Statistical Information Affects Spoken Word Recognition of Tone Languages in Stutterers: Evidence from An Auditory-Perceptual Gating Study

Jiaqiang Zhu [a], Jing Shao [b], Caicai Zhang [a*], Fei Chen [c*], Seth Wiener [d]

[a] Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

[b] Department of English Language and Literature, Hong Kong Baptist University, Hong Kong SAR, China

[c] School of Foreign Languages, Hunan University, Changsha, China

[d] Department of Modern Languages, Carnegie Mellon University, Pittsburgh, United States

*** Correspondence:**

Caicai Zhang: caicai.zhang@polyu.edu.hk

Fei Chen: chenfeianthony@gmail.com

Conflict of Interest: The authors have declared that no competing interests existed at the time of publication.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics Approval Statement: This study involving human participants was reviewed and approved by the ethics review board at the School of Foreign Languages of Hunan University. Written informed consent was signed by each participant.

Funding Statement: This study was supported by grants from the Humanities and Social Science Project of Ministry of Education of China (22YJC740008; 22YJC740093), and the Fundamental Research Funds for the Central Universities of China (531118010660).

Statistical Information Affects Spoken Word Recognition of Tone Languages in Stutterers: Evidence from An Auditory-Perceptual Gating Study

Abstract

Purpose: Previous studies have shown that individuals who stutter exhibit abnormal speech perception in addition to disfluent production as compared with their nonstuttering peers. This study investigated whether adult Chinese-speaking stutterers are still able to use knowledge of statistical regularities embedded in their native language to recognize spoken words, and if so, how much acoustic information is needed to trigger this information.

Method: Seventeen stutterers and 20 typical, nonstuttering controls participated in a gating experiment. All participants listened to monosyllabic words that consisted of syllables and lexical tones and were segmented into eight successive gates. These words differed in syllable token frequency and syllable-tone co-occurrence probability in line with a Chinese spoken word corpus. The correct syllable-only, correct tone-only, correct syllable-tone word, and correct-syllable-incorrect-tone responses were analyzed between the two groups using mixed-effects models.

Results: Stutterers were less accurate overall than controls, with fewer correct syllables, tones, and their combination as words. However, stutterers showed consistent and reliable perceptual patterns triggered by statistical information of speech, as reflected by more accurate responses to high-frequency syllables, high-probability tones, and tone errors all in manners similar to those of nonstuttering controls.

Conclusions: Stutterers’ atypical speech perception is not due to a lack of statistical learning. Stutterers were able to perceive spoken words with phonological tones based on statistical regularities embedded in their native speech. This finding echoes previous production studies of stuttering, and lends some support for a link between perception and production. Implications of pathological, diagnostic, and therapeutic conditions of stuttering are discussed.

Keywords: stuttering; statistical information; gating; spoken word recognition; syllable-tone words

1 Introduction

Stuttering is a neurodevelopmental disorder characterized by audible or silent elementary repetitions or prolongations that disrupt the flow of speech (Etchell et al., 2018; Wingate, 1978). The typical age of stuttering begins from 33 months after birth, and is found in roughly 5%–8% of preschool children at that age (Bloodstein & Bernstein Ratner, 2008). Approximately 80% of these children can recover either with or without therapy (Yairi & Ambrose, 2005). Stuttering in the teenage and adult populations is observed in about 1% where boys/men outnumber girls/women with a gender ratio being around 4:1 (Craig et al., 2002; Smith & Weber, 2017). In this study, we build on previous research related to stutterers' speech processing of Mandarin Chinese—a language that requires pitch perception for word meaning—in order to examine to what degree stutterers are able to track and make use of statistical information of Mandarin speech sounds.

1.1 Stuttering Affects Speech Production

Disfluencies in the flow of speech could occur at times in daily communications, yet stuttering-like disfluencies are distinct from normal disfluencies produced by typically fluent individuals in both quantity and quality (Smith & Weber, 2017). There is a rich literature that has identified the unusual pattern of speech production in individuals who stutter (IWS) as compared with individuals who do not stutter (IWNS). For example, Ambrose and Yairi (1999) was the first study to conduct a largescale and detailed analysis of speech samples among those who do (n = 90) and do not (n = 54) stutter. The production data were audio- and video-taped from two one-week-apart research sessions, where participants interacted with either their family members or the investigators. With the well-controlled observations, Ambrose and Yairi

(1999) drew a conclusion that part-word repetition, single-syllable word repetition, disrhythmic phonation (blocks, prolongations, and broken words), and repetition units were sufficient for differentiation purposes (stuttering versus nonstuttering). This unusual pattern was divergent from normal disfluencies, such as interjections of *um* and *ah* as well as phrase revisions (Yairi & Ambrose, 2005).

Stuttering-like disfluencies with these features have consistently been reported in a large number of subsequent studies (e.g., De Nil et al., 2008; Lu et al., 2010; Smith et al., 2012), with results further clarifying some specific and atypical production behaviors in stutterers: a higher percentage of stuttering for the initial consonant clusters whose phonemes were omitted, substituted, or distorted than the clusters without such errors (Wolk et al., 2000), lower repetitions for single-syllable words than fluently produced words, part-word repetitions, and sound prolongations (Anderson & Byrd, 2008), or the decreased accuracy in the nonword repetition task based on behavioral and kinematic indices (Smith et al., 2012). There is broad agreement that stuttering-like disfluencies are connected with stutterers' atypical control and coordination of their articulatory, respiratory, and laryngeal systems, disrupting the forward movement of speech (Howell et al., 2012; Lu et al., 2010; Walsh et al., 2015; see a meta-analysis, Belyk et al., 2015).

1.2 Stuttering Affects Speech Perception

Although stuttering is commonly diagnosed according to one's atypical speech production, rapidly growing research has also shown that stutterers' perceptual performance, which is the focus of the current study, is abnormal in tasks where the overt speech response was neither encouraged nor required (Bakhtiar et al., 2019; Corbera et al., 2005; Halag-Milo et al., 2016).

For instance, Neef et al. (2012) tested the stability of phoneme percepts by analyzing listeners' ability to identify voiced and voiceless stop consonants. In their study, two syllable continua (/də/-/tə/ and /bə/-/pə/) were synthesized based on voice onset time, the time interval of the burst and the beginning of glottal pulse in stop consonants (Cho & Ladefoged, 1999). Neef et al. (2012) found that in IWS, discriminatory performance was weaker and less stable relative to IWNS, with phoneme boundaries located at longer voice onset times. Likewise, Basu et al. (2018) examined stutterers' vowel-consonant speech syllable recognition (15 consonants paired with a vowel among /a/, /i/, or /u/) in quiet and conditions with masking noise. Results showed that speech syllable recognition in quiet and masked conditions was poor in stutterers.

Furthermore, despite the relatively few studies to date, recent studies have identified that the degraded speech perception at the segmental level can be generalized to the suprasegmental level among stutterers who speak a tone language as their mother tongue (Bakhtiar et al., 2019, 2021; Shao et al., 2022). Tone languages (e.g., Mandarin, Cantonese, and Thai) make use of a speaker's fundamental frequency (F0), which corresponds to the reciprocal of vocal fold vibratory period during speech production (Wang, 1972), to convey lexical semantics in a pitch-to-meaning manner, similar to the linguistic function of segmental phonemes (Gandour, 1983). For example, the syllable-tone combinations of "ma" bearing Tone 1 (T1, high-level tone), Tone 2 (T2, mid-rising tone), Tone 3 (T3, low-dipping tone), and Tone 4 (T4, high-falling tone) refer to "mother", "hemp", "horse", and "scorn", respectively, in Mandarin Chinese (Lee & Wiener, 2020). Ample evidence indicates that variable F0 can be perceived as different but static categories (T1 to T4) by typical, native speakers (Peng et al., 2010; Shen & Froud, 2019). However, results from studies using lexical tone to measure stutterers' perception demonstrated that IWS had difficulty in categorizing linguistic or non-linguistic (pure tone) pitch information, indexed by longer response time latency or lower perceptual accuracy, in quiet and masked conditions (Bakhtiar et al., 2019, 2021; Shao et al., 2022). Altogether, the above-mentioned work documented IWS's abnormally inferior performance to typically fluent controls in perceiving segmental and suprasegmental information of speech, pointing to an auditory-perceptual component in stuttering (Etchell et al., 2018; Howell & Bernstein Ratner, 2018; Neef et al., 2012).

1.3 Stuttering Behaviors Characterized by Statistical Regularities

Previous studies have explored and developed multiple elegant accounts for stuttering behaviors over the decades (Byrd et al., 2007; Corbera et al., 2005; Lescht et al., 2022). For instance, one of the widely accepted explanations referred to stutterers' abnormal (delayed or disrupted) phonological encoding during speech processing (Coalson & Byrd, 2015; Pelczarski et al., 2019; for a review, see Sasisekaran, 2014), such that stutterers' abstract phonological representation remained less robust in comparison to that of the controls (Neef et al., 2012; Shao et al., 2022). Importantly, there was an unstated assumption based on the stuttering data in several prior studies, which involved the possible effect of statistical regularities embedded in the language input on stuttering (Smith & Weber, 2017).

As declared by Smith and colleagues (Smith, 1990, 1999; Smith & Kelly, 1997; Smith & Weber, 2017), a combination of neurological, genetic, linguistic, environmental, emotional, and speech motor factors contributes to stuttering (Smith & Weber, 2017). The surrounding environment is filled with diverse statistical information, which can automatically be tracked by an individual in the absence of explicit feedback (Saffran, 2003; Saffran et al., 1996) in various domains, including speech (Saffran et al., 1996; Wiener & Ito, 2016), music (Peretz et al., 2012), and visual (Brady & Oliva, 2008) domains. The process of tracking these regularities involves statistical learning, which is fundamental to language acquisition (Saffran, 2003; Thiessen, 2017). Among these regularities, word frequency has been implied to play a role in stuttering in a handful of studies, though statistical learning was not primarily measured (Anderson, 2007; Castro et al., 2017; Coalson & Byrd, 2015). Brundage and Bernstein Ratner (2022) indicated that the less frequent a word is in a language, the more likely it will be stuttered or disfluent. For example, Anderson (2007) reported that low word frequency has an effect on stutter events; likewise, Newman and Bernstein Ratner (2007) found that the magnitude of the response time difference was greater in IWS than IWNS when producing words of low versus high frequency (Coalson & Byrd, 2015). A recent study by Coalson and Byrd (2017) additionally revealed the frequency effects on the suprasegmental feature of lexical stress, the relative prominence given to a syllable within a word (Teschner & Whitley, 2004). In English, the trochaic stress pattern occurs more frequently than the iambic stress pattern (e.g., IMport /ˈimpɔrt/ versus imPORT /imˈpɔrt/, Choi et al., 2019). Correspondingly, Coalson and Byrd (2017) found that when no auditory-orthographic cues were provided, IWS were more accurate when recalling trochaic than iambic nonwords in terms of their verbal responses. Nonetheless, most of the mentioned studies were indirect investigations of stutterers’ use of knowledge about statistical information and mainly concentrated on their production behaviors, with few studies using word stimuli that contained phonological tone at the suprasegmental level of speech in a perception task.

1.4 The Present Study

Our goal was to fill the research gap by evaluating whether statistical regularities embedded in the language input would affect stutterers' speech perception. The gating paradigm (Grosjean, 1980, 1996) provides a valuable opportunity for this assessment among stutterers. Firstly, the test stimuli were words from Mandarin Chinese, because syllables with tones can map directly to a morpheme or word and hence ensured the relatively straightforward examination of how speech input affects phonological (syllable and tone) and lexical (syllable-tone combination as a word) processing (Wiener & Ito, 2016; Wiener et al., 2019). Two types of statistical regularities were manipulated using the Mandarin Chinese spoken word corpus SUBTLEX-CH (Cai & Brysbaert, 2010): syllable token frequency (syllable frequency) and syllable-tone co-occurrence probability (tone probability; see more details in the Method section).

Secondly, as shown in Figure 1, the gating task involves the auditory presentation of a stimulus with increasingly longer fragments/gates, through which listeners are required to report the word in response to the heard token. Gating, therefore, allows for an investigation of how much acoustic information is needed for the listeners to access the mental lexicon to correctly recognize the segments, tones, and their combination as words. Notably, in early gates when the acoustic information is insufficient for correct identification, gating forces listeners to draw on their previous language experience and predict likely syllables, tones, and their combinations, that is, draw on their knowledge of the statistical distribution of Chinese speech sounds (Zhu et al., 2022).

[Insert Figure 1 around here]

Wiener and Ito (2016) used the gating paradigm to test native Chinese speakers. The stimuli consisted of syllable-tone words that incorporated high and low syllable token frequencies paired with a tone that was either most or least likely to co-occur with the syllable, based on a Chinese spoken word corpus (SUBTLEX-CH, Cai & Brysbaert, 2010). Participants were required to type the perceived word using the Pinyin romanization system (which specified the syllable and tone number, e.g., “da2”) after listening to each stimulus. The first gate presented only the word onset, while Gates 2 through 7 presented the onset and successive 40-ms increments. The complete word was presented at Gate 8. Results in Wiener and Ito (2016) revealed that correct responses for high-frequency syllable-tone words outnumbered low-frequency syllable-tone words. Besides, the authors analyzed correct-syllable-incorrect-tone responses, reflecting enough acoustic information for segmental identification but not suprasegmental identification. With the limited F0, listeners would either report an acoustically similar tone (e.g., T2 as T3, Moore & Jongman, 1997) or a more probable tone associated with the syllable (T2 more often than T4 on the syllable “ren”, Cai & Brysbaert, 2010). Wiener and Ito (2016) concluded that adult listeners immediately made use of the statistical regularities (i.e., syllable frequency information and tone probability information) for spoken word recognition in order to overcome the truncated speech presented via gating. Similar findings were reported in different populations, including learners of Chinese as a second language (Lee & Wiener, 2020; Wiener & Lee, 2020; Wiener et al., 2019) and individuals with congenital amusia (also known as “tone deafness”, Peretz, 2016; Zhu et al., 2022).

To summarize, the present gating study builds on prior work and adopts the gating stimuli from Wiener and colleagues (Wiener & Ito, 2016; Wiener et al., 2019; Zhu et al., 2022) to test stutterers’ spoken word recognition. Because previous studies showed that stutterers performed worse than typical individuals in speech perception at both segmental and suprasegmental levels (Corbera et al., 2005; Halag-Milo et al., 2016; Neef et al., 2012; Shao et al., 2022), we hypothesized a lower accuracy in perceiving syllables, tones, and/or syllable-tone words in stutterers versus nonstuttering controls. Meanwhile, as implied from the literature (Castro et al., 2017; Coalson & Byrd, 2015; Howell & Bernstein Ratner, 2018; Smith & Weber, 2017), it may be the case that stutterers will be sensitive to statistical regularities embedded in the language. We hence expect that stutterers will exhibit a perceptual pattern similar to the controls; that is, stutterers' spoken word recognition will also be affected by stimulus statistics in that high-frequency speech stimuli would be perceived more accurately than low-frequency stimuli.

2 Method

2.1 Participants

Seventeen individuals who stutter (IWS) and 20 individuals who do not stutter (IWNS) were recruited for this gating study. The two groups were closely matched for age, gender, and education level, such as high school degree, bachelor's degree, and graduate degree (Lescht et al., 2022). The demographic information is displayed in Table 1. All participants spoke Mandarin Chinese as their native language and were naïve to the purposes and methods of the study. None of them reported having any hearing, neurological, or psychological disorders.

Following previous studies (e.g., Bakhtiar et al., 2021; Halag-Milo et al., 2016; Howell et al., 2012), the Stuttering Severity Instrument–Third Edition (SSI-3; Riley, 1994) was used to confirm the stuttering status. The speech sample video-recorded for assessing stuttering severity encompassed an unstructured interview and a reading passage, each containing at least 600 words and independently evaluated by two research assistants majoring in speech therapy. The percentage of stuttered syllables, the average length of the three longest stuttering blocks, and the degree of physical concomitants were calculated based on SSI-3 (Riley, 1994). A high consistency was achieved between the two raters as shown by the reliability analysis (Cronbach's $\alpha = .95$). The severity of the IWS's stuttering ranged from mild to very severe (SSI-3 score: Mean = 15.88, SD = 7.01, range = 9–34). No IWNS demonstrated any stuttering-like disfluencies on the SSI-3 (Riley, 1994).

[Insert Table 1 around here]

Using digit span tests in either forward or reverse order derived from Wechsler Adult Intelligence Scale-Revised by China (Gong, 1992), the participants' working memory was also measured. The independent-samples t tests revealed that IWS did not differ from IWNS in age or working memory (both ps > .05). All participants were right-handed according to a handedness questionnaire adapted from a modified Chinese version of the Edinburgh Handedness Inventory (Oldfield, 1971). Each participant signed the consent form prior to the experiment and was paid for the participation. The study procedures were approved by the ethics review board at the School of Foreign Languages of Hunan University.

2.2 Stimuli

The auditory stimuli were adapted from previous studies (Wiener & Ito, 2016; Wiener et al., 2019; Zhu et al., 2022). On the basis of Chinese phonology (Wang, 1973), all these selected stimuli were legal Chinese monosyllabic words combined by syllables at the segmental level and lexical tones at the suprasegmental level. Correspondingly, syllable token frequency and syllable-tone co-occurrence probability of each stimulus were manipulated, following statistical distribution as determined in the 33.5-million spoken word corpus of SUBTLEX-CH

(Cai & Brysbaert, 2010).

In detail, a median common log frequency at 4.4 was firstly obtained based on the calculation of all token occurrences of a particular syllable independent of tone in SUBTLEX-CH (Cai & Brysbaert, 2010). Then, a syllable whose log frequency was above 4.8 was considered high token frequency (F+; e.g., “da”), whereas a syllable whose log frequency was lower than 4.0 was considered low token frequency (F-; e.g., “niao”). There were 24 syllables including 12 F+ syllables (with a mean log frequency at 5.08) and 12 F- syllables (with a mean long frequency at 3.84). Likewise, tone probability, referring to the most probable (P+) or least probable (P-) tone, was defined by dividing the token count for a given syllable-tone word by the token count of that particular syllable irrespective of tone. The P+ tone occurred in more than 50% of the syllables’ utterances (e.g., T2 on “pang”) in SUBTLEX-CH (Cai & Brysbaert, 2010), whereas P- tone occurred in less than 20% of the syllables’ utterances (e.g., T1 on “pang”) in this corpus. Each syllable carried both tones, resulting in a total of 48 syllable-tone words (24 F+/F- syllables × 2 P+/P- tones) for the present study.

A female native speaker from Beijing, China, with no history of speech, language, or hearing disorders, recorded all the words at 44.1 kHz in a soundproof booth. Five Chinese natives then confirmed that these words and their constituent segments and lexical tones were correct and all sounded natural. These native speakers did not attend the main experiment. Gates were generated using Praat software (Boersma & Weenink, 2018). Each word was fragmented into eight gates, starting from the consonant onset up to the beginning of the first regular periodicity of the vowel (Gate 1). The intermediate gates were developed with six 40-ms gradual increments on the rhyme (Gates 2 to 7), with the last gate (Gate 8) being the full word (see Figure 1). Because a single talker uttered these words (different from the multi-talker sounds, see Wiener & Lee, 2020), some cues, including duration and intensity, of the words were not normalized such that they could be auditorily presented in their original form as the naturistically produced tokens. This serves the current research purpose by estimating effects of stimulus statistics instead of acoustic cues on speech perception. Besides, the 40-ms increment size for the intermediate gates was implemented, which made the current study comparable to previous gating studies (e.g., Wiener & Ito, 2016; Wiener et al., 2019). In summary, there were 384 auditory items in this study (48 words × 8 gates).

2.3 Procedure

Participants were told to give their responses by typing the perceived words using Pinyin, a Romanization system taught in primary school, to specify the syllable and lexical tone. The four lexical tones were inserted as the numbers 1, 2, 3, and 4, whereby listeners felt comfortable and easy to type their answers using the computer keyboard, e.g., for “tie3” the number represented T3. The responses of Chinese characters were avoided because of the existing polyphonic characters with multiple pronunciations (e.g., “薄” has three syllable-tone combinations of “bao2”, “bo2”, and “bo4”), which would otherwise contaminate the results (Wiener & Ito, 2016). Both IWS and IWNS were told that the heard token might be a fragment instead of the entire word in order to avoid the potential floor effect, especially in early gates (Wiener et al., 2019). Moreover, since stutterers have difficulty perceiving speech (Bakhtiar et al., 2019, 2021; Neef et al., 2012), a pretest training was given for the participants to review the constituent initials and finals in a self-paced manner, following previous practice (Zhu et al., 2022). These word elements were listed on a paper handout, with none of the full syllable-tone combinations presented.

The main experiment began after the roughly 15-minute pretest training. Each participant was seated in a quiet room and listened to the tokens over headphones, which were randomly presented in a duration-blocked fashion (i.e., blocks from Gate 1 to Gate 8) via E-Prime 2.0 (Schneider et al., 2002). The order of gates from the first to the last gates was fixed to avoid the carryover effect of speech recognition from full to fragmented acoustic information (Wiener & Ito, 2016). Crucially, the present study served as an untimed task with no time limit designated in each trial, because a task with time pressure could affect IWS in numerous ways (see discussions in Lescht et al., 2022). Meanwhile, we did not provide an experimenter-generated set of targets/answers. Hence, both IWS and IWNS had to self-select items from their mental lexicon, which could benefit the estimation of the use of statistical knowledge when listeners encountered the truncated speech (Lescht et al., 2022; Wiener & Ito, 2015, 2016). Participants could familiarize themselves with the procedures by doing practice trials that did not contain the stimuli occurring in the main experiment. In total, the whole experiment lasted approximately 40–50 minutes.

2.4 Data Preparation Before Analysis

Illegal responses that violated Chinese phonology (nonwords, e.g., “mau1” and “biy2”) were firstly excluded (1%, cf. 4% in Wiener & Ito, 2016). With this criterion, a total of 14,019 responses was eligible for data analysis. To comprehensively clarify how statistical regularities of syllables, tones, and their combination as words affected listeners’ typing answers, a manual check was done for each trial in order to identify correct syllable-only responses (e.g., the response “qi3” to the stimulus “qi2”), correct lexical tone-only responses (e.g., “gang1” to "pang1"), and correct syllable-tone word responses (e.g., "zhou2" to "zhou2").

Each response was additionally checked if it was incorrect with lexical tone, i.e., tone errors. These tone errors were defined as correct-syllable-incorrect-tone responses by stutterers and nonstuttering controls. Two types of tone errors were further separated, namely acoustic-based errors and probability-based errors, in line with the preceding studies (Wiener & Ito, 2016; Wiener et al., 2019; Zhu et al., 2022). Acoustic-based errors involved reporting an acoustically similar tone given the two tones' similar F0 cues. For example, T1 and T4 both start from the high pitch register, and native and non-native listeners commonly confuse this tone pair, especially in early gates where short fragments contained high F0 (Wiener & Ito, 2016; Wiener et al., 2019). Similarly, T2 and T3 could be confused with each other due to their initial acoustic ambiguity (Moore & Jongman, 1997). For example, the response "wu3" to the stimulus "wu2" was coded as an acoustic-based error. The other error type was probability-based errors, which stemmed from reporting the statistically more probable tone associated with the perceived syllable. These errors could occur in spite of the two tones' acoustic dissimilarity. As an instance, the response "da4" to the stimulus "da2" was regarded as a probability-based error, given that these tones started in opposite registers and were not acoustically similar. The responses that did not belong to the two types of errors (e.g., "hong3" to "hong4"), or happened to be either acoustic-based or probability-based error (e.g., "bin1" to "bin4") were not further analyzed (Wiener & Ito, 2016).

With the formula of log[(probability error + 0.5)/(acoustic error + 0.5)], the empirical log of the error ratio was computed for each participant at each gate as a function of syllable frequency (F+ and F−). A positive log ratio reflected that a listener made more probability-based errors than acoustic-based errors, but a negative value indicated that a listener made more acoustic-based errors than probability-based errors (Wiener & Ito, 2016; Wiener & Lee, 2020; Wiener et al., 2019).

2.5 Data Analysis

The data were analyzed by constructing generalized (accuracy) or linear (error log ratio) mixed-effects models with the lme4 package (Bates et al., 2015) in R (R Core Team, 2021). The response to each trial was coded as 0 (incorrect) or 1 (correct) for each participant. Gate-by-gate analyses were carried out to identify the locus of speech statistics effect on speech perception at both segmental and suprasegmental levels. To increase power and reduce the number of models run, gates were collapsed into four windows (e.g., Window 1 by Gates 1 and 2, Window 2 by Gates 3 and 4, Window 3 by Gates 5 and 6, and Window 4 by Gates 7 and 8; Wiener & Ito, 2016; Wiener et al., 2019; Zhu et al., 2022).

For the analysis of correct responses, the models were built with group (IWS and IWNS), syllable frequency (F+ and F−), and tone probability (P+ and P−) acting as fixed factors, with the dependent variables being syllable-only accuracy, lexical tone-only accuracy, or syllable-tone word accuracy in the four windows. For the analysis of tone errors, the models were built with group (IWS and IWNS), syllable frequency (F+ and F−), and window (Windows 1 to 4) acting as fixed factors, with empirical log ratio as the dependent variable. Two-way and three-way interaction terms were also included as the fixed effects in the models. By-subject and by-item random intercepts and slopes for all possible fixed factors were included in the initial model (Barr et al., 2013). Using the analysis of variance function in lmerTest package (Kuznetsova et al., 2017), the initial model was compared with a simplified model that excluded a specific fixed factor for both accuracy and error analyses. Pairwise comparisons with Tukey adjustment were calculated using the lsmeans package (Lenth, 2016).

3 Results

3.1 Accuracy Results

Figure 2 exhibits the mean syllable accuracy, tone accuracy, and syllable-tone word accuracy by IWS and IWNS at different gates. It shows the gate-by-gate improvement across the test tokens for both groups of listeners: more acoustic information of the stimuli led to higher accuracy. At each gate, the same pattern occurred: tones were identified most accurately, followed by syllables and syllable-tone combinations; besides, IWS were less accurate in general than IWNS. The between-group differences as a function of statistical regularities were further analyzed and reported below.

[Insert Figure 2 around here]

3.1.1 Correct Syllable-Tone Combinations

Figure 3 plots the mean syllable-tone word accuracy in IWS and IWNS as faceted by syllable frequency and tone probability at different windows, with error bars showing 1 standard error (SE). The figure shows that IWS were more accurate as more acoustic information was provided, but their performance remained poorer than their nonstuttering peers. The mixed-effects models revealed the significant main effects of group (χ^2 (1) = 5.14, $p < .05$), syllable frequency (χ^2 (1) = 26.10, $p < .001$), and tone probability (χ^2 (1) = 6.73, $p < .01$), with their three-way interaction (χ^2 (1) = 3.95, $p < .05$) at Window 1. Further analysis of this interaction showed that IWNS ($M = 0.12$; $SD = 0.33$) were more accurate than IWS ($M = 0.04$; $SD = 0.20$) when listening to F- syllables combined with P- tones ($\beta = 1.18$, SE = 0.47, $t = 2.54$,

$p < .05$). Besides, P- tones were recognized less accurately than P+ tones when presented on F- syllables for both IWS (β = -2.21, SE = 0.59, t = -3.73, $p < .001$) and IWNS (β = -1.39, SE = 0.55, t = -2.54, $p < .05$). Regardless of tone probability and group, F- syllables were always identified less accurately than F+ syllables (ps < .05). At Window 2, the models revealed a significant main effect of syllable frequency (χ^2 (1) = 18.13, $p < .001$) as well as an interaction between syllable frequency and tone probability (χ^2 (1) = 3.87, $p < .05$). Further analysis of this interaction showed that as compared with P+ tones, P- tones were less accurate when they co-occurred with F- syllables (β = -1.64, SE = 0.68, t = -2.39, $p < .05$); moreover, both IWS and IWNS less accurately recognized F- syllables than F+ syllables in combination with P- tones (β = -3.23, SE = 0.69, t = -4.68, $p < .001$).

[Insert Figure 3 around here]

There was a significant main effect of group at Window 3 (χ^2 (1) = 6.13, $p < .05$), which revealed that IWNS (*M* = 0.76; *SD* = 0.42) had a higher accuracy than IWS (*M* = 0.66; *SD* = 0.47). There were also a significant main effect of syllable frequency (χ^2 (1) = 13.11, $p < .001$), and an interaction between syllable frequency and tone probability (χ^2 (1) = 6.71, $p < .05$). Further analysis of this interaction showed that P- tones were less accurately identified than P+ tones with F- syllables (β = -1.65, SE = 0.57, t = -2.88, $p < .01$), and F- syllables were less accurate than F+ syllables when carrying P- tones (β = -2.67, SE = 0.58, t = -4.62, $p < .001$) by both groups with or without stuttering. Similar to Window 3, the models revealed a significant main effect of group (χ^2 (1) = 5.44, $p < .05$) at the last window, which demonstrated that IWS (*M* = 0.79; *SD* = 0.41) had a lower accuracy than IWNS (*M* = 0.88; *SD* = 0.32) for the test tokens regardless of their statistical characteristics. The significant main effect of syllable frequency (χ^2 (1) = 19.75, p < .001) and interaction between syllable frequency and tone probability (χ^2 (1) = 6.97, p < .01) were found. Further analysis of this interaction showed that F- syllables were less accurately identified than F+ syllables with P- tones (β = -2.43, SE = 0.46, t = -5.30, p < .001), and P- tones were less accurately recognized than P+ tones on F- syllables (β = -1.20, SE = 0.44, t = -2.76, p < .01). Other effects did not reach significance (ps > .05).

In summary, results of syllable-tone word accuracy revealed that IWS were outperformed by IWNS when either the fragments or the entire words were heard, particularly for the tokens of infrequent syllable and improbable tone with minimal acoustic information provided at Window 1. This manifested IWS's reduced spoken word recognition relative to IWNS. Besides, analogous to IWNS, IWS more accurately identified P+ tones on F- syllables (as compared with P- tones) and P- tones on F+ syllables (as compared with F- syllables) across the four windows, regardless of how much acoustic information was auditorily presented. This finding revealed that similarly to the nonstuttering controls, stutterers' perception was affected by statistical regularities embedded in a language with phonological tone. Next, correct syllable-only responses were analyzed and reported below.

3.1.2 Correct Syllable-Only Responses

Figure 4 depicts the mean syllable-only accuracy in IWS and IWNS as faceted by syllable frequency and tone probability at different windows, with error bars showing 1 SE. This figure shows the rising accuracy in both IWS and IWNS when acoustic information increased from the first to the last windows, but IWS showed a lower average accuracy than IWNS. The mixed-effects models revealed that at Window 1, there were significant main effects of group (χ^2 (1) = 6.14, p < .05) and syllable frequency (χ^2 (1) = 37.15, p < .001), which demonstrated the higher accuracy in IWNS (*M* = 0.50; *SD* = 0.50) than IWS (*M* = 0.41; *SD* = 0.49) independent of stimulus statistical characteristics; besides, both groups more accurately recognized F+ syllables than F- syllables co-occurring with either P+ or P- tones. The models revealed a significant main effect of syllable frequency (χ^2 (1) = 33.71, $p < .001$) and an interaction between syllable frequency and tone probability (χ^2 (1) = 3.91, $p < .05$) at the second window. Further analysis of this interaction showed that both groups performed less accurately when listening to F- syllables than F+ syllables with either P- tones (β = -3.97, SE = 0.64, t = -6.24, $p < .001$) or P+ tones (β = -2.17, SE = 0.63, t = -3.46, $p < .001$), and they less accurately recognized P- tones than P+ tones on F- syllables (β = -1.34, SE = 0.62, t = -2.16, $p < .05$) but not F+ syllables (β = 0.46, SE = 0.64, t = 0.72, p = .47).

[Insert Figure 4 around here]

At Window 3, the models uncovered a significant main effect of syllable frequency (χ^2 (1) = 23.46, $p < .001$) and an interaction between syllable frequency and tone probability (χ^2 (1) = 7.58, $p < .01$). Further analysis of this interaction showed that as compared with P+ tones, P- tones were recognized less accurately on F- syllables (β = -1.61, SE = 0.62, t = -2.60, $p < .01$); besides, F- syllables were less accurately recognized than F+ syllables with P- tones (β = -3.74, SE = 0.65, t = -5.78, $p < .001$). The similar main effect of syllable frequency (χ^2 (1) = 21.10, $p < .001$) and interaction between syllable frequency and tone probability (χ^2 (1) = 8.74, $p < .01$) were revealed at Window 4. Further analysis of this interaction showed that both groups less accurately identified P- tones than P+ tones on F- syllables (β = -1.62, SE = 0.55, t = -2.98, $p < .01$), and F- syllables than F+ syllables as bearing P- tones (β = -3.56, SE = 0.69, t = -5.18, $p < .001$). Other effects were not significant (*p*s > .05).

In summary, results of syllable-only accuracy demonstrated reduced accuracy by IWS compared to IWNS in terms of syllable identification at Window 1. Besides, both groups more accurately perceived F+ syllables than F- syllables regardless of tone probability at Windows 1 and 2, yet, as acoustic information further extended, F+ syllables were more accurately recognized than F- syllables with co-occurrence of P- tones alone. Correct lexical tone-only responses were next analyzed and reported.

3.1.3 Correct Lexical Tone-Only Responses

Figure 5 displays the mean lexical tone-only accuracy in IWS and IWNS as faceted by syllable frequency and tone probability at different windows, with error bars showing 1 SE. The figure shows that the starting point of mean lexical tone-only accuracy was about 50%, higher than mean syllable-tone word and syllable-only accuracies, pointing to the relative ease of tone identification; other patterns were similar in that more acoustic information led to higher accuracy, and IWS showed their lower perceptual accuracy than IWNS.

[Insert Figure 5 around here]

The mixed-effects models revealed no significant main or interaction effects involving group, syllable frequency, and tone probability at Window 1. There was a marginally significant main effect of group (χ^2 (1) = 2.97, p = .09) at the second window, which demonstrated a trend that IWS (M = 0.81; SD = 0.39) were outperformed by IWNS (M = 0.83; SD = 0.38). At Window 3, a significant main effect of group (χ^2 (1) = 5.91, p < .05) was found, suggesting that IWNS (M = 0.93; SD = 0.26) had a higher accuracy than IWS (M = 0.88; SD = 0.33) irrespective of speech statistics relating to syllable and tone. The main effect of group was again uncovered (χ^2 (1) = 4.87, p < .05) at Window 4, which reflected that IWS (M = 0.93; SD = 0.26) had a lower accuracy than IWNS (*M* = 0.97; *SD* = 0.17). Other effects were not significant (*p*s > .05).

In summary, results of mean lexical tone-only accuracy exhibited that IWS were poorer than IWNS in tone perception mainly at Windows 3 and 4, where the acoustic information in the auditory items gradually approximated the full word. This hinted that IWS's abnormal speech perception also existed at the suprasegmental level of lexical tones, even though the complete F0 cues were available. Next, incorrect responses of correct-syllable-incorrect-tone errors were reported below.

3.2 Correct-Syllable-Incorrect-Tone Errors

Following Wiener and colleagues (Wiener & Ito, 2016; Wiener et al., 2019; Zhu et al., 2022), the correct syllable but incorrect tone responses were analyzed using the mixed-effects models. This involved the cases of enough acoustic information for correct identification of the syllable, but not necessarily sufficient acoustic information for F0 categorization of the tone. Listeners were hence forced to use the limited F0 information, which led to an acoustic-based error, or rely on their statistical knowledge of the most probable tone associated with the heard syllable, which resulted in a probability-based error.

[Insert Figure 6 around here]

Figure 6 portrays the mean log ratio of errors in IWS and IWNS as faceted by syllable frequency at different windows, with error bars showing 1 SE. The figure shows that the log ratio approximates zero as gates become longer, indicative of the gradually equal (and limited) number of both types of tone errors as stimulus acoustic information increases; besides, more negative log ratios for frequent than infrequent syllables are exhibited for both IWS and IWNS. The mixed-effects models firstly revealed a significant main effect of group (χ^2 (1) = 4.55, *p*

$< .05$) and an interaction between group and syllable frequency (χ^2 (1) = 5.22, $p < .05$). Further analysis of this interaction showed that IWS (*M* = -0.40; *SD* = 0.43) had a more negative log ratio than IWNS (*M* = -0.19; *SD* = 0.41) when listening to F+ syllables (β = 0.22, SE = 0.07, *t* = 3.01, $p < .01$); besides, IWS had a more negative log ratio for tones carried by F+ syllables than F- syllables (β = -0.21, SE = 0.04, *t* = -4.80, $p < .001$). The models also revealed the significant main effects of syllable frequency (χ^2 (1) = 15.80, $p < .001$) and window (χ^2 (1) = 22.14, $p < .001$), with their two-way interaction (χ^2 (1) = 18.01, $p < .001$). Further analysis of this interaction showed that both groups had a more negative log ratio for tones on F+ syllables than F- syllables at Window 1 (β = -0.26, SE = 0.05, *t* = -5.13, $p < .001$), Window 2 (β = -0.17, SE = 0.05, *t* = -3.36, $p < .001$), and Window 3 (β = -0.11, SE = 0.05, *t* = -2.27, $p < .05$). This suggested listeners' reliance on syllable token frequency information. In addition, as tones co-occurred with F+ syllables, the log ratio was more negative at Window 1 than other later Windows 2, 3, and 4, respectively (*p*s < .05); likewise, for F- syllables, the log ratio was more negative at Window 1 than at Window 3 and at Window 2 than at Window 3, respectively (*p*s < .05). This demonstrated that listeners were less likely to make acoustic-based tone errors with increased acoustic signal. Other effects were not significant (*p*s > .05).

Collectively, the results of the mean log ratio of errors showed that both IWS and IWNS primarily made acoustic-based errors on F+ syllables (as compared with F- syllables), but IWS were more likely to make acoustic-based errors on F+ syllables (as compared with IWNS). These errors became less common as the stimulus acoustic information increased.

4 Discussion

The current study examined stutterers' spoken word recognition in the face of impoverished acoustic input using a gating paradigm (Grosjean, 1980, 1996; Wiener & Ito, 2016). The auditory words varied in syllable token frequency and syllable-tone co-occurrence probability, in line with the Chinese spoken word corpus of SUBTLEX-CH (Cai & Brysbaert, 2010). These syllable-tone words were played in increasing gates to individuals who stutter (IWS) in order to examine whether statistical regularities in the auditory input affect their spoken word recognition behavior. Their performance, including both accuracy and tone errors, was compared with a control group of individuals who do not stutter (IWNS) and analyzed with mixed-effects models. Notably, syllable-tone word accuracy, syllable-only accuracy, tone-only accuracy, and correct-syllable-incorrect-tone errors were all analyzed to systematically evaluate whether stutterers' abnormal speech perception co-occurred at both the segmental and suprasegmental levels. We reported two key findings.

First, we corroborated that there was an auditory-perceptual component in stuttering (Halag-Milo et al., 2016; Howell & Bernstein Ratner, 2018; Smith & Weber, 2017), with IWS showing poorer speech perception than IWNS for either segments or suprasegmentals (Bakhtiar et al., 2021; Basu et al., 2018; Corbera et al., 2005). This finding was robust as the IWS were less accurate than IWNS across comparisons of statistical analyses. This finding held for the first window of syllable-only analysis, suggesting that when minimal acoustic information was provided (onset and up to 40 ms of the vowel), IWS struggled to recognize the syllable. For the lexical tone-only analysis, this finding held for Windows 2, 3, and 4 (onset and 80 ms or more of the vowel), which indicated that tone categorization was problematic for IWS even when the majority of F0 information was available (Bakhtiar et al., 2019; 2021; Shao et al., 2022). For the syllable-tone word analysis, IWS again showed less accurate identification than IWNS at all windows except Window 2. The correct-syllable-incorrect-tone analysis additionally revealed that for F+ syllables, IWS were more likely to make acoustic-based tone errors than IWNS across the four windows. Taken together, these findings underline the difficulty that IWS have with speech perception as compared with typical individuals, and that difficulty occurs not only in recognizing syllables and lexical tones, but in recognizing syllable-tone combination as words (Bakhtiar et al., 2019; 2021; Neef et al., 2012; Shao et al., 2022).

Second, we found that in our syllable-only and syllable-tone word analyses, IWS made use of the statistical information to correctly recognize high-frequency (F+) syllables more accurately than low-frequency (F-) syllables. This finding held for both IWS and IWNS across all the four windows in both analyses involving syllable. This strengthens the claim that the syllable plays an important role in Chinese speech perception (and production) for native speakers, including individuals who stutter (Chen et al., 2002; You et al., 2012). Given the roughly 400 unique syllables in Mandarin Chinese (Duanmu, 2007), more frequent syllables are "privileged" than less frequent syllables during lexical processes (Lee & Wiener, 2020).

This syllable result also relates to a recent study which also required participants to respond in a typed mode (Lescht et al., 2022). Lescht et al. reported that while IWS generated fewer words than IWNS in letter fluency performance, they showed a pattern of generating more words with fewer syllables, because shorter words (e.g., bat, cat, and mat) more frequently occur than longer words (e.g., elephant and elegant) in a cluster sharing similar phonemes in English.

More importantly, we found novel evidence that stutterers were able to perceive the syllable-tone words according to syllable-tone co-occurrence probabilities. Similar to IWNS, IWS were more likely to make acoustic-based tone errors on F+ syllables than F- syllables from Windows 1 to 3 (onset and up to 200 ms of the vowel) in our analysis of correct-syllable-incorrect-tone errors, given that F+ syllables were easier to predict than F- syllables (Zhu et al., 2022). Crucially, the most prominent instance was that in our analysis of syllable-tone word accuracy, the interaction between syllable frequency and tone probability emerged in both groups across the windows, irrespective of the stuttering status and the amount of acoustic information. The results showed that both IWS and IWNS more accurately recognized F+ syllables with P- tones as compared with F- syllables, and F- syllables with P+ tones as compared with P- tones. This again corroborated previous findings from gating (typical Chinese speakers, Wiener & Ito, 2016; learners of Chinese as a second language, Wiener & Lee, 2020, Wiener et al., 2019; individuals with congenital amusia, Zhu et al., 2022; spoken word recognition in noise, Wang et al., 2023), and word reconstruction experiments (Wiener, 2020; Wiener & Turnbull, 2016).

This result is also in line with several speech production studies in the stuttering literature, which indirectly connected statistical information of speech with stuttering (Anderson, 2007; Bernstein Ratner et al., 2009; Brundage & Bernstein Ratner, 2022). For example, word frequency had an effect on the occurrence of stutter events and affects the length of reaction time in word production (Anderson, 2007; Newman and Bernstein Ratner, 2007). More recently and relatedly, Coalson and Byrd (2015, 2017) additionally demonstrated that accuracy of verbal responses by IWS was also impacted by the frequency effects even for the suprasegmental unit of lexical stress in English (more frequent trochaic stress pattern versus less frequent iambic stress pattern). However, Coalson and Byrd (2015, 2017) used nonword stimuli whose wordlikeness was rated as "unlike a word" and "neutral", with the frequency of syllable (at the segmental level) being invariant (Baayen et al., 1995).

Taken together, our study revealed a *typical* pattern of statistical regularities in atypical speech perception among stutterers. That is, statistical regularities embedded in the language input impacted on how IWS perceived the syllable-tone words, despite their fewer correct responses as compared with their fluent peers. The above-mentioned results relating to stimulus statistical characteristics hence demonstrated that statistical regularities, either for segments or suprasegmentals, could be a moderating factor in stutterers' speech perception. Furthermore, a potential link between speech perception and production could be supported, given statistical information affects speech perception and production in IWS despite their abnormal speech processing (Anderson, 2007; Brundage & Bernstein Ratner, 2022; Castro et al., 2017; Coalson & Byrd, 2015, 2017). In our perceptual study, IWS showed inferior performance to IWNS across the between-group comparisons; however, they could recognize a word based not only on its syllable token frequency but also on the syllable-tone co-occurrence probability, presumably due to their lifetime exposure to their native language. Previous studies have shown that listeners are able to involuntarily track the embedded patterns from their surrounding environment without explicit instruction, such as the three-syllable linguistic (Omigie & Stewart, 2011; Saffran et al., 1996) or three-motif music word segmentation (Peretz et al., 2012; Saffran et al., 1999) from a non-pause, continuous sound flow. Likewise, in the absence of active feedback, listeners can build a phonological category for a mental lexicon because they automatically capture the frequency and variability of the heard sounds (Escudero & Williams, 2014; Maye et al., 2002). This process of subliminally computing the regularities is important in statistical learning (Erickson & Thiessen, 2015). Although our study did not focus on statistical learning by IWS, the results could indicate their preserved statistical knowledge that was accumulated implicitly via their everyday experience with Chinese. This might contribute to their comparable pattern in perception as IWNS. The recent evidence showing IWS's spared implicit learning further suggested that IWS could manage to perceive speech in a statistical manner despite their degraded speech perception (Alm, 2021; Höbler et al., 2022; Smits-Bandstra & Gracco, 2013). For example, Smits-Bandstra and Gracco (2013) argued that although IWS showed less implicit sequence learning relative to IWNS, their slower reaction times were found only for early but not late learning trials, which implied delays but not deficiencies in general learning. It was thus likely to conclude that IWS perceived the gated tokens by drawing on their implicit knowledge of statistical information, i.e., syllable frequency and tone probability, in their native language.

Stutterers' degraded speech perception appears to be primarily attributed to their aberrant lexical access to the mental lexicon (Lescht et al., 2022; McGill et al., 2016; Newman & Bernstein Ratner, 2007). The gating experiment delicately controls the length of each gate and evaluates the amount of acoustic information needed to access the mental lexicon (Grosjean, 1996; Wiener & Ito, 2016). Our study followed previous gating studies with the same test tokens and replicated the finding in terms of the performance profile in typical, healthy individuals (Wiener & Ito, 2016; Wiener & Lee, 2020; Wiener et al., 2019). However, IWS were outperformed by the nonstuttering controls with their fewer correct responses; besides, they were more likely to make acoustic-based tone errors on F+ syllables. This could reflect that IWS's processing of stimulus acoustic signal was diminished (Basu et al., 2018; Corbera et al., 2005), hence they required more acoustic input to access the mental lexicon and correctly identify the syllable-tone word as compared with IWNS. Meanwhile, there are multiple functions, such as conceptualization, selection, and determination of a lexicon (Levelt, 2001; Levelt et al., 1999), involved in lexical access. Although not measured directly, some cognitive functions related to lexical access could remain abnormal in IWS (Maxfield et al., 2015; McGill et al., 2016; Pellowski, 2011), especially considering their poorer performance than IWNS when the full word was played in the current study. For example, it has been articulated that IWS have difficulties in selecting a target word among a cohort of candidate words (Lescht et al., 2022; Maxfield, 2020). The limited lexical selection in IWS was partially ascribed to their abnormal inhibitory control system, which refers to the capacity to plan and suppress inappropriate responses under instructions or in novel or uncertain situations (Choi et al., 2013) and afflicts both child and adult stutterers (Eggers et al., 2010; Maxfield, 2020). Relatedly, Anderson and Wagovich (2017) found that IWS's inhibitory control abilities were different from IWNS in suppressing a dominant response while executing a conflicting response, manifesting a "less controlled response style" (Eggers et al., 2013; see a meta-analysis, Ofoe et al., 2018). Consequently, IWS were graded lower than IWNS, as exhibited in the current auditory-perceptual gating study. In addition to aberrant inhibitory control of lexical selection, future experiments may want to define whether other sub-component(s) of lexical access or executive functions in stutterers could be divergent from typical individuals, which interferes with their word retrieval among the stored lexical items and reduces the efficient operation of their lexical access skills (Etchell et al., 2018; Lescht et al., 2022; Maxfield et al., 2012).

An alternative account referred to IWS's poor phonological processing abilities (Byrd et al., 2007; Halag-Milo et al., 2016; Sasisekaran, 2014). Previous studies have consistently found the group differences in phonological categorization of speech sounds, including segmental phonemes and suprasegmental lexical tones, between IWS and IWNS. For example, Neef et al. (2012) uncovered that IWS perceived categories of phonemes less distinctly (Basu et al., 2018). The recent studies replicated Neef et al.'s results and further revealed that although being native tonal-language speakers, IWS still had difficulty in categorizing lexical tones (e.g., Bakhtiar et al., 2019, 2021; Shao et al., 2022). These studies identified that in categorical perception, between-category comparisons were discriminated more easily than within-category comparisons by both groups; however, a significantly lower score obtained by IWS than IWNS was observed in processing between-category comparisons, which indicated IWS's less robust phonological representation than IWNS (Bakhtiar et al., 2019, 2021; Shao et al., 2022; see Xu et al., 2006 for discussions). Our study did not give a pool of experimenter-generated targets for listeners to select a response (e.g., Lescht et al., 2022); instead, both IWS and IWNS self-selected the items from their mental lexicon (Grosjean, 1996; Wiener & Ito, 2016). It was likely that IWS's unstable phonological representation led to their greater ambiguity in speech perception (Neef et al., 2012; Shao et al., 2022), as shown by the current findings of IWS's worse performance than IWNS.

With respect to ramifications for clinical intervention, our finding of preserved statistical knowledge of language in stutterers is informative for researchers, clinicians, and speech therapists to design the training programs to rehabilitate stuttering. Emotional or temperamental factors, such as negative quality of mood (Johnson et al., 2010), are tightly connected with stuttering. Two stages can be divided in terms of treatment efforts: at the early stage stutterers could practice using high-frequency speech stimuli more often so as to build their confidence (i.e., due to their higher accuracy or fluency), but at the later stage there could also be the low-frequency speech materials with which stutterers have severer problems. By harnessing the statistical information embedded in speech, the training efforts may firstly guide IWS to learn to perceive/produce the high-frequency words, or design some situations where the conversational speech contains the high-frequency words as many as possible. Correspondingly, statistical characteristics of speech elements, including segments and suprasegmentals, may be cautiously considered, because they interacted with each other as shown in our analyses (e.g., IWS more accurately identified F+ syllables with P- tones but not P+ tones, as compared with F- syllables). With this "statistical" training method, the accurate speech perception/production that IWS might achieve is beneficial to soften their anxiety of making errors and establish their confidence, possibly reducing their avoidance behaviors in the challenging situations (with low-frequency or stuttered tokens) and reaching more communication effectiveness (Byrd et al., 2022). Future studies or intervention programs with a component relating to stimulus statistical regularities (i.e., separating or choosing stimuli based on their statistics) could be worthy of a try in an aim to mitigate stuttering.

We conclude by noting the limitations of the study. Firstly, our study did not record participants' response time. Response time serves as an important index of the efficiency of how listeners process the tokens (Strange et al., 2011), which has been widely used for between-group comparisons between IWS and IWNS in the stuttering literature. After controlling the typing speed (e.g., with baseline typing tasks, Lescht et al., 2022), the introduction of response time into data analysis would draw a more comprehensive picture, such as the speed-accuracy trade-off (Bakhtiar et al., 2019; Howell & Bernstein Ratner, 2018), of stuttering behaviors. Secondly, our stimuli were not produced by different speakers. Future studies may want to exploit the gated tokens that were uttered by both male and female speakers (e.g., Wiener & Lee, 2020) to improve the ecological validity of the lab-based setting and better mimic the real-life speech situations. Thirdly, as stimulus acoustic cues play a significant role in speech perception, it would be interesting to explore how stimulus statistical and acoustic information interact to influence stutterers' perceptual performance, as compared with nonstuttering controls. A study manipulating both types of information benefits such an estimation. Lastly, our study did not design a production task in addition to the perception task. Further experiments may employ both tasks so as to directly evaluate the effects of statistical information on speech processing and address the perception-production link in IWS more clearly. Besides, with the aid of brain imaging tools, the cerebral correlates of this link are expected to be uncovered. This is crucial to the understanding of the pathology of stuttering as a multifaceted disorder involving both speech perception and production (Smith & Weber, 2017), and is heuristic for the treatment efforts to remediate stuttering in that IWS's perception skills also need to improve as their production abilities.

In conclusion, individuals who stutter showed degraded speech perception for Chinese spoken words. They showed lower accuracy for Chinese syllables, tones, and their combination as syllable-tone words; moreover, they were more likely to make acoustic-based tone errors. Nevertheless, despite their impoverished performance, stutterers exhibited the reliable pattern similar to typical, nonstuttering control listeners, indicative of their preserved statistical knowledge of tracking regularities embedded in speech. In all, our findings manifest an auditory-perceptual component involving phonological tone in stuttering, but there remains a typical pattern in stutterers' atypical speech perception in relation to statistical information of speech.

5 References

Alm, P. A. (2021) The dopamine system and automatization of movement sequences: A review with relevance for speech and stuttering. *Frontiers in Human Neuroscience, 15*, 661880. https://doi.org/10.3389/fnhum.2021.661880

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research, 42*(4), 895–909. https://doi.org/10.1044/jslhr.4204.895

Anderson, J. D. (2007). Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. *Journal of Speech, Language, and Hearing Research, 50*(1), 229–247. https://doi.org/10.1044/1092-4388(2007/018)

Anderson, J. D., & Byrd, C. T. (2008). Phonotactic probability effects in children who stutter. *Journal of Speech, Language, and Hearing Research, 51*(4), 851–866. https://doi.org/10.1044/1092-4388(2008/062)

Anderson, J. D., & Wagovich, S. A. (2017). Explicit and implicit verbal response inhibition in preschool-age children who stutter. *Journal of Speech, Language, and Hearing Research, 60*(4), 836–852. https://doi.org/10.1044/2016_JSLHR-S-16-0135

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* [CD]. Philadelphia, PA: University of Pennsylvania, Linguistics Data Consortium.

Bakhtiar, M., Shao, J., Cheung, M. N., & Zhang, C. (2021). Categorical perception of speech sounds in adults who stutter. *Clinical Linguistics & Phonetics, 35*(6), 560–576. 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), e0216124. 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., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belyk, M., Kraft, S. J., & Brown, S. (2015). Stuttering as a trait or state–an ALE meta-analysis of neuroimaging studies. *European Journal of Neuroscience, 41*(2), 275–284. https://doi.org/10.1111/ejn.12765

Bernstein Ratner, N., Newman, R., & Strekas, A. (2009). Effects of word frequency and phonological neighborhood characteristics on confrontation naming in children who stutter and normally fluent peers. *Journal of Fluency Disorders, 34*(4), 225–241. https://doi.org/10.1016/j.jfludis.2009.09.005

Bloodstein, O., & Ratner, N. (2008). *A Handbook on Stuttering* (6th ed.). Thompson Delmar Learning, Boston, MA.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer*. http://www.praat.org

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes extracting categorical regularities without conscious intent. *Psychological Science, 19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Brundage S. B., & Bernstein Ratner N. (2022). Linguistic aspects of stuttering: Research updates on the language–fluency interface. *Topics in Language Disorders, 42*(1), 5–23. https://doi.org/10.1097/TLD.0000000000000269

Byrd, C. T., Coalson, G. A., & Young, M. M. (2022). Targeting communication effectiveness in adults who stutter: A preliminary study. *Topics in Language Disorders, 42*(1), 76–93. https://doi.org/10.1097/TLD.0000000000000270

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: Holistic versus incremental processing. *American Journal of Speech-Language Pathology, 16*(1), 43–53. https://doi.org/10.1044/1058-0360(2007/006)

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE, 5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Castro, N., Pelczarski, K. M., & Vitevitch, M. S. (2017). Using network science measures to predict the lexical decision performance of adults who stutter. *Journal of Speech, Language, and Hearing Research, 60*(7), 1911–1918. https://doi.org/10.1044/2017_JSLHR-S-16-0298

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781. https://doi.org/10.1006/jmla.2001.2825

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Choi, D., Conture, E. G., Walden, T. A., Lambert, W. E., & Tumanova, V. (2013). Behavioral inhibition and childhood stuttering. *Journal of Fluency Disorders, 38*(2), 171–183. https://doi.org/10.1016/j.jfludis.2013.03.001

Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: Tone language, experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners. *Cognition, 189*, 188–192. https://doi.org/10.1016/j.cognition.2019.04.004

Coalson, G. A., & Byrd, C. T. (2015). Metrical encoding in adults who do and do not stutter. *Journal of Speech, Language, and Hearing Research*, *58*(3), 601-621. https://doi.org/10.1044/2015_JSLHR-S-14-0111

Coalson, G. A., & Byrd, C. T. (2017). Nonword repetition in adults who stutter: The effects of stimuli stress and auditory-orthographic cues. *PLOS ONE*, *12*(11), e018811. https://doi.org/10.1371/journal.pone.0188111

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

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of stuttering in the community across the entire life span. *Journal of Speech, Language, and Hearing Research, 45*(6), 1097–1105. https://doi.org/10.1044/1092-4388(2002/088)

De Nil, F., Beal, D. S., Lafaille, S. J., Kroll, R. M., Crawley, A. P., & Gracco, V. L. (2008). The effects of simulated stuttering and prolonged speech on the neural activation patterns of stuttering and nonstuttering adults. *Brain and Language, 107*(2), 114–123. https://doi.org/10.1016/j.bandl.2008.07.003

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York: Oxford University Press.

Eggers, K., De Nil, L., & Van den Bergh, B. (2010). Temperament dimensions in stuttering and typically developing children. *Journal of Fluency Disorders, 35*(4), 355–372. https://doi.org/10.1016/j.jfludis.2010.10.004

Eggers, K., De Nil, F. L., & Van den Bergh, B. (2013). Inhibitory control in childhood stuttering. *Journal of Fluency Disorders, 38*(1), 1–13. https://doi.org/10.1016/j.jfludis.2012.10.001

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37,* 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition, 133*(2), 408–413. https://doi.org/10.1121/1.3629144

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

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*(2), 149–175. https://doi.org/10.1016/s0095-4470(19)30813-7

Gong, Y. X. (1992). *Wechsler Adult Intelligence Scale-Revised in China Version*. Changsha: Hunan Medical College.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267–283. https://doi.org/10.3758/BF03204386

Grosjean, F. (1996). Gating. *Language and Cognitive Processes, 11*(6), 597–604. https://doi.org/10.1080/016909696386999

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. https://doi.org/10.1016/j.nicl.2016.02.017

Höbler, F., Bitan, T., Tremblay, L., & De Nil, L. (2022). Differences in implicit motor learning between adults who do and do not stutter. *Neuropsychologia, 174*, 108342. https://doi.org/10.1016/j.neuropsychologia.2022.108342

Howell, T. A., & Bernstein Ratner, N. (2018). Use of a phoneme monitoring task to examine lexical access in adults who do and do not stutter. *Journal of Fluency Disorders, 57*, 65–73. https://doi.org/10.1016/j.jfludis.2018.01.001

Howell, P., 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. https://doi.org/10.1016/j.bandl.2012.09.010

Johnson, K., Walden, T., Conture, E., & Karrass, J. (2010). Spontaneous regulation of emotions in preschool-age children who stutter: Preliminary findings. *Journal of Speech Language, and Hearing Research, 53*(6), 1478–1495. https://doi.org/10.1044/1092-4388(2010/08-0150)

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, C. Y., & Wiener, S. (2020). Acoustic-based and knowledge-based processing of Mandarin tones by native and non-native speakers. In H. M. Liu, F. M. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition: Multidisciplinary approaches in Chinese languages* (Vol. 11, pp. 37–57). Springer Nature. https://doi.org/10.3389/fpsyg.2020.00214

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lescht, E., Dickey, M. W., Stockbridge, M. D., & Bernstein Ratner, N. (2022). Adults who stutter show diminished word fluency, regardless of mode. *Journal of Speech, Language, and Hearing Research, 65*(3), 906–922. https://doi.org/10.1044/2021_JSLHR-21-00344

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy* *of Sciences, 98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Lu, C., Chen, C., Ning, N., Ding, G., Guo, T., Peng, D., Yang, Y., Li, K., & Lin, C. (2010). The neural substrates for atypical planning and execution of word production in stuttering. *Experimental Neurology, 221*(1), 146–156. https://doi.org/10.1016/j.expneurol.2009.10.016

Maxfield, N. D. (2020). Inhibitory control of lexical selection in adults who stutter. *Journal of Fluency Disorders, 66*, 105780. https://doi.org/10.1016/j.jfludis.2020.105780

Maxfield, N. D., Morris, K., Frisch, S. A., Morphew, K., & Constantine, J. L. (2015). Real-time processing in picture naming in adults who stutter: ERP evidence. *Clinical Neurophysiology, 126*(2), 284–296. https://doi.org/10.1016/j.clinph.2014.05.009

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

McGill, M., Sussman, H., & Byrd, C. T. (2016). From grapheme to phonological output: Performance of adults who stutter on a word jumble task. *PLOS ONE, 11*(3), e0151107. https://doi.org/10.1371/journal.pone.0151107

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America, 102*(3), 1864–1877. https://doi.org/10.1121/1.420092

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)

Newman, R.S., & Bernstein Ratner, N. (2007). The role of selected lexical factors on confrontation naming accuracy, speed, and fluency in adults who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 50*(1), 196–213. https://doi.org/10.1044/1092-4388(2007/016)

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. https://doi.org/10.1044/2018_JSLHR-S-17-0372

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Omigie, D., & Stewart, L. (2011). Preserved statistical learning of tonal and linguistic material in congenital amusia. *Frontiers in Psychology, 2*, 109. https://doi.org/10.3389/fpsyg.2011.00109

Pelczarski, K. M., Tendera, A., Dye, M., & Loucks, T. M. (2019). Delayed phonological encoding in stuttering: Evidence from eye tracking. *Language and Speech, 62*(3), 475–493. https://doi.org/10.1177/0023830918785203

Pellowski, M. W. (2011). Word-finding and vocabulary abilities of adults who do and do not stutter. *Contemporary Issues in Communication Science and Disorders, 38*, 126–134. https://doi.org/10.1044/cicsd_38_F_126

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*(5), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences, 20*(11), 857–867. https://doi.org/10.1016/j.tics.2016.09.002

Peretz, I., Saffran, J., Schon, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences, 1252*(1), 361–366. https://doi.org/10.1111/j.1749-6632.2011.06429.x

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Riley, G. D. (1994). *Stuttering severity instrument for children and adults* (3rd ed.). Pro-Ed.

Saffran, J. R. (2003). Musical learning and language development. *Annals of the New York Academy of Sciences, 999*(1), 397–401. https://doi.org/10.1196/annals.1284.050

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Sasisekaran, J. (2014). Exploring the link between stuttering and phonology: A review and implications for treatment. *Seminars in Speech and Language, 35*(2), 95–113. https://doi.org/10.1055/s-0034-1371754

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's Guide*. Pittsburgh, PA: Psychlogy Software Incorporated.

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. https://doi.org/10.1044/2022_JSLHR-21-00276

Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition, 22*(2), 253–265. https://doi.org/10.1017/S136672891800038X

Smith, A. (1990). Factors in the etiology of stuttering. In J. A. Cooper (Ed.), *Research needs in stuttering: Roadblocks and future directions* (pp. 39–47). Rockville, MD: American Speech-Language-Hearing Association.

Smith, A. (1999). Stuttering: A unified approach to a multifactorial, dynamic disorder. In N. Bernstein Ratner & E. C. Healey (Eds.), *Stuttering research and practice: Bridging the gap* (pp. 27–44). Hove, United Kingdom: Psychology Press.

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: Evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of Fluency Disorders, 37*(4), 344–358. https://doi.org/10.1016/j.jfludis.2012.06.001

Smith, A., & Kelly, E. (1997). Stuttering: A dynamic, multifactorial model. In R. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (pp. 204–217). Boston, MA: Allyn & Bacon.

Smith, A., & Weber, C. (2017). How stuttering develops: The multifactorial dynamic pathways theory. *Journal of Speech, Language, and Hearing Research, 60*(9), 2483–2505. https://doi.org/10.1044/2017_JSLHR-S-16-0343

Smits-Bandstra, S., & Gracco, V. (2013). Verbal implicit sequence learning in persons who stutter and persons with Parkinson's disease. *Journal of Motor Behavior, 45*(5), 381–393. https://doi.org/10.1080/00222895.2013.812058

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Teschner, R. V., & Whitley, S., M. (2004*). Pronouncing English: A stress-based approach with CD-ROM.* Washington DC: Georgetown University Press.

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160056. https://doi.org/10.1098/rstb.2016.0056

Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of Neurodevelopmental Disorders, 7*, 27. https://doi.org/10.1186/s11689-015-9123-8

Wang, W. S. Y. (1972). The many uses of F0. In *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre,* ed. by Albert Valdman, 487–503. The Hague: Mouton.

Wang, W. S. Y. (1973). The Chinese language. *Scientific American, 228*(2), 50–60.

https://doi.org/10.1038/scientificamerican0273-50

Wang, X., Lee, C. Y., & Wiener, S. (2023). Non-native disadvantage in spoken word recognition is due to lexical knowledge and not type/level of noise. *Speech Communication, 149,* 29–37. https://doi.org/10.1016/j.specom.2023.03.004

Wiener, S. (2020). Second language learners develop non-native lexical processing biases. *Bilingualism: Language and Cognition, 23*(1), 119–130. https://doi.org/10.1017/S1366728918001165

Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition & Neuroscience, 30*(9), 1048–1060. https://doi.org/10.1080/23273798.2014.946934

Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics, 56*, 38–51. https://doi.org/10.1016/j.wocn.2016.02.001

Wiener, S., & Lee, C. Y. (2020). Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology, 11*, 214. https://doi.org/10.3389/fpsyg.2020.00214

Wiener, S., Lee, C. Y., & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning, 69*(3), 527–558. https://doi.org/10.1111/lang.12342

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, *59*(1), 59–82. https://doi.org/10.1177/0023830915578000

Wingate, M. E. (1978). Disorders of fluency. In: Skinner, P., Sheldon, R. (Eds.), *Speech, language, hearing: Normal processes and disorders*. Reading, MA: Addison-Wesley.

Wolk, L., Blomgren, M., & Smith, A. B. (2000). The frequency of simultaneous disfluency and phonological errors in children: A preliminary investigation. *Journal of Fluency Disorders, 25*(4), 269–281. https://doi.org/10.1016/S0094-730X(00)00076-0

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering: For clinicians by clinicians*. Pro-Ed.

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLOS ONE, 7*(10), e46595. https://doi.org/10.1371/journal.pone.0046595

Zhu, J., Chen, X., Chen. F., & Wiener, S. (2022). Individuals with congenital amusia show impaired speech perception but preserved statistical learning for tone languages. *Journal of Speech, Language, and Hearing Research, 65*(1), 53–69. https://doi.org/10.1044/2021_JSLHR-21-00383

Tables and Figures

Table 1. The demographic information of IWS and IWNS.

Subject information	**IWS**	**IWNS**
No. of participants (Male)	17 (12)	20 (12)
Age in years, Mean (SD)	26.76 (4.42)	24.30 (3.73)
Working memory, Mean (SD)	13.71 (2.57)	14.35 (0.99)
Education (HS:BA:MA)	5:10:2	5:11:4

Note. IWS = individuals who stutter; IWNS = individuals who do not stutter; HS = high school degree; BA = bachelor's degree; MA = master's degree.

Figure 1. The schematic illustration of the gating paradigm. The length of the colored bar represents the amount of stimulus acoustic information. The stimulus acoustic information extends with the increasingly longer gates from the syllable onset (Gate 1) to the full word (Gate 8), with the 40-ms incremental size for the intermediate gates (Gates 2 to 7).

Figure 2. Mean correct tone, syllable, and syllable-tone word responses by IWS and IWNS at each gate.

Figure 3. Mean correct syllable-tone word responses as faceted by syllable frequency and tone probability by IWS and IWNS at each window (error bar = ±1 *SE*).

Figure 4. Mean correct syllable-only responses as faceted by syllable frequency and tone probability by amusics and typical listeners at each window (error bar = ±1 *SE*).

Figure 5. Mean correct lexical tone-only responses as faceted by syllable frequency and tone probability by IWS and IWNS at each window (error bar = ±1 *SE*).

Figure 6. Mean empirical log ratio of tone errors as faceted by syllable frequency by IWS and IWNS at each window (error bar = ±1 *SE*).

Figure 1

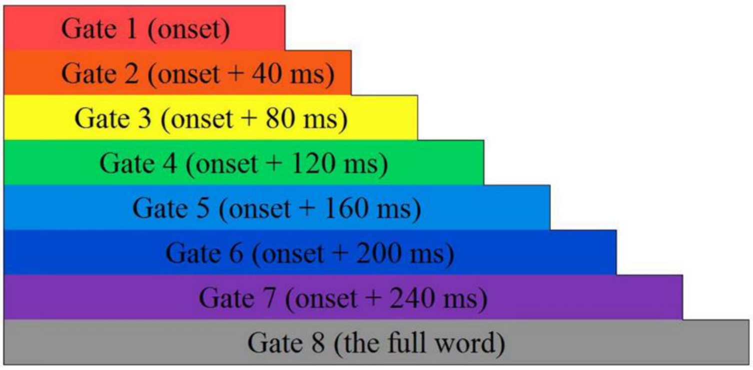
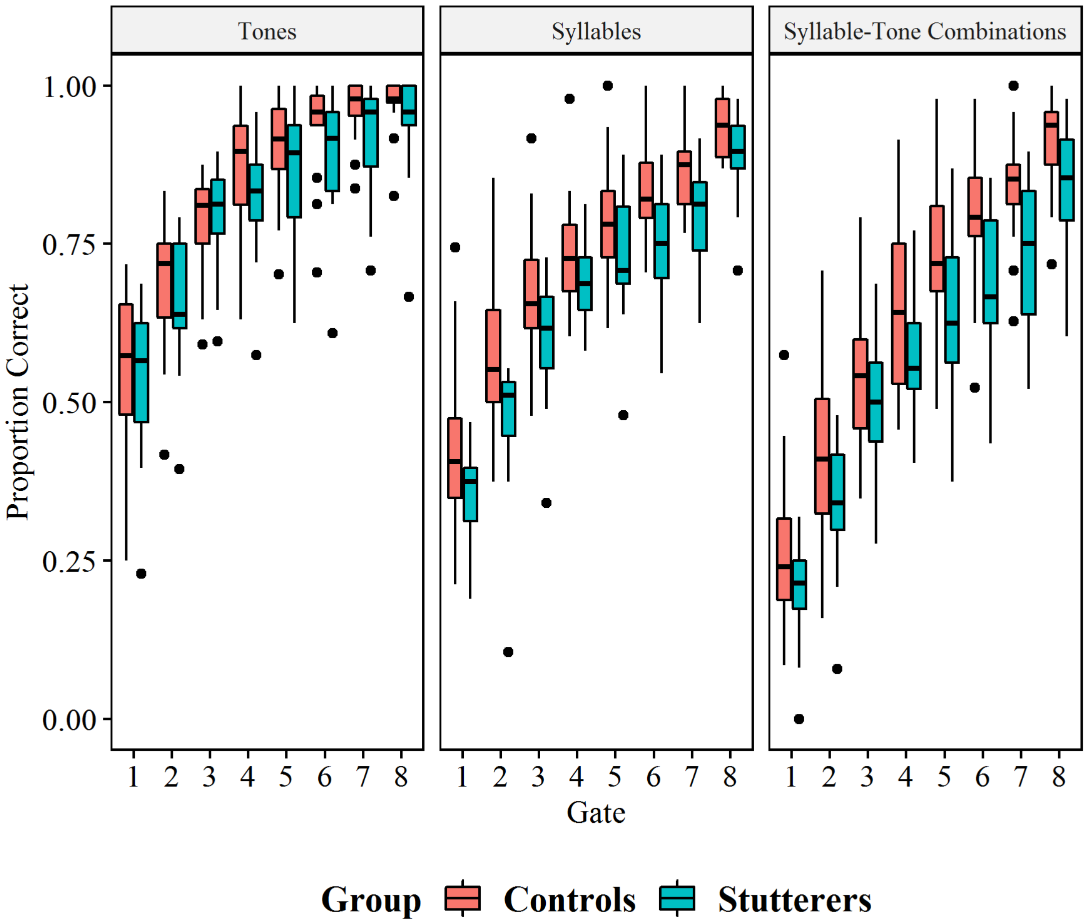
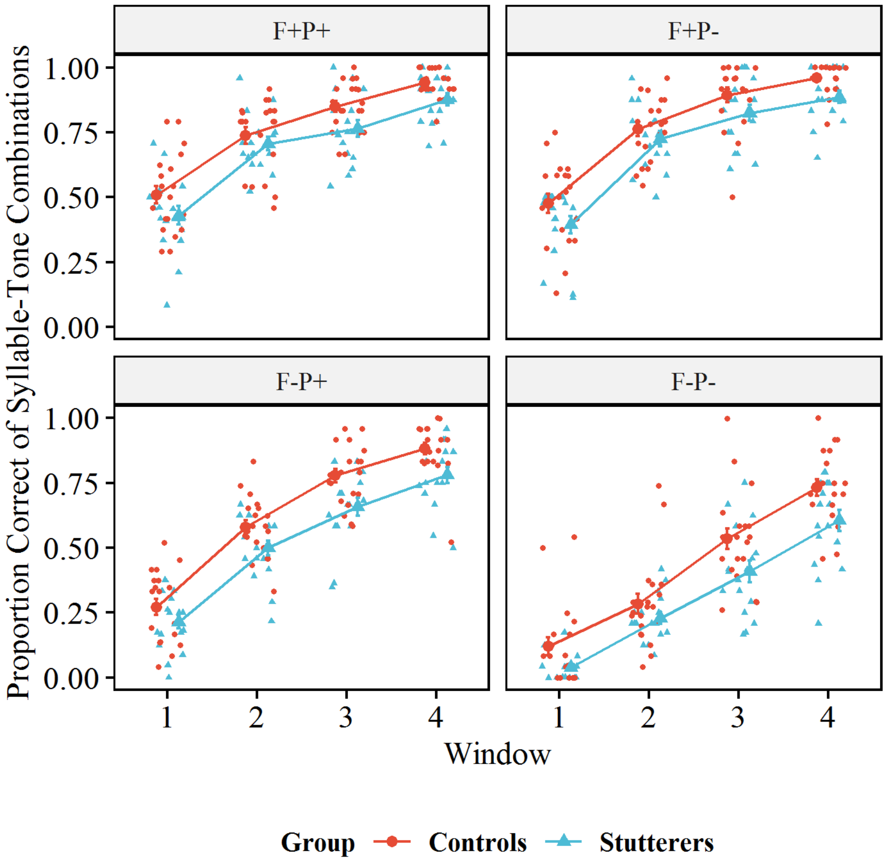

Gate 1 (onset)
Gate 2 (onset + 40 ms)
Gate 3 (onset + 80 ms)
Gate 4 (onset + 120 ms)
Gate 5 (onset + 160 ms)
Gate 6 (onset + 200 ms)
Gate 7 (onset + 240 ms)
Gate 8 (the full word)

Figure 2

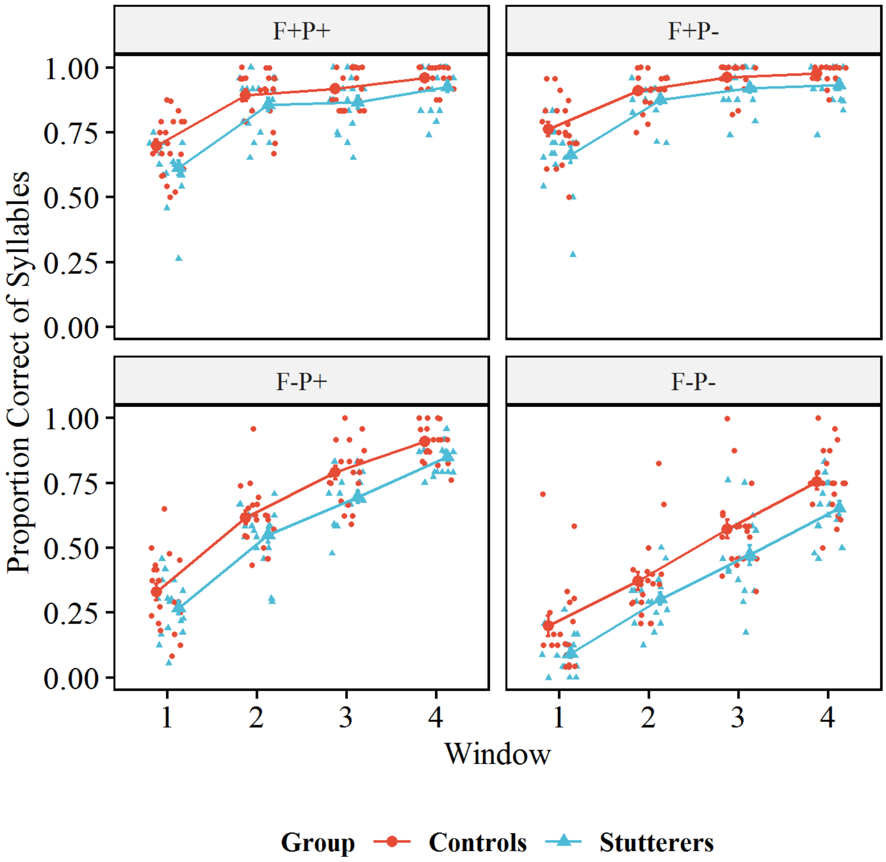


Tones
Syllables
Syllable-Tone Combinations
Proportion Correct
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8
Gate
Group
Controls
Stutterers

Figure 3


F+P+
F+P-
F-P+
F-P-
Proportion Correct of Syllable-Tone Combinations
1.00
0.75
0.50
0.25
0.00
1
2
3
4
Window
Group
Controls
Stutterers

984 Figure 4



985

986

Figure 5

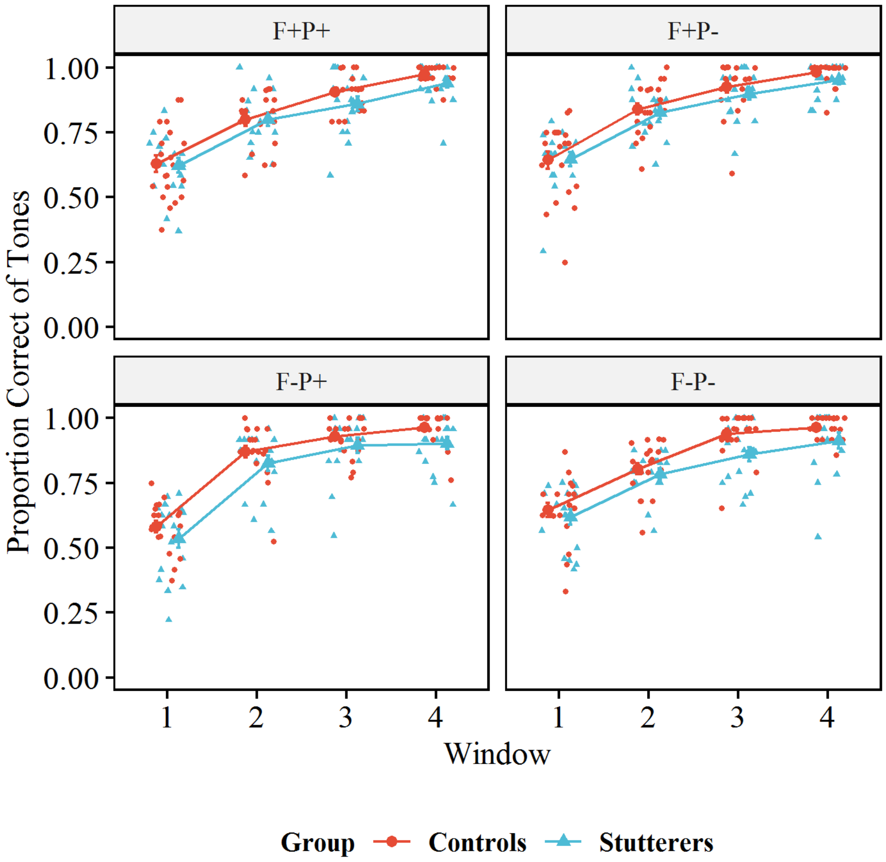
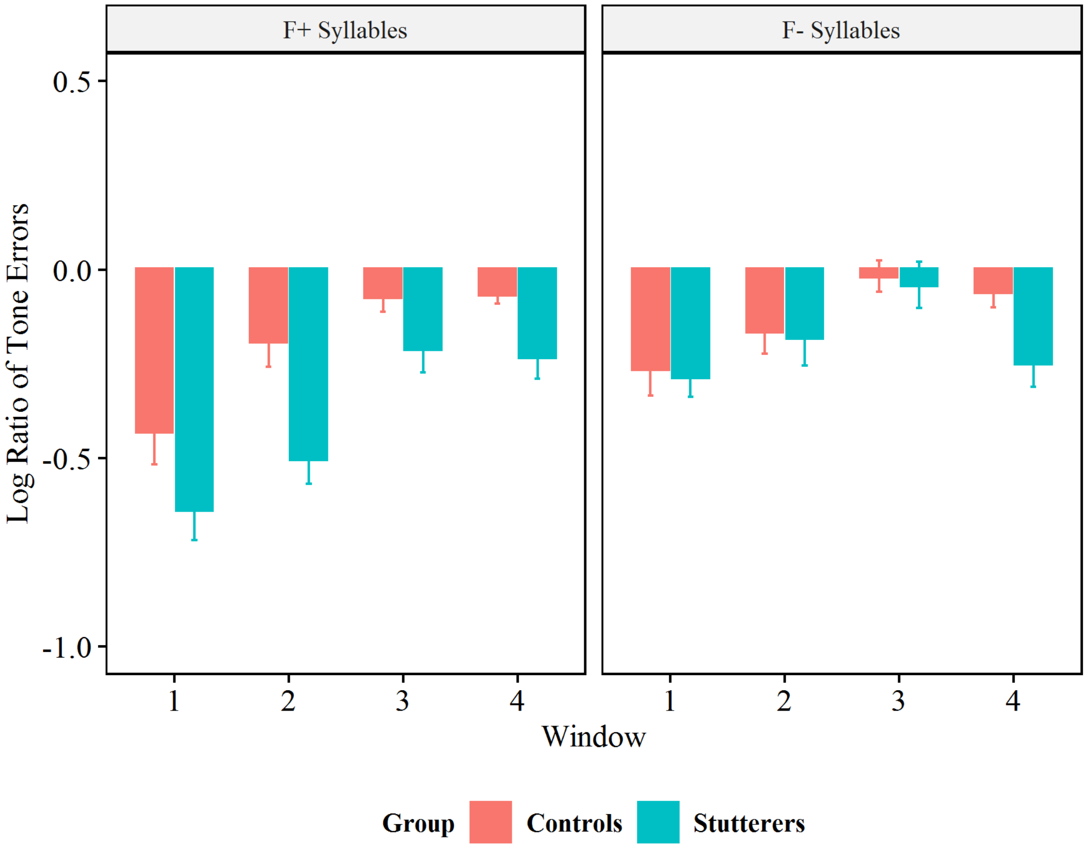

F+P+
F+P-
F-P+
F-P-
Proportion Correct of Tones
1.00
0.75
0.50
0.25
0.00
1
2
3
4
Window
Group
Controls
Stutterers

Figure 6


F+ Syllables
F- Syllables
Log Ratio of Tone Errors
0.5
0.0
-0.5
-1.0
1
2
3
4
1
2
3
4
Window
Group
Controls
Stutterers