOT condition only, AWS had a larger boundary width across all stimulus types, implying that this group had less abrupt response shifts across all conditions. Analysis of the d' score, a measure of perceptual sensitivity, also suggested that AWS performed significantly worse than AWNS in between-category discrimination across all stimulus types. The boundary width and d' score are two crucial indices of categorical perception. Typical categorical perception is thought to demonstrate an abrupt response shift across the categorical boundary in the identification task (as indexed by the boundary width) and more discernible discrimination of stimulus pairs that cross the categorical boundary in the discrimination task (Repp, 1984). In contrast, the boundary location may not be a direct index of the robustness of categorical perception (Zhang et al., 2017); rather, it might suggest perceptual bias towards the aspirated sound (/p^ha/) in AWS. Therefore, we conclude that AWS might exhibit impaired categorical perception across all stimulus types, which is not confined to the VOT condition only.

In addition to these linguistic elements, our findings also suggest that AWS exhibit impaired processing of non-linguistic stimuli. This finding somewhat diverges from that of Corbera et al. (2005) but is consistent with that reported by Hampton and Weber-Fox (2008). Corbera et al. (2005) revealed that AWS and control subjects showed similar MMN amplitudes in response to simple sound contrasts (frequency-based pure tones), indicating that the perception of simple sound features and their representations in neural traces within the auditory cortex were preserved in AWS. However, other studies have demonstrated reduced MMN in response to pure tones in at least a subset of AWS, indicating for some PWS, a disruption in auditory signal encoding may contribute to reduced efficiency in processing and monitoring auditory feedback, with potential effects on fluency (Hampton & Weber-Fox, 2008). However, our results regarding nonspeech stimuli must be interpreted with caution. First, the pure tone stimuli are analogues of the lexical tone stimuli and thus are highly acoustically similar to lexical tones. Furthermore, during the experiment, the participants were required to identify the pure tones as either high level (/ji55/, 'to treat/cure') or high rising tones (/ji25/, 'chair') in Cantonese during the identification task. This requirement may have enhanced their access of phonological knowledge to some extent. The discrimination data also showed higher d' scores for between-category discrimination than for within-category discrimination across all stimulus types, including the nonspeech stimuli, which implied that the nonspeech stimuli may have undergone categorical processing to some extent. Future studies using nonspeech stimuli with less acoustic similarity to the linguistic stimuli and a different task are required to distinguish PWS' impairment in the perception of non-linguistic versus linguistic processing.

Although some compelling results are reported herein, the behavioural design of the study cannot reveal anything about the neural mechanisms that underly categorical perception. Further neuroimaging studies may be needed to explore these mechanisms. In addition, the majority of the AWS in this study had mild stuttering severity. Future studies could include more participants with a wider range of stuttering severity. Additionally, our stimuli in the VOT and lexical tone conditions also had a lexical component (i.e., noun vs. verb), which may have influenced categorical perception, especially in the identification task. Future research could use pseudowords to avoid the possible effect of lexicality on categorical perception. The experimental design of future studies could also be improved by measuring the RT from the onset of each stimulus to capture the real-time nature of the response more efficiently. Lastly, although our findings suggest that AWS may have deficits in both speech and nonspeech perception, our paradigm incorporates backward masking, which is thought to reflect central auditory perception. For this reason, we cannot separate deficits in speech perception from those in central auditory perception using the available data. Further research intended to address auditory perception and speech perception directly is needed to better understand this matter.

Conclusion

Taken together, the findings of this study provide robust behavioural evidence that AWS exhibit impaired categorical perception of both speech and nonspeech sounds, supporting the perspective that reduced access to phonemic representations might be associated with speech disfluency in this population. However, this deficit was only evident under the backward masking condition, suggesting that the observed deficit in accessing phonemic representations may only manifest under more demanding listening conditions. Furthermore, the impaired categorical perception observed in AWS was not confined to VOT (i.e., temporal processing) but also influenced the perception of frequency/spectral aspects of speech and of non-linguistic stimuli. The results of this study highlight the roles of speech and nonspeech perception in stuttering, suggesting that this disorder may be a manifestation of a more complex dysfunction. In other words, stuttering may interfere with not only fluent speech production but also auditory and speech perception in a more complex manner.

Acknowledgements: We thank Man Na Cheung, a former speech therapy student, for assistance with the data collection and fluency assessments. We also thank all of the participants in this study.

References

Anderson, J. D., & Wagovich, S. A. (2010). Relationships among linguistic processing speed, phonological working memory, and attention in children who stutter. *Journal of Fluency Disorders*, *35*(3), 216–234. https://doi.org/10.1016/j.jfludis.2010.04.003

Bakhtiar, M., Shao, J., Cheung, M. N., & Zhang, C. (2020). Categorical perception of speech sounds in adults who stutter. *Clinical Linguistics & Phonetics*, 1–17. https://doi.org/10.1080/02699206.2020.1803407

Bakhtiar, M., Zhang, C., & Ki, S. S. (2019). Impaired processing speed in categorical perception: Speech perception of children who stutter. *PLoS ONE*, *14*(4), 1–18. https://doi.org/10.1371/journal.pone.0216124

Basu, S., Schlauch, R. S., & Sasisekaran, J. (2018). Backward masking of tones and speech in people who do and do not stutter. *Journal of Fluency Disorders*, *57*, 11–21. https://doi.org/10.1016/j.jfludis.2018.07.001

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42. 2011.

Beal, D S, Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & De Nil, L. F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. *NeuroImage*, *52*(4), 1645–1653. https://doi.org/10.1016/j.neuroimage.2010.04.277

Beal, D S, Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & De Nil, L. F. (2011).

Speech-induced suppression of evoked auditory fields in children who stutter. *NeuroImage*, *54*(4), 2994–3003. https://doi.org/10.1016/j.neuroimage.2010.11.026

Beal, Deryk S., Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & De Nil, L. F. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. *NeuroImage*, *54*(4), 2994–3003. https://doi.org/10.1016/j.neuroimage.2010.11.026

Bent, T., & Holt, R. F. (2013). The influence of talker and foreign-accent variability on spoken word identification. *The Journal of the Acoustical Society of America*, *133*(3), 1677–1686. https://doi.org/10.1121/1.4776212

Bloodstein, O. (1995). *A Handbook on Stuttering, Ed 5*. San Diego, CA: Singular Publishing Group.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program], 2018.

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529. https://doi.org/10.1162/jocn.2009.21306

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception and Psychophysics*, *62*(8), 1572–1595. https://doi.org/10.3758/BF03212157

Buss, E., Hall, J. W., Grose, J. H., & Dev, M. B. (1999). Development of adult-like performance in backward, simultaneous, and forward masking. *Journal of Speech, Language, and Hearing Research*, *42*(4), 844–849. https://doi.org/10.1044/jslhr.4204.844

Chang, S.-E., Kenney, M. K., Loucks, T. M. J., & Ludlow, C. L. (2009). Brain activation abnormalities during speech and non-speech in stuttering speakers. *NeuroImage*, *46*(1), 201–212. https://doi.org/10.1016/j.neuroimage.2009.01.066

Chao, Y.-R. (1930). A system of tone letters. *Le Maître Phonétique*, *45*, 24–27.

Chou, F.-C., Zebrowski, P., & Yang, S.-L. (2015). Lexical tone and stuttering loci in Mandarin: Evidence from preschool children who stutter. *Clinical Linguistics \& Phonetics*, *29*(2), 115–130.

Corbera, S., Corral, M.-J., Escera, C., & Idiazábal, M. A. (2005). Abnormal speech sound representation in persistent developmental stuttering. *Neurology*, *65*(8), 1246–1252. https://doi.org/10.1212/01.wnl.0000180969.03719.81

Elliott, L. L. (1962). Backward masking: Monotic and dichotic conditions. *The Journal of the Acoustical Society of America*, *34*(8), 1108–1115.

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2018). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *Journal of Fluency Disorders*, *55*, 6–45. https://doi.org/10.1016/j.jfludis.2017.03.007

Goncalves, I. C., Andrade, C. R. F. de, & Matas, C. G. (2015). Auditory processing of speech and non-speech stimuli in children who stutter: electrophysiological evidences. *Brain Disord Ther*, *4*(5), 199.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Guitar, B. (2013). *Stuttering: An integrated approach to its nature and treatment*. Lippincott

Williams \& Wilkins.

Hall, J. W., & Jerger, J. (1978). Central auditory function in stutterers. *Journal of Speech and Hearing Research*, *21*(2), 324–337.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

Hampton, A., & Weber-Fox, C. (2008). Non-linguistic auditory processing in stuttering: Evidence from behavior and event-related brain potentials. *Journal of Fluency Disorders*, *33*(4), 253–273. https://doi.org/10.1016/j.jfludis.2008.08.001

Harnad, S. R. (1987). *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

Hilger, A. I., Zelaznik, H., & Smith, A. (2016). Evidence that bimanual motor timing performance is not a significant factor in developmental stuttering. *Journal of Speech, Language, and Hearing Research*, *59*(4), 674–685. https://doi.org/10.1044/2016_JSLHR-S-15-0172

Hill, P. R., Hogben, J. H., & Bishop, D. M. V. (2006). Auditory Frequency Discrimination in Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1136–1146. https://doi.org/10.1044/1092-4388(2005/080)

Hill, Penelope R., Hartley, D. E. H., Glasberg, B. R., Moore, B. C. J., & Moore, D. R. (2006). Auditory Processing Efficiency and Temporal Resolution in Children and Adults. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1022–1029. https://doi.org/10.1044/1092-4388(2004/076)

Howell, P, Rosen, S., Hannigan, G., & Rustin, L. (2000). Auditory backward-masking performance by children who stutter and its relation to dysfluency rate. *Perceptual and Motor Skills*, *90*(2), 355–363. https://doi.org/10.2466/pms.2000.90.2.355

Howell, P, & Williams, S. M. (2004). Development of auditory sensitivity in children who stutter and fluent children. *Ear and Hearing*, *25*(3), 265–274. https://doi.org/10.1097/01.AUD.0000130798.50938.EB

Howell, Peter, Davis, S., & Williams, S. M. (2006). Auditory abilities of speakers who persisted, or recovered, from stuttering. *Journal of Fluency Disorders*, *31*(4), 257–270. https://doi.org/10.1016/j.jfludis.2006.07.001

Howell, Peter, Jiang, J., Peng, D., & Lu, C. (2012). Neural control of rising and falling tones in Mandarin speakers who stutter. *Brain and Language*, *123*(3), 211–221.

Jansson-Verkasalo, E., Eggers, K., Järvenpää, A., Suominen, K., den Bergh, B., De Nil, L., & Kujala, T. (2014). Atypical central auditory speech-sound discrimination in children who stutter as indexed by the mismatch negativity. *Journal of Fluency Disorders*, *41*(C), 1–11. https://doi.org/10.1016/j.jfludis.2014.07.001

Kaganovich, N., Wray, A. H., & Weber-Fox, C. (2010). Non-linguistic auditory processing and working memory update in pre-school children who stutter: An electrophysiological study. *Developmental Neuropsychology*, *35*(6), 712–736. https://doi.org/10.1080/87565641.2010.508549

Kramer, M. B., Green, D., & Guitar, B. (1987). A comparison of stutterers and nonstutterers on masking level differences and synthetic sentence identification tasks. *Journal of Communication Disorders*, *20*(5), 379–390. https://doi.org/10.1016/0021-9924(87)90026-8

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & others. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th ed.). Boston: Cengage Learning.

Law, T., Packman, A., Onslow, M., To, C. K.-S., Tong, M. C.-F., & Lee, K. Y.-S. (2018). Lexical tone and stuttering in Cantonese. *Clinical Linguistics \& Phonetics*, *32*(4), 285–297.

Lenth, R., & Lenth, M. R. (2018). Package ‘lsmeans.’ *The American Statistician*, *34*(4), 216–221.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*. https://doi.org/10.1037/h0020279

Liotti Mario, M., Ingham, J. C., Takai, O., Paskos, D. K., Perez, R., & Ingham, R. J. (2010). Spatiotemporal dynamics of speech sound perception in chronic developmental stuttering. *Brain and Language*, *115*(2), 141–147. https://doi.org/10.1016/j.bandl.2010.07.007

Lotfi, Y., Dastgerdi, Z. H., Farazi, M., Moossavi, A., & Bakhshi, E. (2020). Auditory temporal processing assessment in children with developmental stuttering. *International Journal of Pediatric Otorhinolaryngology*, *132*(February). https://doi.org/10.1016/j.ijporl.2020.109935

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user’s guide*. Mahwah: Lawrence Erlbaum Associates.

Massaro, D. W., & Simpson, J. A. (2014). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Psychology Press.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The Essential Role of Premotor Cortex in Speech Perception. *Current Biology*, *17*(19), 1692–1696. https://doi.org/10.1016/j.cub.2007.08.064

Natke, U., Grosser, J., Sandrieser, P., & Kalveram, K. T. (2002). The duration component of the stress effect in stuttering. *Journal of Fluency Disorders, 27*(4), 305–318.

Natke, U., Sandrieser, P., van Ark, M., Pietrowsky, R., & Kalveram, K. T. (2004). Linguistic stress, within-word position, and grammatical class in relation to early childhood stuttering. *Journal of Fluency Disorders*, *29*(2), 109–122.

Neef, N. E., Sommer, M., Neef, A., Paulus, W., von Gudenberg, A. W., Jung, K., & Wüstenberg, T. (2012). Reduced speech perceptual acuity for stop consonants in individuals who stutter. *Journal of Speech, Language, and Hearing Research*, *55*(1), 276–289. https://doi.org/10.1044/1092-4388(2011/10-0224)

Ofoe, L. C., Anderson, J. D., & Ntourou, K. (2018). Short-term memory, inhibition, and attention in developmental stuttering: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1626–1648.

Olander, L., Smith, A., & Zelaznik, H. N. (2010). Evidence that a motor timing deficit is a factor in the development of stuttering. *Journal of Speech, Language, and Hearing Research*, *53*(4), 876–886. https://doi.org/10.1044/1092-4388(2009/09-0007)

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

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

Porter, H. L., Spitzer, E. R., Buss, E., Leibold, L. J., & Grose, J. H. (2018). Forward and backward masking of consonants in school-age children and adults. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1807–1814. https://doi.org/10.1044/2018_JSLHR-H-17-0403

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In *Speech and language: Advances in basic research and practice* (Vol. 10, pp. 243–335). New York: Cambridge University Press.

Riley, G. D. (1994). No Title. *Stuttering Severity Instrument for Children and Adults (3rd Ed.)*.

Rosen, S., Adlar, A., & Van Der Lely, H. K. J. (2009). Backward and simultaneous masking in children with grammatical specific language impairment: no simple link between auditory and language abilities. *Journal of Speech, Language, and Hearing Research*, *52*(2), 396–411. https://doi.org/10.1044/1092-4388(2009/08-0114)

Rosen, S., & Manganari, E. (2001). Is There a Relationship between Speech and Nonspeech Auditory Processing in Children with Dyslexia? *Journal of Speech, Language, and Hearing Research*, *44*(1–4), 720–736. https://doi.org/10.1044/1092-4388(2001/057)

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime (version 2.0). *Computer Software and Manual]. Pittsburgh, PA: Psychology Software Tools Inc*.

Teesson, K., Packman, A., & Onslow, M. (2003). The Lidcombe Behavioral Data Language of stuttering. *Journal of Speech, Language, and Hearing Research : JSLHR*, *46*(4), 1009–1015. https://doi.org/10.1044/1092-4388(2003/078)

Toscher, M. M., & Rupp, R. R. (1978). A study of the central auditory processes in stutterers using the synthetic sentence identification (SSI) test battery. *Journal of Speech and Hearing Research*, *21*(4), 779–792.

Turkeltaub, P. E., & Branch Coslett, H. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*(1), 1–15. https://doi.org/10.1016/j.bandl.2010.03.008

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*(1), 1.

Wiener, M., Turkeltaub, P., & Coslett, H. B. (2010). The image of time: A voxel-wise meta-analysis. *NeuroImage*, *49*(2), 1728–1740. https://doi.org/10.1016/j.neuroimage.2009.09.064

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, *38*(2), 66–87. https://doi.org/10.1016/j.jfludis.2012.11.002

Zhang, C, Shao, J., & Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PLoS ONE*, *12*(8). https://doi.org/10.1371/journal.pone.0183151

Zhang, Caicai, Peng, G., & Wang, W. S.-Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *Journal of the Acoustical Society of America*, *132*(2), 1088. https://doi.org/10.1121/1.4731470

Zhang, Caicai, Shao, J., & Chen, S. (2018). Impaired perceptual normalization of lexical tones in Cantonese-speaking congenital amusics. *The Journal of the Acoustical Society of America*, *144*(2), 634–647.

Zhang, Caicai, Shao, J., & Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PloS One*, *12*(8), e0183151.

Figure Captions:

Figure 1. Schematic diagram of stimulus continua divided into seven stimuli (adopted from Zhang, Shao, & Huang, 2017). (A): VOT continuum. (B): Lexical tone continuum. (C): Vowel continuum.

Figure 2. Identification curves for the AWS and AWNS groups across the four stimulus continua in the identification task. Top-left panel: Rate of /pa55/ (疤 'scar') responses in the VOT continuum. Top-right panel: Rate of /fu55/ (膚 'skin') responses in the vowel continuum. Bottom-left panel: Rate of /ji55/ (醫 'doctor') responses in the lexical tone continuum. Bottom-right panel: Rate of high-level pitch responses in the nonspeech continuum.

Figure 3. Comparison of boundary location and boundary width between the AWS group and AWNS group in the four stimulus continua in the identification task.

Figure 4. Boundary location and boundary width in the AWS group and AWNS group averaged across the four stimulus continua in the identification task.

Figure 5. Plots for the identification RTs of the AWS and AWNS groups across the categorical boundaries (left) and four stimulus types (right).

Figure 6. The d' scores of each stimulus continuum in the discrimination task for the AWS and AWNS groups.

Figure 7. Interaction plot of the average d' scores of between- and within-category pairs for the AWS and AWNS participants (left) and of the four stimulus continua in the discrimination task (right).

Figure 8. Plots for the discrimination RTs of the AWS and AWNS groups across the categorical boundaries (left) and four stimulus types (right).

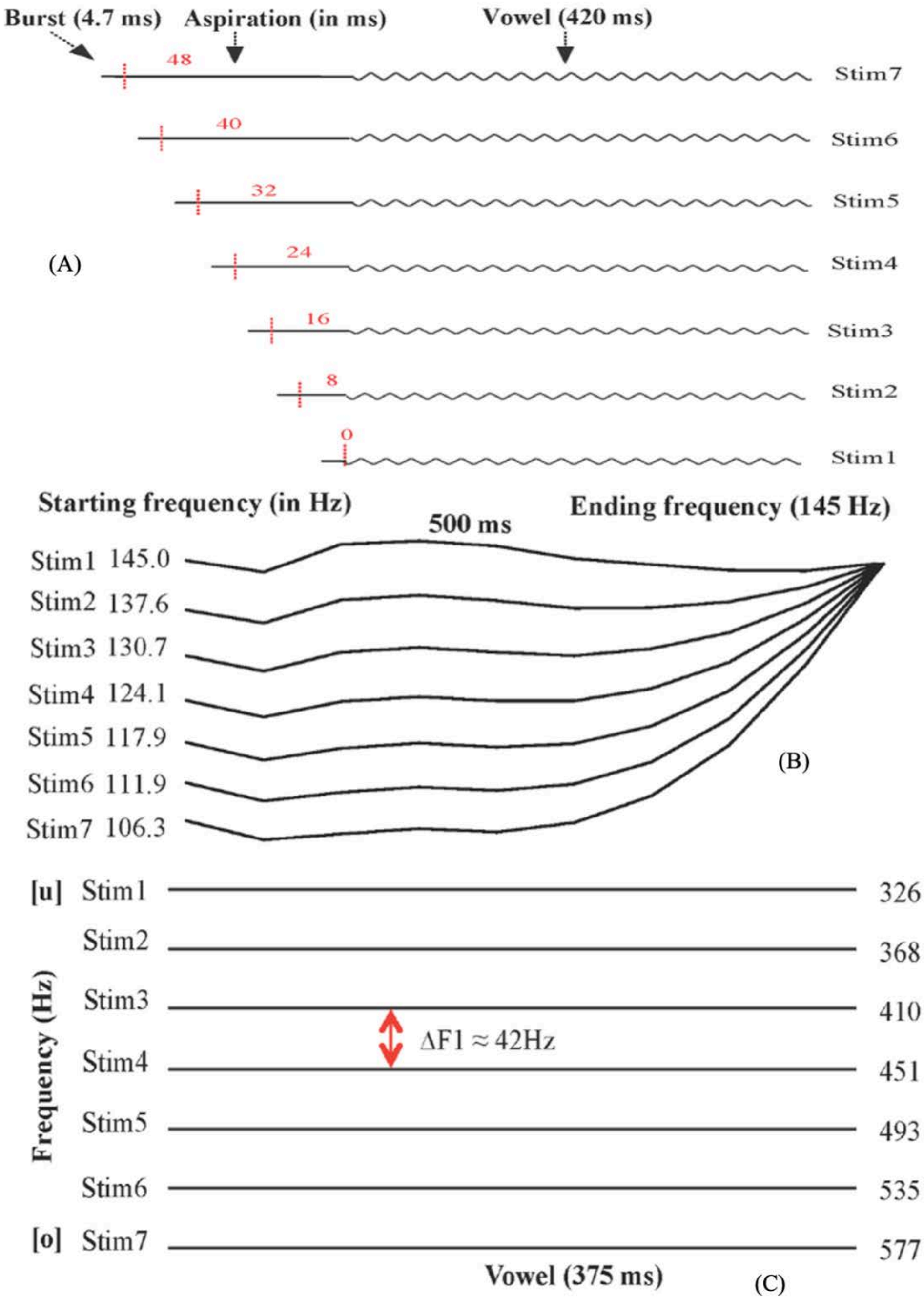
Table 1. Demographic information for the AWS group and the AWNS group

Participants	Age	Gender	Handedness	Education	SSI-3 TOS	SSI-3 Severity
AWS1	21	M	Right	HD	20	Mild
AWS2	21	M	Right	BA	22	Mild
AWS3	23	M	Right	BA	22	Mild
AWS4	20	M	Right	BA	14	Mild
AWS5	24	M	Right	BA	16	Mild
AWS6	25	M	Right	BA	16	Mild
AWS7	27	M	Right	BA	16	Mild
AWS8	34	M	Right	BA	11	Very Mild
AWS9	37	F	Right	BA	15	Mild
AWS10	25	M	Right	BA	10	Very Mild
AWS11	29	M	Right	BA	13	Mild
AWS12	22	M	Right	BA	18	Mild
AWS13	27	F	Right	BA	10	Very Mild
AWS14	24	M	Right	BA	19	Mild
AWS15	25	M	Right	BA	33	Severe
AWNS1	20	M	Right	BA	4	-
AWNS2	22	M	Right	BA	2	-
AWNS3	22	M	Right	BA	4	-
AWNS4	20	M	Right	BA	6	-
AWNS5	25	M	Right	BA	6	-
AWNS6	25	M	Right	BA	2	-
AWNS7	26	M	Right	BA	8	-
AWNS8	32	M	Right	BA	6	-
AWNS9	35	F	Right	BA	4	-
AWNS10	25	M	Right	BA	2	-
AWNS11	30	M	Right	BA	2	-
AWNS12	20	M	Right	BA	2	-
AWNS13	26	F	Right	BA	4	-
AWNS14	25	M	Right	BA	4	-
AWNS15	26	M	Right	BA	2	-

Note: HD = higher diploma; BA = bachelor's degree; SSI-3 = Stuttering Severity Instrument-3; TOS = total overall scores.

Table 2. Descriptive data for the boundary position and boundary width in the identification task.

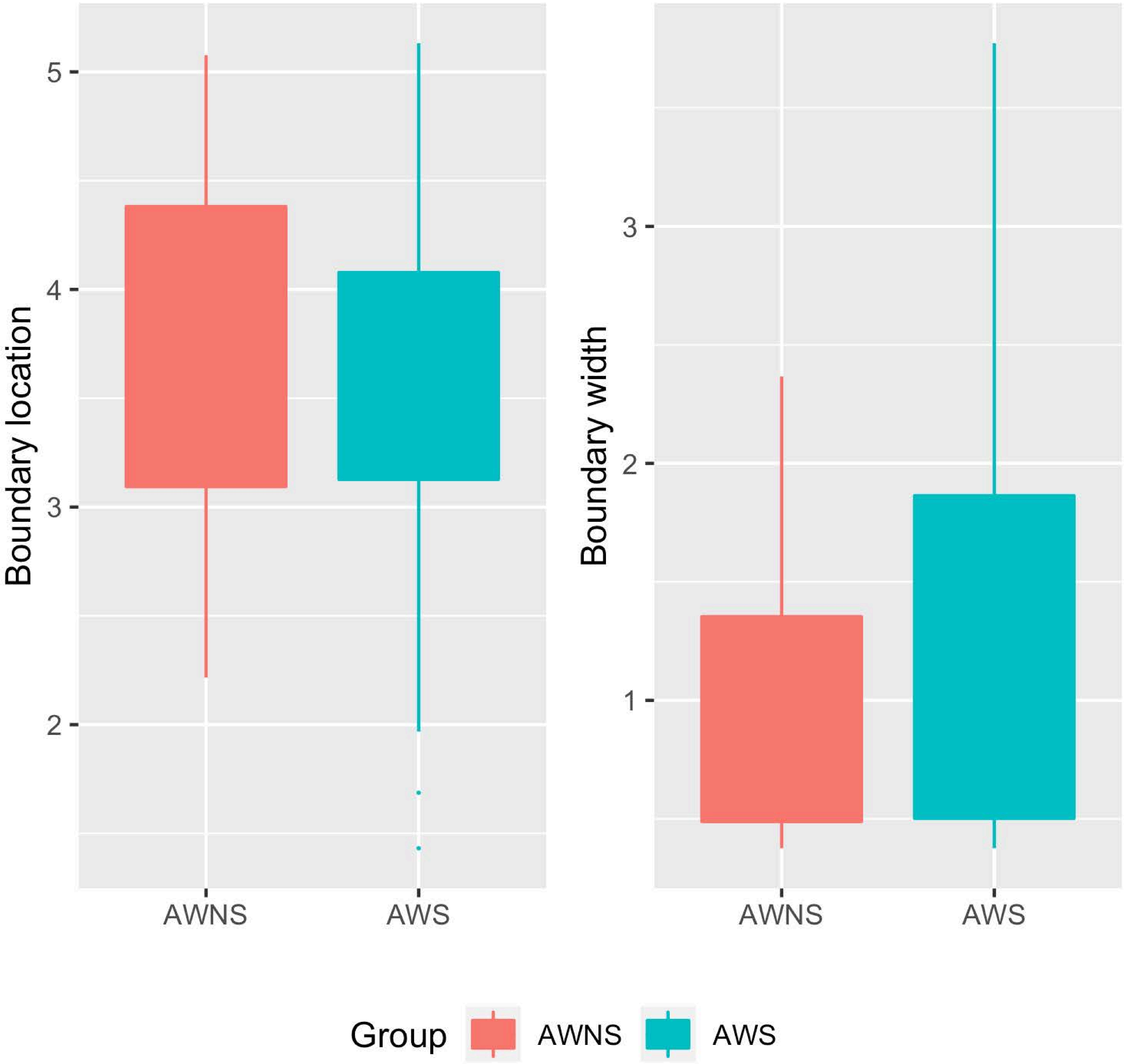
Group	**Condition**	**Boundary Position**			**Boundary Width**		
		Mean	SD	Range	Mean	SD	Range
AWS	VOT	3.208	0.982	1.432-5.132	1.826	1.053	0.453-3.773
AWNS	VOT	3.985	0.719	2.216-5.077	1.212	0.404	0.582-1.814
AWS	Vowel	4.368	0.386	3.572-4.893	0.864	0.498	0.489-1.747
AWNS	Vowel	4.379	0.445	2.912-4.893	0.833	0.464	0.464-1.552
AWS	Lexical tone	3.316	0.527	2.351-4.500	0.905	0.828	0.376-3.379
AWNS	Lexical tone	3.393	0.565	2.261-4.340	0.779	0.346	0.380-1.358
AWS	Nonspeech	3.274	0.574	1.687-4.049	1.722	0.616	0.418-2.425
AWNS	Nonspeech	3.180	0.474	2.519-4.006	1.022	0.734	0.418-2.425
AWS	Overall	3.542	0.801	1.432-5.132	1.329	0.882	0.376-3.773
AWNS	Overall	3.734	0.726	2.216-5.077	0.961	0.525	0.380-2.367



Burst (4.7 ms)
Aspiration (in ms)
Vowel (420 ms)
48
Stim7
40
Stim6
32
Stim5
(A)
24
Stim4
16
Stim3
8
Stim2
0
Stim1
Starting frequency (in Hz)
Ending frequency (145 Hz)
500 ms
Stim1 145.0
Stim2 137.6
Stim3 130.7
Stim4 124.1
Stim5 117.9
Stim6 111.9
Stim7 106.3
(B)
[u] Stim1
326
Stim2
368
Stim3
410
ΔF1 ≈ 42Hz
Stim4
451
Stim5
493
Stim6
535
[o] Stim7
577
Frequency (Hz)
Vowel (375 ms)
(C)

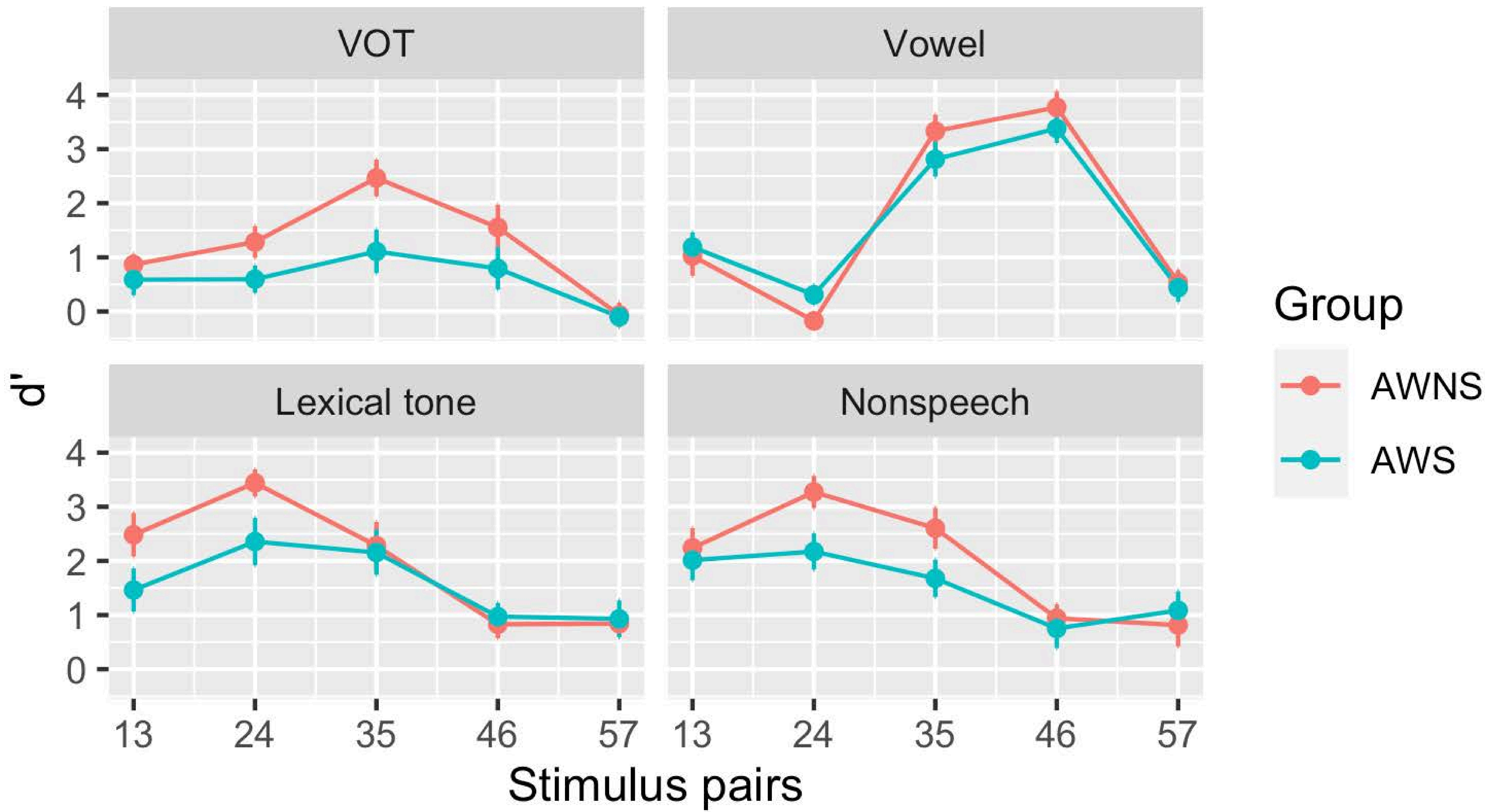
VOT
Vowel
Lexical tone
Nonspeech
Identification curves
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
Stimulus numbers
Group
AWNS
AWS


Boundary location
5
4
3
2
VOT
Vowel
Lexical tone
Nonspeech
Stimulus types
Boundary width
3
2
1
VOT
Vowel
Lexical tone
Nonspeech
Stimulus types
Group
AWNS
AWS



Boundary location
5
4
3
2
AWNS
AWS
Boundary width
3
2
1
AWNS
AWS
Group
AWNS
AWS

Identification RT (ms)
2000
1000
0
Between
Within
Category
Identification RT (ms)
2000
1000
0
VOT
Vowel
Lexical tone
Nonspeech
Stimulus types
Group
AWNS
AWS



VOT
Vowel
Lexical tone
Nonspeech
d'
0
1
2
3
4
13
24
35
46
57
Stimulus pairs
Group
AWNS
AWS


d'
4
2
0
-2
Between
Within
Category
VOT
Vowel
Lexical tone
Nonspeech
Stimulus types
Group
AWNS
AWS

Discrimination RT (ms)
0
1000
2000
Between
Within
Category
Discrimination RT (ms)
0
1000
2000
VOT
Vowel
Lexical tone
Nonspeech
Stimulus types
Group
AWNS
AWS

Table 1. T-test results comparing the AWS's performance in the first half and second half of the discrimination experiments.

Condition	Mean d' -first half	Mean d' - second half	*t*	*p*
VOT	2.07	1.91	0.508	0.613
Vowel	3.26	3.61	-1.12	0.267
Lexical tone	2.79	2.71	0.228	0.820
Nonspeech	2.71	2.92	-0.582	0.563