Categorical Perception of the Speech Sounds in Adults who Stutter

Mehdi Bakhtiar[a]*, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
Shao Jing[b], School of Humanities, Shanghai Jiao Tong University.
Man Na Cheung[a], Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
Caicai Zhang[a], Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

* Corresponding authors:
Mehdi Bakhtiar, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.
Email: m.bakhtiar@polyu.edu.hk

Received 27 Apr 2020,

Accepted 27 Jul 2020,

Abstract

Stuttering is often attributed to the impaired speech production system, however, there is growing evidence implicating issues in speech perception. Our previous research (Bakhtiar, Zhang, & Ki, 2019) showed that children who stutter have similar patterns but slower categorical perception (i.e. the ability to categorise different acoustic variations of the speech sounds into the same or different phonemic categories) compared to the children who do not stutter.

This study aimed to extend our previous research to adults who stutter (AWS) using the same categorical perception paradigm. Fifteen AWS and 15 adults who do not stutter (AWNS) were recruited to complete identification and discrimination tasks involving acoustic variations of Cantonese speech sounds in four stimulus contexts: consonants (varying in voice onset times, VOTs), lexical tones, vowels and pure tones. The results showed similar categorical perception between the two groups in terms of the boundary position and width in the identification task and between-category benefits in the discrimination task. However, there were some trends for lower discrimination accuracy (overall d' scores) and slower discrimination of the between-category stimuli versus within-category stimuli for AWS than AWNS. These results partially confirm our previous finding on children in terms of a comparable pattern of categorical perception between the two groups, but slower processing speed to access the phonemic representations in speech perception among AWS than AWNS.

Keywords: Stuttering, Speech perception, Categorical perception, Identification, Discrimination

Introduction

For decades, several theories from different disciplines attempt to explain the nature of persistent stuttering. Though stuttering is often attributed to impaired speech motor control, there is growing evidence supporting the presence of issues in speech perception skills (Corbera, Corral, Escera, & Idiazabal, 2005; Halag-Milo et al., 2016; Neef et al., 2012). Several studies have also found a relation between stuttering and some acoustic measures related to speech production such as the longer voice onset times (VOTs), stop gap durations, vowel durations, and durations of consonant–vowel transition in certain phonetic context (for a review see Bloodstein, 1995). Evidence from neuroimaging studies showed reduced neural activation in the auditory regions of people who stutter (PWS) compared to the people who do not stutter (PWNS) during both speech perception and speech production tasks (for a review see Etchell, Civier, Ballard, & Sowman, 2018).

Recent sensorimotor models of speech production have related stuttering to unstable or insufficiently activated internal models of speech sounds and a weaker feedforward control system (Max, Guenther, Gracco, Ghosh, & Wallace, 2004). It is proposed that the internal models hold the sensory consequences of speech sounds including the auditory and somatosensory information (Hickok, Houde, & Rong, 2011). Instability of the internal models are caused by failure in establishing the correct mappings between motor commands and sensory consequences during the early stages of speech acquisition, which may lead to incorrectly prepared feedforward motor commands or an inability to accurately predict the sensory consequences of prepared motor commands (Hickok et al., 2011; Max, Guenther, Gracco, Guitar, & Wallace, 2004). These models also postulates that stuttering could be caused by a heavy reliance on the external auditory and somatosensory feedbacks due to a weak feedforward control system. Since there is always a time lag for receiving external feedback following the execution of the motor commands, this may create some instabilities in the speech production system giving rise to stuttering symptoms (Civier, Tasko, & Guenther, 2010; Max et al., 2004).

The above models point to the complications in internal sensorimotor representations in stuttering and underscore an impaired integrative relationship between speech perception and production. Different behavioral studies have investigated any deficit in the internal representation of speech sound (i.e. phoneme representation) in PWS, as proposed by the above models, through different tasks such as phoneme segmentation, phoneme monitoring and nonword repetition (Bakhtiar, Dehqan AhmadAbad, & Seif Panahi, 2007; Hakim & Ratner, 2004; Sasisekaran & Byrd, 2013; Sasisekaran & De Nil, 2006; Sasisekaran, De Nil, Smyth, & Johnson, 2006). However, some of these tasks assess the metalinguistic knowledge rather than the genuine phoneme representations during the speech perception. Furthermore, the nonword repetition task may fall short in disentangling the deficits in speech perception versus articulatory speech motor skills. Neef et al. (2012) were the first to test the phoneme representations in speech perception in adults who stutter (AWS) using a categorical perception paradigm. The participants were asked to identify the speech syllables (e.g. /be/ versus /pe/) that were systematically modified in terms of their voice onset time (VOT). The results revealed a weaker discriminatory performance in AWS, and that the phonemic boundaries of the VOT continuum (e.g. /be/-/pe/) were placed at longer intervals in the AWS compared to the control group. The authors concluded that the phonemic representation in AWS is less stable or insufficiently accessed and underlined the role of speech perception deficits in persistent stuttering. In a recent study (Bakhtiar, Zhang, & Sze Ki, 2019), we used the categorical perception paradigm to examine the presence and extent of any phonemic representation deficit among children who stutter (CWS). We expanded the investigation case from consonants varying in VOT to include vowels and lexical tones to provide a more comprehensive study of the phonemic representation deficits. The results revealed important differences from Neef et al. (2012), as there were no significant differences in term of the boundary position and boundary width, signalling comparable phonemic representations in CWS and CWNS. However, the CWS group showed slower processing speed especially on the perception of the acoustic stimuli located across the categorical boundaries than the ones within the categorical boundaries, suggesting a possible inefficiency in accessing the phonemic representations in a timely manner (Bakhtiar et al., 2019).

The two studies above revealed impairments in two different aspects of categorical perception, namely the stable access of phonemic representations and processing speed during such access. It is not clear what factors lead to the discrepancy. One possibility is the target population: whereas Neef et al. (2012) studied adults who have demonstrated persistent stuttering, Bakhtiar et al. (2019) examined CWS. As a portion of CWS grow out of stuttering, it is possible that those with persistent stuttering into adulthood might demonstrate most severe impairments, therefore demonstrating the impeded access of phonemic representations instead of the processing speed. Furthermore, previous studies have shown that children's performance is usually more varied than adults, as children usually have more difficulty in sustaining their attention to the acoustic stimuli than adults and they may vary in their developmental trajectories regarding the formation of auditory perceptual abilities (Basu, Schlauch, & Sasisekaran, 2018; F. Chen, Peng, Yan, & Wang, 2017; Liu, Chen, & Tsao, 2014).

Therefore, to segregate the above possibility and shed light on the speech perception deficiencies of PWS, the current research aimed to examine the speech perception of AWS using a similar categorical perception paradigm that we used on CWS (Bakhtiar et al., 2019). In addition to the three speech stimuli types (i.e. consonants varying in VOT, lexical tone, and vowel) that we used in the previous study, a non-speech pure tone stimulus was included to examine the general auditory perception and speech perception skills in AWS versus AWNS. It is expected that the adults' data would be less noisy as they would have longer attention span and their auditory perceptual abilities should have reached maturation as compared to the children. Furthermore, we examined the response time (RT) in categorical perception in addition to the accuracy, which would inform us of any processing speed impairment in accessing the phonemic representation in AWS versus AWNS. In order to control for possible overall slowness in general decision making and hand motor coordination in AWS, which may contribute to slower processing speed in categorical perception, we also conducted a simple speeded letter decision task (Reich, Till, & Goldsmith, 1981).

Methodology

Participants

Fifteen AWS (13 males; age in years: 20-37; M= 25.60; SD = 4.75) and 15 controls (13 males; age in years: 20 – 35; M = 25.27; SD=4.38) participated in the study. The AWNS participants were matched with the AWS on age, gender, handedness, level of education and language profile. None of the participants reported any history of neurological or psychological problems, and learning difficulties including dyslexia. Demographic information of the participants is shown in Table 1. The study was approved by the Human Subjects Ethics Committee of The Hong Kong Polytechnic University, and written consents were obtained from all the participants before commencement of the study. All participants received monetary remuneration for their participation in the study.

Screening tests and tasks

Hearing test

All participants were asked to report their hearing condition and undergo a hearing screening test using pure tone audiometry at different frequencies including 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz. All participants showed normal hearing acuity.

Stuttering assessment

In order to assess the stuttering severity a minimum connected speech sample of 600 syllables, and a passage reading sample of 300 words (phonetically balanced) were collected using the video recording. The samples were analysed independently by the third author who is a graduate of Master of speech therapy (MST) program and trained in the fluency clinic of Hong Kong Polytechnic University. Stuttering severity was estimated by calculating the percentage of syllables stuttered (%SS), the average length of the three longest stuttering durations, and the degree of physical concomitant based on Stuttering Severity Instrument-3 (Riley, 1994). To determine the interrater reliability of the %SS, one-third of the speech samples were randomly selected, and independently evaluated by another trained MST student in Hong Kong Polytechnic University. The intraclass correlation coefficient of 0.995 (95% confidence interval) was achieved, indicating a very high inter-rater reliability. The measurement of stuttering severity for each participant is also shown in Table 1.

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

Participant	Age	Gender	Education	%SS(reading)	%SS(narration)	SSI-3 TOS	SSI-3 Severity
AWS1	21	M	HD	1.84	6.92	20	Mild
AWS2	21	M	BA	3.68	4.24	22	Mild
AWS3	23	M	BA	2.45	5.84	22	Mild
AWS4	20	M	BA	1.53	3.95	14	Mild
AWS5	24	M	BA	1.23	4.09	16	Mild
AWS6	25	M	BA	0.00	3.07	16	Mild
AWS7	27	M	BA	3.94	2.25	16	Mild
AWS8	34	M	BA	2.42	1.97	11	Very Mild
AWS9	37	F	BA	1.84	1.82	15	Mild
AWS10	25	M	BA	0.31	2.08	10	Very Mild
AWS11	29	M	BA	0.61	2.53	13	Mild
AWS12	22	M	BA	2.45	4.90	18	Mild
AWS13	27	F	BA	1.23	2.56	10	Very Mild
AWS14	24	M	BA	2.15	4.82	19	Mild
AWS15	25	M	BA	2.15	14.36	33	Severe
AWNS1	20	M	BA	0.92	0.59	4	-
AWNS2	22	M	BA	0.31	0.42	2	-
AWNS3	22	M	BA	0.00	0.67	4	-
AWNS4	20	M	BA	0.61	1.14	6	-
AWNS5	25	M	BA	0.61	0.93	6	-
AWNS6	25	M	BA	0.00	0.42	2	-
AWNS7	26	M	BA	0.00	1.34	8	-
AWNS8	32	M	BA	0.92	1.42	6	-
AWNS9	35	F	BA	0.31	0.62	4	-
AWNS10	25	M	BA	0.61	0.27	2	-
AWNS11	30	M	BA	0.61	0.42	2	-
AWNS12	20	M	BA	1.23	0.66	2	-
AWNS13	26	F	BA	1.23	1.21	4	-
AWNS14	25	M	BA	0.61	0.89	4	-
AWNS15	26	M	BA	0.00	0.28	2	-

Note: Higher Diploma =HD, Bachelor degree=BA, SSI-3=Stuttering severity instrument-3, TOS=Total overall scores, AWS= Adults who stutter, AWNS=Adults who do not stutter

Letter Decision Task

Previous studies indicated that some subgroups of PWS may show poorer non-speech motor coordination and slower motor initiation (Olander, Smith, & Zelaznik, 2010; Webster, 1989). In order to control for any possible effects of manual motor coordination on the categorical perception task, the two groups completed a letter decision task in which they were instructed to press different buttons on a computer mouse according to different letters (i.e., Letter 'X' and Letter 'O') displayed on the screen and their RT and accuracy responses were collected accordingly.

Stimuli

Four types of stimulus continua—consonants varying in VOT, pure tones (nonspeech), lexical tones, and vowels—were constructed for this study following the methodology of our previous studies (Bakhtiar et al., 2019; Zhang, Shao, & Huang, 2017). Three pairs of Cantonese words, which were minimally contrastive were chosen: /pa55/ (疤 'scar') vs. /p^{h}a55/ (趴 'to lie down') for the consonant continuum; /ji55/ (醫 'to treat/cure') vs. /ji25/ (椅'chair') for the lexical tone continuum and /fu55/ (膚 'skin') vs. /fɔ 55/ (科 'section') for the vowel continuum. These three minimal pairs were all meaningful words in Cantonese. The pure tone continuum is the nonspeech analogue of the lexical tone continuum. A male native Cantonese speaker was recorded reading aloud the words in isolation naturally. Each word was repeated six times and one clear token was selected for each pair to generate the stimulus continuum.

For the lexical tone continuum, we first normalized the duration of the two selected words (/ji55/ 醫 'doctor' and /ji25/ 椅 'chair') to 500 ms, and their mean intensity to 60 dB using Praat (Boersma and Weenink, 2014). We then measured the F0 at 11 time points at 10% intervals across the entire duration of /ji55/ and /ji25/. The F0 distance between /ji55/ and /ji25/ at each time point was then calculated. The distance at each time point was evenly divided into seven steps in semitones ($\Delta F0 \approx 0.74$ semitone at the onset of the stimuli, which decreased toward the end of the stimuli). At last, the original F0 contour of syllable /ji55/was replaced with the seven equally distanced F0 contours respectively using the overlap-add re-synthesis in Praat and a continuum of seven equally distanced pitch trajectories between high level tone and high rising tone was generated.

The pure tone continuum were nonspeech analogues of the lexical tone continuum. We first generated a 500-ms pure tone sound with the mean intensity at 75 dB at the frequency of 145 Hz, which is close to the mean F0 of /ji55/. The seven equally distanced F0 contours in the lexical tone continuum were then extracted and superimposed on the pure tone sound, generating a continuum of seven pure tone stimuli.

As for the vowel continuum, the two words (/fu55/ 膚 'skin' and /fo55/ 科 'section') were normalized to 500 ms in duration and 60 dB in mean intensity in Praat. We then segmented the words into consonant /f/ and following vowel (/u/ or /o/). The frequencies of the first formant (F1) were measured at 11 time points at 10% intervals across the vowel /u/ and /o/. The smallest F1 value in the measurements of /u/ and the largest F1 value in the measurements of /o/ were selected as the two end points of the F1 continuum, which was then equally divided into seven steps in Hz ($\Delta F1 \approx 42$Hz). As for the frequencies of F2-F4, the mean frequencies of /u/ and /o/ were used. Seven stimuli were synthesized by setting the frequencies of F1-F4 to the designated values in seven steps using Praat with /u55/ as the basis of manipulation. The seven synthesized stimuli were concatenated with the preceding consonant /f/, generating a continuum of seven stimuli between /fu55/ and /fo55/.

For the consonant (varying in VOT) continuum, the word /p^{h}a55/ was normalized in mean intensity to 60 dB using Praat. It was then segmented and divided into three parts: the burst release (~4.7 ms), aspiration (~36 ms), and vowel /a55/ (~420ms). The aspiration part was manipulated to vary between 0 and 36 ms in seven steps (ΔVOT = 6 ms), by shortening it proportionally using the overlap-add re-synthesis in Praat. The seven lengths of the aspiration part were concatenated with the preceding burst release and the following vowel, generating a continuum of seven stimuli that varied in VOT between /pa55/ and /p^{h}a55/.

Procedures

Each stimulus continuum was presented in an identification task and a discrimination task using the E-prime 2 software.

In the identification task, each stimulus continuum (i.e., consonants (VOTs), pure tones (nonspeech), lexical tones, and vowels) was presented in a separate block in which the seven steps of the continuum (stimuli 1–7) were repeated eight times in a random order, resulting in a total of 56 randomly ordered trials. The stimuli were presented to the participants binaurally one at a time through headphones. The participants were instructed to identify the sound they heard as one of the two minimally contrastive words for each continuum by pressing the corresponding button on the Chronos response box. Take the consonant condition for example, the instruction was: "*Please press the first button on the Chronos if you think the sound is /p/, and press the second button on the Chronos if you think the sound is /p^h/.* They were required to respond as quickly as possible within five seconds; if no response was detected within 5 seconds, the experiment would proceed to the next trial automatically. Practice trials were given to each participants to familiarize them with the procedure.

In the discrimination task, each stimulus continuum was also presented within a separate block. In each continuum, a total of 12 pairs were created for discrimination, with seven same pairs (i.e., stimuli pairs 1–1, 2–2, 3–3, 4–4, 5–5, 6–6, and 7–7) and 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 in backward order (i.e., stimuli pairs 3–1, 4–2, 5–3, 6–4, and 7–5). In each set, the same pairs were repeated 5 times, and different pairs were repeated 7 times, generating 70 trials in total. The interval between the two stimuli in each pair was fixed to be 500ms. The auditory stimuli were presented to the participants binaurally through headphones, and they were instructed to discriminate whether the two stimuli were the same or different by pressing buttons on the Chronos response box (the leftmost button for "same" responses and the second left button for "different" responses) as quickly as possible within 3 seconds. Practice trials were given to each participant to familiarize them with the task procedures.

The presentation order of the identification task and discrimination task was counterbalanced with half of the participants receiving the identification task first and the other half receiving the discrimination task first. Within each task, the orders of the four blocks were randomized. The block orders for AWNS participants were kept identical with those corresponding matched controls. Subjects were given a break between the tasks. The accuracy and response times (RT), which were measured from the offset of the stimuli, were recorded.

Data analysis

In the current study, for the identification performance, three outcome measures including the boundary position, boundary width and RT responses were analyzed. The boundary position, and boundary width were initially calculated based on the accuracy results for each participant in each stimulus continuum using probit analysis (Hallé, Chang, & Best, 2004). For instance, for the set of created sound stimuli varying in VOT (e.g., stimuli 1 through 7; see figure 1), there is a slight change from one phoneme category i.e. /pa55/ to another phoneme category i.e. /p^{h}a55/, which is called a continuum. When we hear the continuum, we may hear the stimuli 1 through 3 as examples of /pa55/**,** and stimuli 4 through 7 as examples of /p^{h}a55/. Stimuli 3 to 4 may be the boundary of the identification, called as boundary position, which indicates the position of the perception shift across two categories. Whereas, the boundary width represents the steepness of the response shift across the categorical boundaries. Therefore, in our study the boundary position was defined as the 50% crossover point in a continuum and boundary width was defined as the distance in the stimulus step between 25% and 75% of the identification responses determined by the probit analysis.

The discrimination performance was analysed using the sensitivity index d' and RT responses. The d' was formulated as the z-score of the hit rate ("different" responses to different pairs) minus the z-score of the false alarm rate ("different" responses to identical pairs) for pairs in each stimulus continuum per subject. In addition, for each subject, the pairs were classified as between-category or within-category pairs based on the boundary position obtained from their performance in the identification task. For example, if the boundary position was 3.5, the two-step pairs (i.e. 2-4 and 3-5) will be classified into the between-category group, while the remaining pairs will be classified into the within-category group.

For RT analysis, in both identification and discrimination tasks, RTs below 50 ms were first removed; RTs larger or smaller than 3SD of the mean were also removed. These calculations were conducted for each individual’s data. The discard rate was 9.04% for identification, and 5.5% for discrimination. Furthermore, the identification RT were classified into two groups—between-category and within-category—based on the boundary position obtained from the probit analysis for each participant for each stimulus continuum. For instance, if the boundary position for a particular participant was 4.5, then stimuli 4 and 5 would be labelled as the between-category stimuli and the rest of the stimuli (i.e., 1, 2, 3, 6, and 7) would be classified as the within-category stimuli. Similarly, the discrimination RT data were also classified into between-category and within-category groups according to the boundary position.

Statistical analysis

We used Linear mixed-effect (LME) models to analyze different outcome measures for each task. LME is a robust statistical analysis that has become increasingly popular in psychological sciences and psycholinguistics. LME allows modeling the fixed effects as well as random effects including the random intercepts of the items and participants and random slopes, which increase the generalisability and allow population level inferences to extend beyond limited numbers of participants and items (Baayen, Davidson, & Bates, 2008).

Regarding the identification task we used two LME models to compare the boundary position and boundary width across the groups (AWS, AWNS), stimulus types (nonspeech, consonant varying in VOT, vowel, lexical tone) and their interactions (groups × stimulus types). The model also includes the random intercept of the subjects.

For the analysis of RTs in the identification task, LME models were constructed to compare the identification performance across the groups (AWS, AWNS), stimulus types (nonspeech, consonant varying in VOT, vowel, lexical tone) and categories (between-category, within-category). The maximal model was first fitted including the above variables and their three-way interactions (i.e. groups × stimulus types × categories) as the fixed factors, and the random factors including the random intercepts of the stimuli and subjects, and the random slope of groups per stimuli, and the random slope of categories by stimulus types per subjects. Then, random intercepts and slopes were removed one by one to reach to the final simpler model. At last, a model with the random slope of stimulus types per subject, and the random slope of groups per stimuli was selected, as this model did not show any significant differences with the maximal model.

Regarding the discrimination task, LME analysis was conducted to compare the d' scores across the groups (AWS, AWNS), stimulus types (nonspeech, consonants varying in VOT, vowel, lexical tone) and categories (between-category and within-category). The final LME model was constructed including the above variables and their three-way interactions (groups × stimulus types × categories) and the random slope of stimulus types per subjects.

For the analysis of RTs in the discrimination task, LME models were constructed to compare the discrimination across the groups (AWS, AWNS), stimulus types (nonspeech, consonants varying in VOT, vowel, lexical tone) and categories (between-category and within-category). The maximal model was first fitted including the above variables and their three-way interactions (groups × stimulus types × categories) as the fixed factors, and the random factors including the random intercepts of the tones and subjects, and random slope of the groups per tones, and random slope of the categories by stimulus types per subjects. Then, random intercepts and slopes were removed one by one to reach the final simpler model. At last, a model with random slope of stimulus types per subjects was selected, as this model did not show any significant differences from the maximal model.

Results

Letter decision task

The Mann-Whiney U test revealed no group differences in term of the accuracy (U=75 , p = 0.105) and RT responses (U= 96, p = 0.494) in the letter decision task between AWS and AWNS, confirming that manual movements involved in button press are comparable between the two groups.

Identification task

The identification curves of the consonants varying in VOT, pure tones (nonspeech), lexical tones, and vowels across the seven stimuli steps for the AWS group compared with the AWNS group are shown in Fig 2, and the boundary position and width across the four stimulus continua for the AWS group compared with the AWNS group are shown in Table 2 and Fig 3.

Insert Figure 2 about here

LME models showed no significant group difference in terms of either the boundary location (Estimate = -0.089, Std. Error = 0.245, t = -0.363, p = 0.717) or the boundary width (Estimate = -0.225, Std. Error = 0.264, t = -0.851, p = 0.397). The effects of stimulus types (Estimate = -0.152, Std. Error = 0.244, t = -0.621, p = 0.536, in the boundary location analysis; Estimate = -0.0236, Std. Error = 0.237, t = -0.099, p = 0.921, in the boundary width analysis) and its interaction with groups (Estimate = 0.008, Std. Error = 0.346, t = 0.026, p = 0.980 in the boundary location analysis; Estimate = 0.404, Std. Error = 0.336, t = 1.203, p = 0.232, in the boundary width analysis) were not significant as well.

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

Group	Stimulus type	Boundary Position			Boundary Width		
		Mean	SD	Range	Mean	SD	Range
AWS	Consonant	3.182	0.965	1.240-5.077	1.280	0.594	0.380-2.202
AWNS		3.471	0.698	2.282-4.645	1.296	0.790	0.405-2.777
AWS	Nonspeech	3.247	0.603	2.209-4.393	1.228	0.663	0.390-2.423
AWNS		3.327	0.651	2.500-5.119	1.049	0.895	0.383-3.914
AWS	Tone	3.390	0.449	2.769-4.282	0.847	0.564	0.376-2.404
AWNS		3.479	0.520	2.516-4.565	1.072	0.971	0.380-4.137
AWS	Vowel	3.919	0.803	2.436-4.842	1.541	0.767	0.476-2.895
AWNS		3.849	0.741	2.045-4.726	1.303	0.655	0.476-2.502

Insert Figure 3 about here

Regarding the RT responses, results mainly demonstrated that the mean RTs for between-category stimuli (400 ms) were significantly slower than the within-category stimuli (Mean= 291 ms, Estimate = -73.093, Std. Error = 16.637, *t* = -4.394, *p* < 0.001). Concerning the effect of stimulus types, RTs for the nonspeech stimuli were significantly longer than those of the consonants varying in VOT (Estimate = -50.957, Std. Error = 21.176, *t* = -2.406, *p* = 0.01) and vowel stimuli (Estimate = -86.013, Std. Error = 22.184, *t* = -3.877, *p* < 0.001). There were no significant effect of the groups (Estimate = 33.830, Std. Error = 32.310, *t* = 1.047, *p* = 0.303) nor any interactions between the groups and categories (Estimate = 11.020, Std. Error = 38.363, *t* = 0.287, *p* = 0.776) or stimulus types (Estimate = -39.781, Std. Error = 48.562, *t* = -0.819, *p* = 0.418). RTs for the groups across different categories and stimulus types are presented in Table 3 and Figure 4.

Table 3. Descriptive data for the response time (RTs) in the identification and discrimination task.

			Identification Task			Discrimination Task		
Group	stimulus type	Category	Mean	SD	Range	Mean	SD	Range
AWS	Consonant	Between	369.55	288.49	51-1847	448.70	354.72	53-2083
		Within	334.89	325.81	50-2108	396.61	346.23	50-2276
		Overall	344.95	315.63	50-2108	406.40	348.24	50-2276
AWNS		Between	392.16	323.73	59 - 1439	418.22	290.23	53-1409
		Within	296.64	283.22	50-1745	362.05	303.26	50-1996
		Overall	323.92	298.27	50-1745	373.64	301.33	50-2276
AWS	Nonspeech	Between	462.68	333.64	51-1891	556.84	349.19	55-2052

		Within	332.37	310.76	52-1973	466.05	334.23	51-2342
		Overall	369.84	322.73	51-1973	484.33	339.08	51-2342
AWNS		Between	409.09	290.64	58-1585	451.57	318.94	66-2130
		Within	272.82	228.43	51-1375	387.75	273.73	51-1813
		Overall	311.45	254.97	51-1585	400.51	284.33	51-2130
AWS	Tone	Between	454.04	318.11	70-1671	494.42	334.28	63-2021
		Within	284.88	235.71	52-2008	430.06	302.78	51-2223
		Overall	334.65	273.53	52-2008	443.12	310.33	51-2223
AWNS		Between	434.50	322.79	59-1971	405.49	351.77	50-2169
		Within	249.14	220.13	51-1636	371.70	306.34	50-2098
		Overall	304.92	268.98	51-1971	378.59	316.23	50-2169
AWS	Vowel	Between	354.71	331.25	50-1838	430.34	278.72	62-1743
		Within	299.28	298.89	51-2183	387.09	263.84	51-2371
		Overall	315.28	309.37	50-2183	394.29	263.96	51-2371
AWNS		Between	325.56	260.71	50-1561	387.50	258.46	52-1742
		Within	260.21	219.51	50-1526	384.29	260.76	53-1879
		Overall	279.17	233.92	50-1561	384.92	260.19	52-1879

Insert Figure 4 about here

Discrimination Task

LME analysis conducted to compare the d' showed that the difference between AWS and AWNS was approaching significant (Estimate = -0.590, Std. Error = 0.329, *t* = -1.791, *p* = 0.07), with AWS performing lower d' scores than AWNS (i.e. 1.408 vs. 1.740) indicating that the overall discrimination ability of AWS was degraded compared with that of AWNS. Moreover, the d' scores for the between-category pairs were significantly higher than those for within-category pairs (Estimate = -1.622, Std. Error = 0.294, *t* = -5.508, *p* < 0.001), providing evidence for a categorical mode of perception. However, the interaction between groups and categories was not significant (Estimate = -12.650, Std. Error = 44.037, *t* = -0.287, *p* = 0.775). Other observed significant effect concerns the stimulus types, in that the d' in the consonant stimuli was lower than that in the lexical tone stimuli (Estimate = -0.898, Std. Error = 0.20647, *t* = -4.351, *p* < 0.001) (see Figs 5 & 6).

Insert figures 5 & 6 about here

Regarding the RT responses, results showed that there was a trend for interaction between groups and categories (Estimate = -30.714, Std. Error = 18.612, *t* = -1.650, *p* = 0.098), indicating AWS employed longer RT in discriminating the between-category pairs than within-category pairs (481 ms vs. 454 ms, see Fig 7). There was also a significant main effect of stimulus types in which the RT responses of the nonspeech stimuli were significantly longer than the lexical tone (Estimate = -53.534, Std. Error = 43.760, *t* = -1.930, *p* = 0.05), consonants varying in VOT (Estimate = -67.811, Std. Error = 27.196, *t* = -2.493, *p* = 0.01) and vowel stimuli (Estimate = -101.647, Std. Error = 32.208, *t* = -3.156, *p* = 0.002).

Insert figure 7 about here

Discussion

As discussed earlier our previous research showed comparable pattern of categorical perception between Cantonese-speaking CWS and CWNS, which was in contrast with a relatively similar study that compared German-speaking AWS and AWNS (Neef et al. 2012). Therefore, in order to resolve this inconsistency, the present study aimed to extend our previous research to Cantonese-speaking AWS by examining the categorical perception of three types of speech sound distinctions (i.e., consonants varying in VOT, lexical tones, and vowels) and one type of pure tone distinction. We used two popular tasks that are being used in categorical perception studies, i.e. identification and discrimination.

The results of this study revealed comparable categorical perception between AWS and AWNS in terms of the boundary position and boundary width for the identification task and higher d' scores for the between-category compared with the within-category stimuli for the discrimination task across different stimulus types (i.e. the speech and non-speech sounds). These findings cannot support a robust deficit in categorical perception of the speech sounds in AWS. These results are in line with our previous findings in children in which similar categorical perception was found across the two groups (Bakhtiar et al., 2019). However, the results of this study is in contrast with Neef et al. (Neef et al., 2012), as we did not find any group differences in terms of the boundary position and boundary width in the identification task. Different assumptions can be provided to explain this discrepancy based on the differences between the two studies in terms of the study design, and linguistic features of the acoustic stimuli. Firstly, using an adaptive procedure Neef et al. (2012) found the individualised boundary location for each subject and created 20 VOT continua in the step sizes of 1 ms intervals (+10 ~ -10 ms) around the boundary location. However, the VOT continua in our study are created in the step sizes of 8 ms intervals, which is much wider than their manipulation. Therefore, it is hypothesized that their stimuli manipulation might be to more sensitive for revealing small group differences in the categorical perception of the VOT perception. Another possible reason for this discrepancy might be related to the fact that our stimuli for VOT continuum includes speech syllables that are referring to the real words in Cantonese i.e. /pa55/ (疤 'scar') vs. /p^{h}a55/ (趴 'to lie down'). However, the speech syllables used in Neef et al. (2012) were mainly pseudowords that does not contain any semantic cues. Therefore, it is hypothesized that the speech syllables in our study would have activated the top-down processes via accessing the lexico-semantic information, which may subsequently facilitate the phonemic recognition at the lower processing levels (Levelt, 1993). Lastly, it might be important to take into account the linguistic features of the two languages that were explored in aforementioned studies, i.e., German versus Cantonese. Cantonese (unlike German) is a tonal language in which the speech syllables carries different lexical tones that systematically distinguishes lexical meaning using pitch patterns. Perhaps this feature can provide extra cues to the listeners in Cantonese, which might be absent for Indo-European languages such as German.

Regarding the RT responses, although AWS (342 milliseconds) were found to be consistently slower than AWNS (305 milliseconds) in the identification of different stimuli types and categories (see Table 3), the differences were not statistically significant. Bakhtiar et al. (2019) found significant group differences in terms of the RT responses for identification task. This difference may be explained in term of the developmental perspective. We speculate that the development of language and attention skills in AWS might provide sufficient resources to cope with the task demands needed for the identification of the speech sounds. However, it is notable that a trend for slower RT responses in discriminating the between-category pairs than within-category pairs were found for the AWS group. These findings may indicate that longer processing time is needed for AWS to discriminate the stimuli that belongs to different phonemic categories than the stimuli within the same phonemic categories. Furthermore, AWS showed marginally lower discrimination accuracy (i.e. overall d' score) than AWNS. The results may support ERP studies showing that PWS have poorer central auditory discrimination processing when discriminating the deviant sounds from the frequently presented standard sounds (Corbera et al., 2005; Jansson-Verkasalo et al., 2014; Kaganovich, Wray, & Weber-Fox, 2010). It is notable that the trends for group differences were only found in the discrimination task but not in the identification task. It is hypothesized that the discrimination task would be more demanding in terms of auditory perceptual processing and working memory as it includes recognition of two auditory stimuli in time, whereas the identification task requires to identify one auditory stimulus at a given time.

In summary, this study found comparable categorical perception between AWS and AWNS and did not support a robust deficit in phoneme representation evaluated through the speech perception among AWS in Cantonese . These results are in line with our previous findings in CWS. However, these results partially confirm the previous finding in terms of slower processing speed in discrimination task. It is notable that AWS and AWNS groups were comparable based on their manual RT responses on a simple letter decision task. Therefore, the observed trend for slowness in discrimination task cannot be attributed to a general slowness in hand motor control. Although the data from adults might be less noisy than children due to more developed auditory perceptual abilities and attention span (Basu et al., 2018), the lack of robust differences between the groups in terms of categorical perception could be also explained by some limitations of the current study. Firstly, the creation of the stimuli continua for the categorical perception in our study were based on relatively wide acoustic intervals, which may not be sufficiently sensitive to determine the more subtle categorical perception issues in AWS. Furthermore, our speech stimuli included the syllables that refer to the real words, which might provide further semantic cues and facilitate the phoneme recognition in AWS. Therefore, future studies might use the speech stimuli with more subtle acoustic intervals and limited semantic cues to investigate the categorical perception of AWS in more detail in Chinese. Furthermore, future ERP studies examining the online cognitive processes of categorical speech perception in this population might provide further insights as well.

Acknowledgments: We would like to thanks all participants of this study. This project was supported by the Departmental General Research Fund granted to the first author from the Hong Kong Polytechnic University [project ID number: P0014009].

Statement of Interest: The authors report no conflict of interest.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Bakhtiar, M., Dehqan AhmadAbad, A., & Seif Panahi, M. S. (2007). Nonword repetition ability of children who do and do not stutter and covert repair hypothesis. *Indian Journal of Medical Sciences*, *61*(8), 462. http://doi.org/10.4103/0019-5359.33711

Bakhtiar, M., Zhang, C., & Sze Ki, S. (2019). Impaired processing speed in categorical perception: Speech perception of children who stutter. *Plos One*, *14*(4), e0216124. http://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. http://doi.org/10.1016/j.jfludis.2018.07.001

Bloodstein, O. (1995). A Handbook on Stuttering. (5 ed.). San Diego: Singular.

Boersma, P., and Weenink, D. (2014). Praat: Doing Phonetics by Computer [Computer Software]. Version 5.3.84.

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, *44*(6), 1413–1434. http://doi.org/10.1017/S0305000916000581

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders*, *35*(3), 246–279. http://doi.org/10.1016/j.jfludis.2010.05.002

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

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. http://doi.org/10.1016/j.jfludis.2017.03.007

Hakim, H. B., & Ratner, N. B. (2004). Nonword repetition abilities of children who stutter: An exploratory study. *Journal of Fluency Disorders*, *29*(3), 179–199.

Halag-Milo, T., Stoppelman, N., Kronfeld-Duenias, V., Civier, O., Amir, O., Ezrati-Vinacour, R., & Ben-Shachar, M. (2016). Beyond production: Brain responses during speech perception in adults who stutter. *NeuroImage: Clinical*, *11*, 328–338. http://doi.org/10.1016/j.nicl.2016.02.017

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.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, *69*(3), 407–422. http://doi.org/10.1016/j.neuron.2011.01.019

Jansson-Verkasalo, E., Eggers, K., Järvenpää, A., Suominen, K., Van 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*, 1–11. http://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. http://doi.org/10.1080/87565641.2010.508549

Levelt, W. (1993). Speaking: From intention to articulation. Cambridge: MIT Press.

Liu, H.-M., Chen, Y., & Tsao, F.-M. (2014). Developmental Changes in Mismatch Responses to Mandarin Consonants and Lexical Tones from Early to Middle Childhood. *Plos One*, *9*, e95587–. http://doi.org/10.1371/journal.pone.0095587

Max, L., Guenther, F. H., Gracco, V. L., Guitar, B., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders*, *31*, 105–122.

Neef, N. E., Sommer, M., Neef, A., Paulus, W., Gudenberg, von, 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. http://doi.org/10.1044/1092-4388(2011/10-0224)

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. http://doi.org/10.1044/1092-4388(2009/09-0007)

Reich, A., Till, J., & Goldsmith, H. (1981). Laryngeal and manual reaction times of stuttering and nonstuttering adults. *Journal of Speech and Hearing Research*, *24*(2), 192–196.

Riley, G. (1994). Stuttering severity instrument for children and adults. Pro-ed.

Sasisekaran, J., & Byrd, C. (2013). Nonword repetition and phoneme elision skills in school-age children who do and do not stutter. *International Journal of Language & Communication Disorders*, *48*(6), 625–639. http://doi.org/10.1111/1460-6984.12035

Sasisekaran, J., & De Nil, L. F. (2006). Phoneme monitoring in silent naming and perception in adults who stutter. *Journal of Fluency Disorders*, *31*(4), 284–302. http://doi.org/10.1016/j.jfludis.2006.08.001

Sasisekaran, J., De Nil, L. F., Smyth, R., & Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *Journal of Fluency Disorders*, *31*(1), 1–21. http://doi.org/10.1016/j.jfludis.2005.11.005

Webster, W. G. (1989). Sequence initiation performance by stutterers under conditions of response competition. *Brain and Language*, *36*(2), 286–300. http://doi.org/10.1016/0093-934x(89)90066-7

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), e0183151. http://doi.org/10.1371/journal.pone.0183151

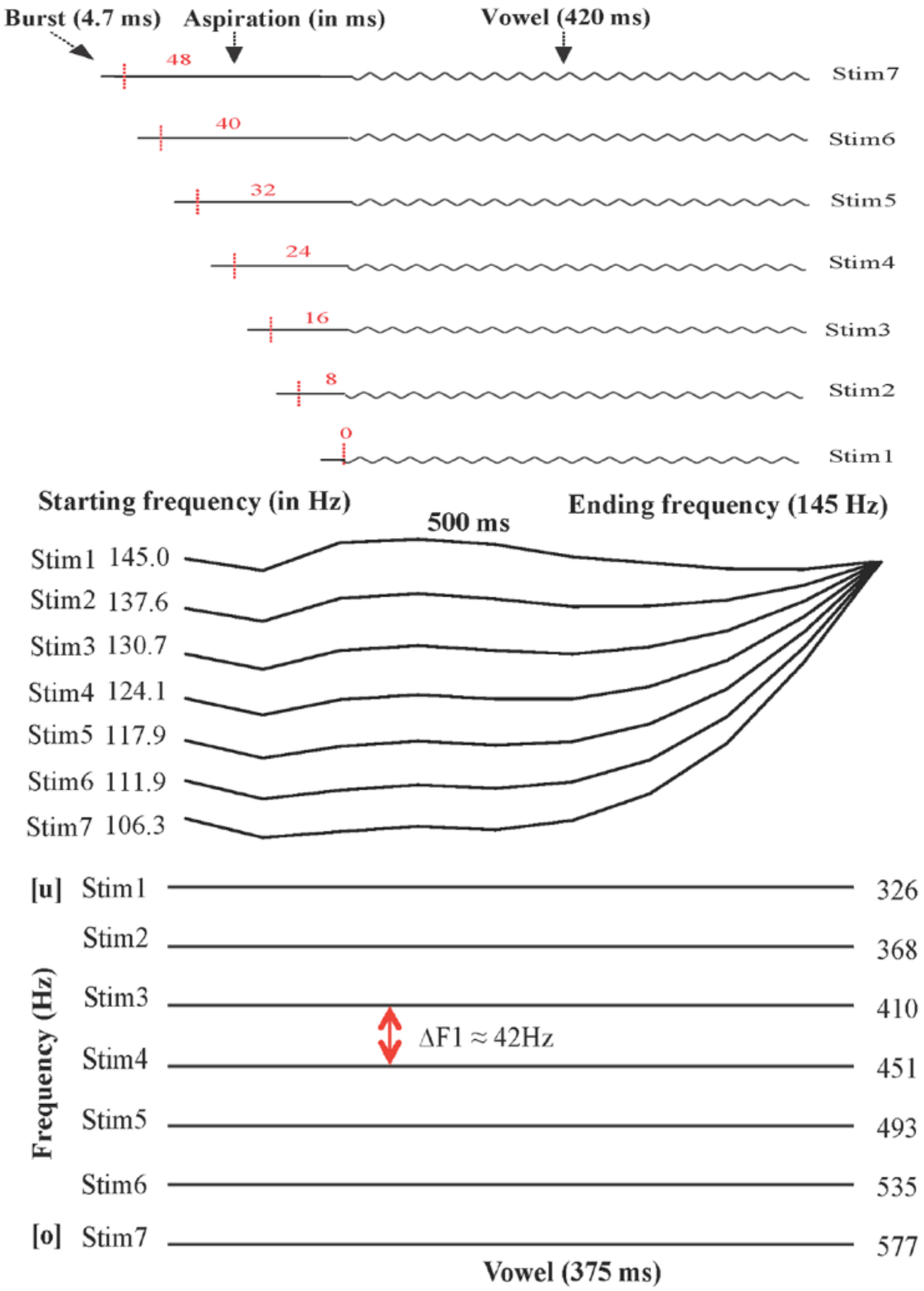

Figure 1. Schematic diagram of stimulus continua divided into seven stimuli (adopted from Zhang, Shao, & Huang, 2017). Top graph: VOT continuum. Middle graph: Lexical tone continuum. Bottom graph: Vowel continuum.

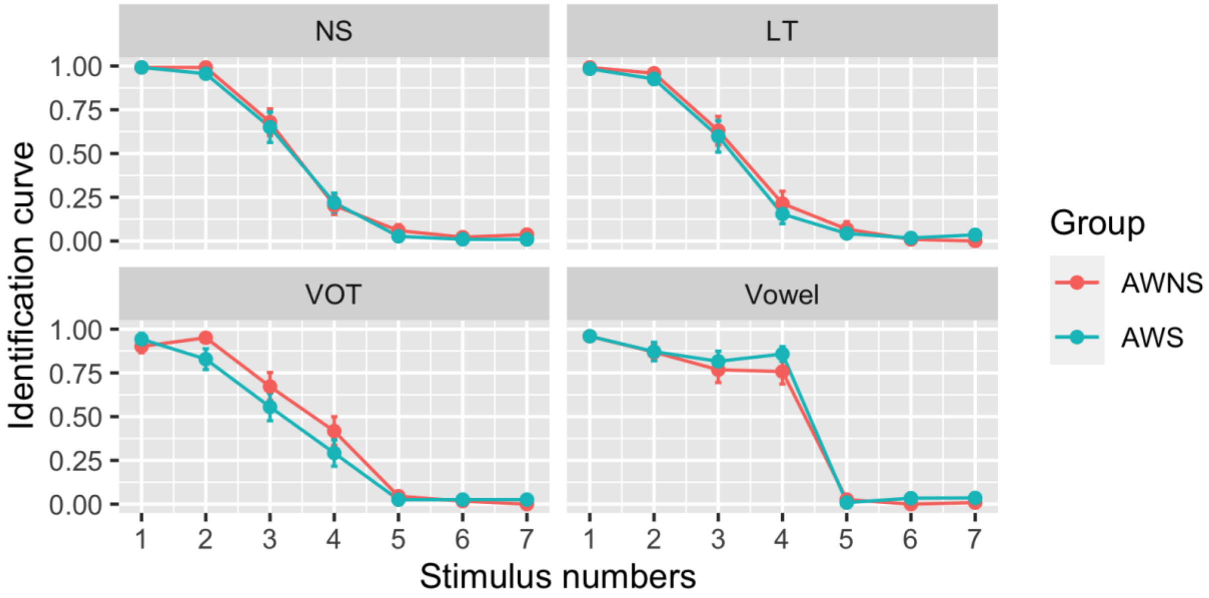

Figure 2. Identification curves for the AWS and AWNS groups across the four stimulus continua in the identification task.

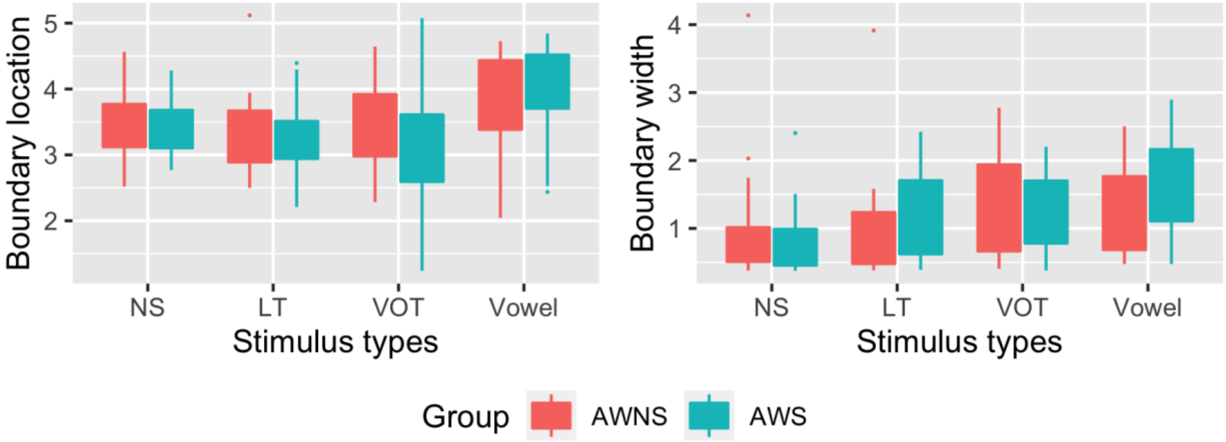

Figure 3. Boundary position and width for the AWS group compared with the AWNS group across four stimulus continua in the identification task.

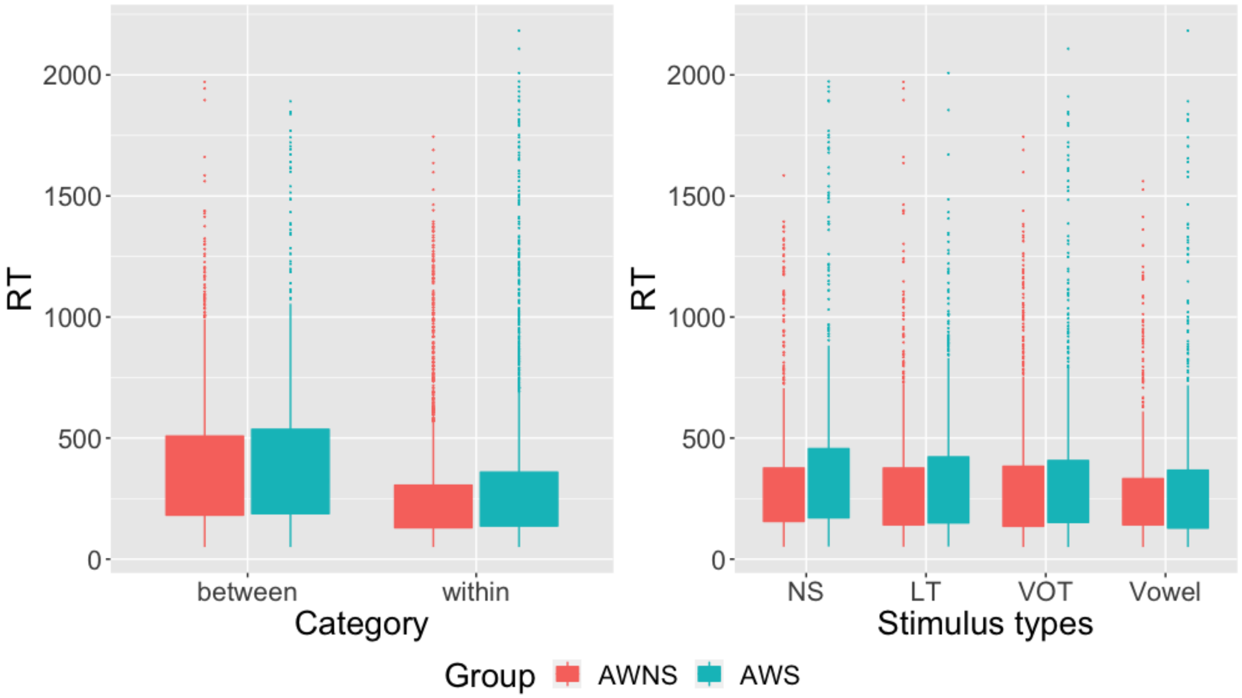

Figure 4. Interaction plots for the response times of the AWS and AWNS groups across the categorical boundaries (left) and four stimulus continua (right) in the identification task.

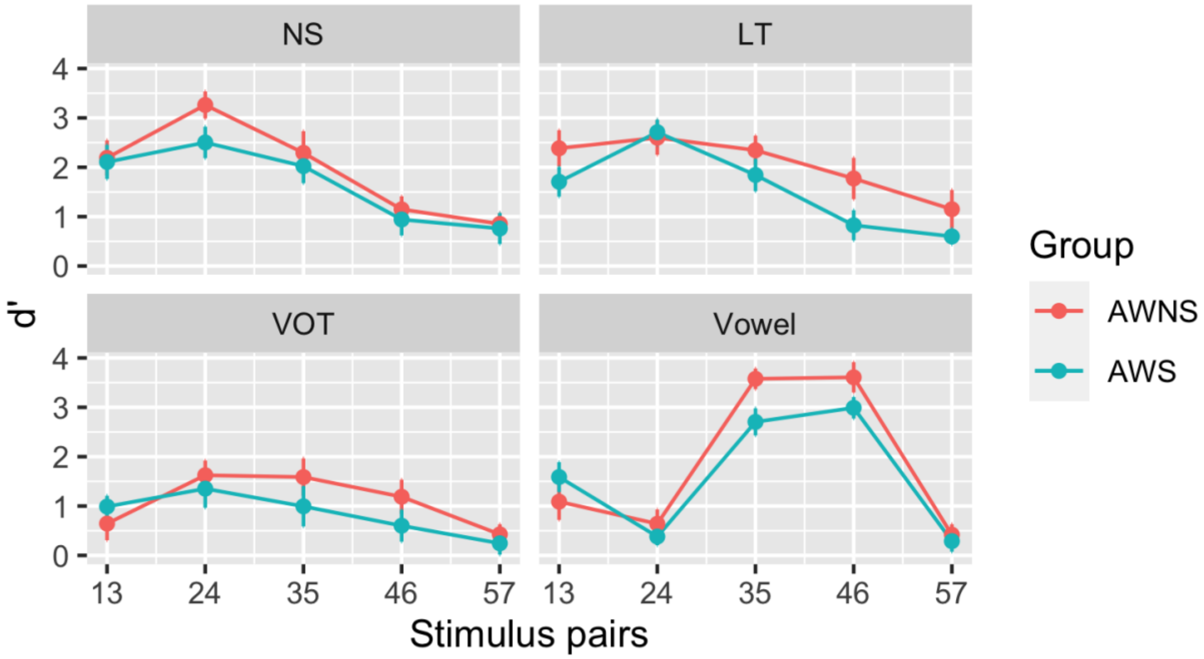

Figure 5. The d’ scores of each stimulus continuum for the AWS and AWNS groups in the discrimination task.

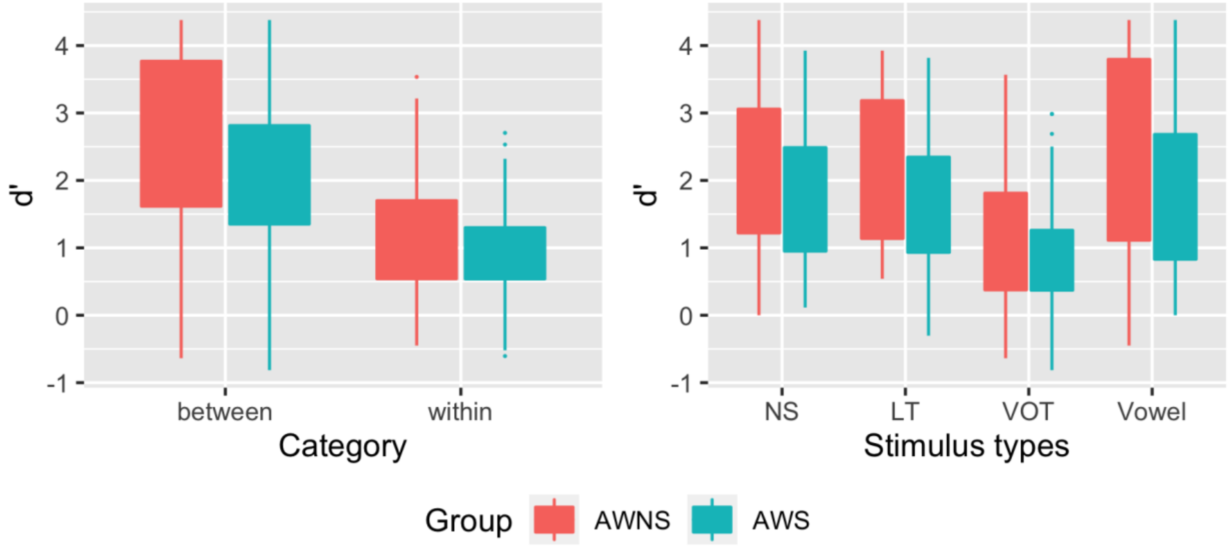

Figure 6. The interaction plot of the averaged d' across the between-category versus within-category pairs (left) and across four stimulus continua (right) for the AWS and AWNS groups in the discrimination task.

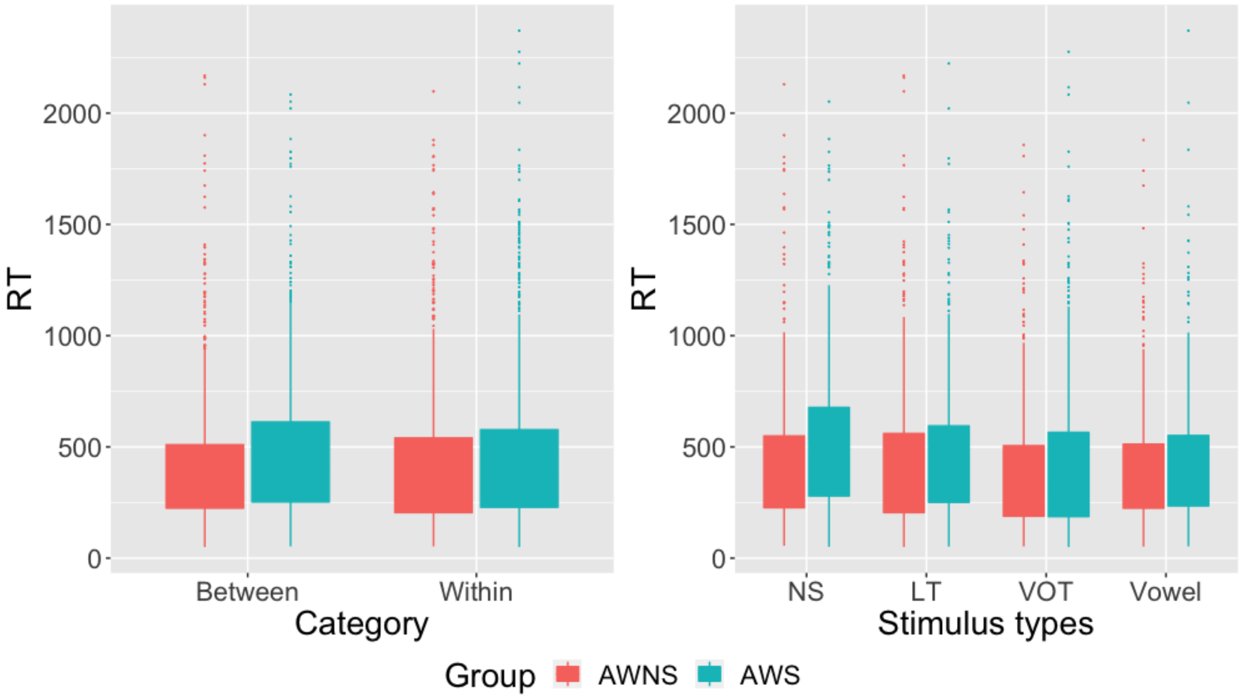

Figure 7. Interaction plots for the response times of the AWS and AWNS groups across the categorical boundaries (left) and stimulus continua (right) in the discrimination task.