Auditory-Motor Mapping Training Facilitates Speech and Word Learning in Tone-Language-Speaking Children with Autism: An Early Efficacy Study

Jinting Yan[1#], Fei Chen[2#, *], Xiaotian Gao[1], and Gang Peng[3]

[1] *College of Qiyue Communication & Cangzhou Research Centre for Child Language Rehabilitation, Cangzhou Normal University, Cangzhou, China*

[2] *School of Foreign Languages, Hunan University, Changsha, China;*

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

[#]The first two authors contribute equally to this study.

*** Corresponding author: Fei Chen**

School of Foreign Languages, Hunan University, Changsha, China.

E-mail addresses: chenfeianthony@gmail.com

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

Funding Statement: This work was in part supported by the Major Program of National Social Science Foundation of China (18ZDA293), the Natural Science Foundation of China overseas collaboration grant (31728009), the General Research Fund of the Research Grants Council of Hong Kong (15610321), and a key project from *Cangzhou Normal University* (xnjjw1907).

Abstract

Purpose: It has been reported that tone-language-speaking children with autism demonstrate speech-specific lexical tone processing difficulty although they have intact or even better-than-normal processing of nonspeech/melodic pitch analogues. In this early efficacy study, we evaluated the therapeutic potential of an Auditory-Motor Mapping Training (AMMT) in facilitating speech and word output for Mandarin-speaking nonverbal and low-verbal children with autism, in comparison with a matched non-AMMT-based control treatment.

Method: Fifteen Mandarin-speaking nonverbal and low-verbal ASD children participated and completed all the AMMT-based treatment sessions by intoning (singing) and tapping the target words delivered via an app, while another fifteen participants received control treatment. Generalized linear mixed-effects models were created to evaluate the speech production accuracy and word production intelligibility across different groups and conditions.

Results: Results showed that the AMMT-based treatment provided a more effective training approach in accelerating the rate of speech (especially lexical tone) and word learning in the trained items. More importantly, the enhanced training efficacy on lexical tone acquisition remained at two weeks post-therapy, and generalized to untrained tones that were not practiced. Furthermore, the low-verbal participants showed higher improvement compared to the nonverbal participants.

Conclusions: These data provide the first empirical evidence for adopting the AMMT-based training to facilitate speech and word learning in Mandarin-speaking nonverbal and low-verbal children with autism. This early efficacy study holds promise for improving lexical tone production in Mandarin-speaking children with autism but should be further replicated in larger-scale randomized studies.

Keywords: autism, nonverbal, low-verbal, Mandarin, lexical tones, Melodic Intonation Therapy

Auditory-Motor Mapping Training Facilitates Speech and Word Learning in Tone-Language-Speaking Children with Autism: An Early Efficacy Study

1 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder identified by a constellation of early-appearing social communication deficits and restricted, repetitive sensory-motor behaviours (American Psychiatric Association, 2013). The delayed/atypical language development – historically linked to a defining feature – has been removed from the diagnostic criteria and is now regarded as one of the co-occurring characteristics of autism. The goal of this study is to integrate recent advances in speech therapy to facilitate speech and word learning in tone-language-speaking children with ASD.

Nonverbal and Minimally Verbal Children with ASD and Behavioral Interventions

Estimates vary, with 25%–46% of children with ASD reported as minimally verbal past the age of 5 years (Norrelgen et al., 2015; Rose et al., 2016; Tager-Flusberg & Kasari, 2013), meaning that they have an expressive vocabulary fewer than 20 (Kasari et al., 2013), or 30 intelligible words (Bal et al., 2016; Plesa-Skwerer et al., 2016). Besides, some individuals with ASD even remain nonverbal with a complete absence of functional speech, and lack the ability to communicate with others using spoken language (Klinger et al., 2002; Koegel et al., 2009; Turner et al., 2006). While children with ASD showed deficits in social communication and social interaction, this is further exacerbated in nonverbal and minimally verbal children with ASD due to their severely delayed speech development. Increased speech and word output is considered a positive prognostic indicator of outcomes for nonverbal or minimally verbal children with ASD (Lord et al., 2006). Recently, the growing availability of computers and smartphones/tablets has generated a great deal of enthusiasm for their potential to help ameliorate speech deficits in minimally-verbal children with ASD, such as the computer-assisted 3-D virtual pronunciation tutor (Chen et al., 2019), and the Proloquo2Go iPad app (King et al., 2014). In consideration of autistic features especially among nonverbal and low-verbal individuals who begin treatment with limited or no spoken words, available behavioral interventions often tried to present speech sounds with computerized “high-tech” solutions (Tager-Flusberg & Kasari, 2013).

Mandarin Chinese as a Tonal Language and the State of the ASD Research in China

A very recent review estimated that ASD prevalence in China was comparable to the Western world (about 108 per 10,000) using standardized case identification protocol (Sun et al., 2019). However, the situation of ASD in China lags considerably behind those in the West in terms of public awareness, education opportunities, and life outcomes of people with autism (Liu et al., 2016; Yu et al., 2020). It is compelling to come up with a language-specific training approach for children with ASD who live in a country hosting nearly 20% of the world’s population. Modern Mandarin is a tone language with a relatively simple syllable structure, consisting of initials (consonants), finals (including monophthong, diphthong, triphthong, or nasal finals with a vowel nucleus), as well as lexical tones (Chen et al., 2017). Specifically, Mandarin Chinese is the official and widely spoken language in China and is also used in some other countries/regions, which phonologically differs from English. For example, Mandarin is a syllable-timed tone language that exploits variations in both pitch height and pitch direction at the syllable level to distinguish lexical meanings (Figure 1), while English is a stress-timed non-tone language (Mok & Dellwo, 2008). There are four citation tones in Mandarin Chinese: high-level Tone 1 (T1, [55]), mid-rising Tone 2 (T2, [35]), low-falling-rising Tone 3 (T3, [214]), and high-falling Tone 4 (T4, [51]). These four lexical tones are essential elements of Mandarin speech sounds, and are used to differentiate lexical meanings. For instance, “i” spoken with the four distinct tones can respectively mean “doctor” (T1), “move” (T2), “rely on” (T3), and “easy” (T4). Thus, changing the lexical tone in a tone language has the same kind of effect as changing a vowel or a consonant. In contrast, variation in pitch contours in non-tone languages mainly conveys different moods or intonations without changing the word content (Wang, 1973). Thus, speech therapy in tone-language-speaking children with ASD should not only aim at enhancing consonant and vowel production (segmental elements of speech), but additionally, also aim at improving lexical tone production (supra-segmental element of speech).

[Figure 1 about here]

Superior Music and Nonspeech Processing but Impaired Lexical Tone Processing in ASD

Music is one of the most meaningful and popular forms of nonspeech sound; like speech, it has developed to take advantage of the efficiencies of the human auditory system (Baldwin, 2012). In the research into auditory processing, several studies demonstrated that children with ASD prefer musical or nonspeech stimuli over speech and attend to them more (Dawson et al., 1998; Kuhl et al., 2005). Many children with ASD showed enhanced capacity of music processing and exhibited strong interest in learning and making music (Buday, 1995; Hairston, 1990; Heaton et al., 1998). Irrespective of the tone inventory (Mandarin or Cantonese), the enhanced or at least preserved non-linguistic pitch perception skill in ASD generalized to those from a tone language background (Cheng et al., 2017; Wang et al., 2017; Yu et al., 2015). Nevertheless, such an enhanced acoustic pitch perception skill was conversely changed to the compromised perception of lexical tones when the pitch changes are phonologically relevant for the tone language speakers with ASD (Chen et al., 2016; Lau et al., 2020; Wang et al., 2017; Yu et al., 2015). In terms of speech production, when compared with age-matched typically-developing (TD) children, the ASD group aged 3–6 years showed an apparent speech delay in the production of Mandarin initials, finals, as well as lexical tones (Wu et al., 2020). Since both music notes and lexical tones share the same psychoacoustical attribute of pitch information, it would be reasonable to take advantage of the relative strength of musical skills to compensate for the relative weakness of speech sound, especially lexical tone acquisition, for tone-language-speaking children with ASD.

The Intonation-Based Interventions in ASD

Given the behavioral resemblance between singing and speaking as well as neural overlap in responses to speech and musical stimuli (Peretz et al., 2015), researchers have begun to examine the therapeutic effects of singing/intonation, and how it can potentially ameliorate some of the speech deficits associated with neurological disorder such as ASD (Wan et al., 2010). Two earlier case studies have described the positive role of singing (intoned rather than spoken verbal stimulus) in facilitating speech development of children with autism (Hoelzley, 1993; Miller & Toca, 1979). Recently, Melodic Intonation Therapy (MIT; Albert et al., 1973; Sparks et al., 1974), initially designed for improving spoken language in left-hemisphere stroke patients with severe nonfluent aphasia, has been introduced into speech therapy in ASD. The MIT approach involves the musical elements of both melody and rhythm through the use of pitched vocalization or singing in combination with left-hand rhythmic tapping to provide cueing for syllable production (Norton et al., 2009). An modified version of MIT, called Auditory-Motor Mapping Training (AMMT), was initially proposed by Wan et al. (2011) to facilitate speech output for English-speaking nonverbal children with autism. The intonation-based

AMMT combines intonation (singing) of bi-syllabic words or phrases and the use of a pair of tuned drums to activate bimanual motor activities. All six English-speaking nonverbal children with autism who participated in Wan et al.'s (2011) study showed noticeable improvements in their ability to articulate several word approximations. Furthermore, the efficacy of AMMT has been further corroborated in English-speaking minimally verbal children with autism (Chenausky et al., 2016) and one more-verbal child with autism (Chenausky et al., 2017) when compared with a control treatment. In this study, we apply and assess the effect of intonation-based AMMT, with proven efficacy as an intervention for English-speaking children with ASD (Chenausky et al., 2016, 2017; Sandiford et al., 2013; Wan et al., 2011), on Mandarin-speaking children with ASD.

To this end, the present study evaluated the therapeutic potential of AMMT in facilitating speech output for tone-language-speaking children with ASD. Moreover, as suggested by Wang (1978), children do not learn speech by acquiring units like phonemes or allophones, but rather by gradually adding lexical items to their repertoire. Two important studies (Ferguson & Farwell, 1975; Hsieh, 1972) investigated the development of phonological production in relation to the acquisition of words, and showed that there was a primacy of word learning during speech development. Thus, in this study, our intervention based on a smartphone/iPad app called Music-Mediated and Lexicon-Integrated (MMLI) training, tries to combine speech and word learning. The current training study aimed to assess the efficacy of AMMT-based MMLI in comparison to a matched non-AMMT-based control treatment, Speech Repetition Therapy (SRT) (Chenausky et al., 2016). Specifically, this study aimed to address the following research questions:

a) Over the intensive treatment sessions of MMLI, would the nonverbal and low-verbal Mandarin-speaking children with ASD show any improvement in speech and word productions?
b) Compared with the control group, would the MMLI group show a greater improvement in the speech production accuracy of initials, finals, and lexical tones, as well as the word production intelligibility?
c) If significant improvements were observed in MMLI, could the benefits be retained after the cessation of daily treatment sessions and generalize to untrained items?

2 Methods

2.1 Participants

Upon entering the treatment sessions, 30 participants with autism were randomly assigned to one of two treatment groups. The MMLI group (n = 15, two girls) represented the experimental group, and the SRT group (n = 15, two girls) acted as the active control group. They were recruited from *Cangzhou Research Centre for Child Language Rehabilitation*. Permission to conduct this study was obtained from by the local institutional review board of the *Hong Kong Polytechnic University*, ensuring appropriate adherence to informed consent procedures. The parents provided informed, written permission for their children to participate before any assessments were administered.

The clinical diagnosis of ASD was established according to the DSM-5 criteria for ASD (American Psychiatric Association, 2013), and further confirmed (16 out of 30 participants) using the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012) by pediatricians and child psychiatrists in child hospitals. The clinical diagnosis of ADOS-2 was not conducted at the time of data collection, and the average time between the study procedures and the diagnosis was 8.14 months (*SD* = 5.29 months). Since the Mandarin version of ADOS-2 has not been officially validated and widely adopted in China (Sun et al. 2013; Yu et al., 2015), for those cases where an ADOS report was absent (14 out of 30 participants), we confirmed diagnoses using the Chinese version of the Gilliam Autism Rating Scale–Second Edition (GARS-2; Gilliam, 2006) under the supervision of a licensed clinical psychologist. Other inclusion criteria were: (1) the ability to imitate two piano notes; (2) the ability to sit in a chair and take part in instructed activities for around 15 minutes at a time; (3) the ability to imitate gross motor activities such as clapping hands, and imitate oral motor movements; (4) not having the following comorbidities: cerebral palsy or tuberous sclerosis, hearing/sight impairment, Down's syndrome, uncontrolled seizures, and organic impairment of oral or laryngeal structures (Chenausky et al., 2016; Wan et al., 2012). An additional 11 children with autism were found to be ineligible and were excluded from the current study because they either did not meet the four inclusion criteria or could not regularly attend all the required treatment sessions.

Furthermore, 30 TD children were recruited from one local kindergarten to investigate the speech, language and cognitive levels of our participants with autism before training in reference to age-matched neuro-typical ones. The TD children did not participate in the training sessions. Although the TD children were not specially screened for autism symptoms via standardized instruments, they met none of the diagnostic DSM-5 criteria for ASD from an interview with their parents or teachers. As shown in Table 1, the participants with ASD did show significant speech, language and cognitive delays compared to age-matched TD children. The children with autism were divided into two subgroups in the current study: Nonverbal status was defined as having complete absence of intelligible words before training; low-verbal status was defined as using expressive vocabulary of no more than 50 words, based on parent reports as well as the pretest measures. The term “low-verbal” was used in several previous studies (Yoder & Stone, 2006; Kasari et al., 2008) to describe individuals with ASD with severe speech, language and cognitive delays.

[Table 1 about here]

2.2 Study Design

2.2.1 Language and Cognitive Evaluation

A randomized control design was used in an effort to determine the effectiveness of the experimental treatment (MMLI) and control for various external factors. Prior to training, all participants were firstly assessed with their language ability (Ning, 2013), then nonverbal IQ using the *Primary Test of Nonverbal Intelligence* (PTONI, Ehrler & McGhee, 2008), and then working memory (Millman & Mattys, 2017). The two treatment groups did not differ from each other in terms of chronological age, language ability, nonverbal IQ, as well as working memory (Table 2).

[Table 2 about here]

Language Ability: The overall language ability (Chen et al., 2017; Ning, 2013) was evaluated for each Mandarin-speaking participant, which consists of five subtests (including Test of Mandarin Grammar, Word Definition Test, Rapid Automatized Naming, Narrative Test, and Sentence Comprehension Test). These subtests evaluated both language comprehension and language expression, and aimed to assess different aspects of language abilities such as phonology, lexicons, grammar, and semantics. The administration time is around 30 minutes.

Primary Test of Nonverbal Intelligence: PTONI (Ehrler & McGhee, 2008) is a theoretically sound, research-based method of assessing reasoning abilities in young children aged from 3:0 to 9:11. The PTONI was normed on a culturally diverse demographic sample from diverse language backgrounds. Testing takes approximately 15 minutes.

Working Memory: In order to evaluate the short-term phonological working memory (Millman & Mattys, 2017), we administered a forward digit span task, In the task, a series of numbers were played to participants auditorily and they were asked to repeat them immediately. For each digit length (two to nine digits), there were two separate items. The response for each item was regarded as correct and awarded 0.5 point only when the participants could correctly repeat every digit in the right order. The full score for the test of forward digit span is 8.

2.2.2 Probe Assessments

For low-verbal participants, probe assessments were conducted in the Pretest (Test 1), Midtest (Test 2), Immediate Posttest (Test 3), and Delayed Posttest (Test 4). After the first 12 treatment sessions, three nonverbal children with ASD from each treatment group (six in total) demonstrated no progress at all, and still remained nonverbal. They received the second-round 12 treatment sessions in the same modality as the first 12 sessions, with 24 treatment sessions in total (Figure 4). The change in protocol was not determined a priori. We decided to conduct more training sessions for these nonverbal participants, in an effort to figure out whether the null training efficacy was due to the relatively short-term training duration or due to the failure of the current approach. For those six nonverbal participants, probe assessments were performed not only in Tests 1–4, but also in the second-round Midtest (Test 5), and Immediate Posttest (Test 6).

During each probe assessment, the same picture naming task was utilized, with trained (Set 1) and untrained (Set 2) picture stimuli intermixed and presented in random order. Set 1 consisted of 60 lexical items that were presented during probe assessments and also practiced during therapy sessions. Set 2 included 12 high-frequency items, which were not practiced during treatment, but presented during the probes (see Supplemental Material S1). The untrained stimuli in Set 2 were used to assess the transfer of learning to untrained stimuli. The children were encouraged to make more than one attempt (at most three times in total) if they did not respond, or failed to produce the target word correctly, in an effort to guarantee that the spontaneous productions could truly reflect production capacity in children with autism. Researchers who administered the probe assessments were blind to treatment condition. However, it should be noted that during the probe assessments, no correct demonstrations were provided. The cueing sentence is "WHAT IS THIS?" without any information about the naming of the target pictures. The spontaneous productions from each child were recorded in a sound-isolated room for further analyses. As shown in Table 2, the two treatment groups did not differ from each other in the production accuracy of speech sounds (initials, finals, and tones) and word production intelligibility in the Pretest (i.e., at baseline assessment).

2.2.3 The Development of a Smartphone/iPad App

The app was developed based on the Ionic Framework (https://ionicframework.com/docs), which uses a user interface (UI) toolkit for building mobile and desktop apps via web technologies. It is an open-source front-end framework for HTML5 hybrid mobile application to help developers use the same source code to generate app files for both Android and iOS-based platforms. The primary function of the current app is to assist children's speech learning as well as word learning. The home page presented six themes, including vegetables, fruits, animals, daily necessities, snacks, and toys (Figure 2a). Under each theme, a total of 10 high-frequency lexical items were chosen, resulting in 60 trained lexical items in total. These 60 trained words were disyllabic nouns relevant to children's early-acquired vocabulary. These trained items contained all the 21 initials, 39 finals, four citation tones as well as neutral tone, and T3 tone sandhi in Mandarin phonology. The untrained words included 12 high-frequency lexical items, which included early-developing ([m, n, t, t^h]), middle-developing ([k, k^h, $tɕ^h$, ɕ]), and late-developing Mandarin consonants ([ts, ts^h, s, tʂ, $tʂ^h$, ʂ]), monophthongs ([u, ɿ, y]), diphthongs ([uᴀ, ɑu, ou]), triphthongs ([iɑu, iou]), finals with a nasal coda ([an, iaŋ, ʊŋ, iŋ, əŋ]), as well as all the tonal categories in Mandarin. Please refer to Supplemental Material S1 for details.

[Figure 2 about here]

Then, for each word item, one corresponding picture was presented in the center of the interface as the visual cue, and two piano icons on the left and right sides (Figure 2b). The natural speech sounds of 60 trained words (120 syllables) were recorded from one female announcer with standard Mandarin pronunciation. To construct the piano-timbre nonspeech sounds, first, a piano note (C4) of 261 Hz frequency was created, then the level pitch tier was replaced with the pitch contour extracted from each syllable (120 syllables in total) using the Pitch-Synchronous Overlap Add implanted in Praat (Boersma & Weenink, 2016). In this way, the piano-timbre nonspeech sounds share the same pitch contours as those in natural speech sounds (Figure 3). All the piano-timbre nonspeech sounds were normalized to 500 ms, and equally for root-mean-square intensity level, at 70 dB SPL. During MMLI training, when tapping the icons from left to right, the piano-timbre nonspeech sounds that match the pitch contours of natural lexical tones for the first and second syllables would be played, one tap per syllable. As for the source code, please refer to https://github.com/introfei/VoiceTrain and for compiling, packaging, and uploading issues, please refer to https://github.com/introfei/Blog/issues/3 to see more details.

[Figure 3 about here]

2.2.4 Treatment Protocol

Therapists were female undergraduate students majoring in speech rehabilitation from the *Cangzhou Research Centre for Child Language Rehabilitation*, and were trained specifically to provide both MMLI and SRT training. If one therapist taught a child participant in the MMLI group, she would need to teach another child participant in the SRT group. They had an average of 1.73 years of experience in teaching children with autism prior to this study. Therapists firstly learned from the first author on how to perform and follow the treatment protocol. They were also required to make a practice of teaching other children with autism who were not participating in this study until they were familiar with the whole protocol. The treatment protocol of MMLI and SRT is shown in Table 3. They were conducted with intensive repetition in a highly structured environment. The SRT is designed to be similar in several respects to conventional forms of speech therapy, while lacking the key elements of MIT (Chenausky et al., 2016). While SRT also presents verbal stimuli through the app interface and contains the same steps and speech outputs as MMLI, in SRT the verbal stimuli are spoken, not intoned; and there is no bimanual hand tapping on piano icons. The video demonstrations of one trial for the two groups (MMLI and SRT) have been shown in Supplemental Material S2. To monitor the treatment fidelity, all treatment sessions were videotaped to evaluate therapists' adherence to the protocol. We reviewed five videotaped sessions (41.67 % of all treatment sessions) selected at random from each child (75 treatment sessions for the MMLI group, and 75 treatment sessions for the SRT control group). Each treatment session contains 30 trials (10 trials x 3 repetitions), and each trial contains five steps (Table 3). On all the 2,250 MMLI trials, therapists intoned the target word and tapped the piano icons, and neither of the 2,250 SRT trials were intoned or tapped. Furthermore, over a total of 11,250 steps assessed in MMLI group, 91 (0.81 %) had repeated steps and 44 (0.39%) had omitted steps. Over the total 11,250 steps assessed in SRT group, 103 (0.92 %) had repeated steps and 38 (0.34%) had omitted steps.

[Table 3 about here]

2.2.5 Training Procedure

The therapy sessions were conducted in clinical treatment rooms at *Cangzhou Research Centre for Child Language Rehabilitation*. The child participants in both MMLI and SRT groups received short-term intensive training – 12 treatment sessions 6 times per week, over a 2-week period. However, after the first round of training, three nonverbal children with ASD (out of the 15 participants) from each treatment group demonstrated no progress at all, and still remained nonverbal. These six nonverbal participants received the second-round 12 treatment sessions, with 24 treatment sessions in total (Figure 4). Each treatment session began with a warm-up stage, followed by 10 lexical trials (each trial was repeated three times) of one specific theme.

The three repetitions of each trial were blocked together, and all the five steps of each trial were repeated. The sequence of steps was the same and strictly followed the protocol regardless of the response from the child (i.e., if the child did not respond or did not pronounce the word correctly). Each session lasted about 50–60 minutes, including breaks, which occurred every ten to fifteen minutes, based on the child's stamina. The order of six training themes within one intervention phase was randomized using a Latin Square among participants. The training order in the second phase is a repetition of that in the first intervention phase for each child. While receiving MMLI or SRT, the child participants were not allowed to engage in any other speech therapy activities in regular school programs. As required in the consent form, parents agreed to withhold other interventions while testing an unproven intervention in order to satisfy scientific objectives.

[Figure 4 about here]

2.3 Outcome Measures

2.3.1 Speech and Word Production Measures

The recorded productions were transcribed offline by Mandarin-speaking transcribers who were totally blind to the current study design in order to minimize experimental bias. Before transcription, an expert majoring in phonetics picked out the best sample from each child's utterances if more than one attempt was produced for one specific target word. The criterion is to choose the one with higher mean accuracy in the productions of initial, final, and lexical tone. If the nonspeech-like vocalization was produced for a certain target, the nonspeech-like vocalization would also be selected and sent to transcription. After selection, there were in total 3,634 words (7,268 syllables) produced by all the child participants from both training groups across all the probe assessments. The transcribers were told about the corresponding requirements in advance, but without being told about the nature and purpose of the current study. The transcribers needed to transcribe all the 3,634 words (7,268 syllables) and they were allowed to listen to the sound stimuli as many times as they wanted until they were confident to make a transcription. The order of presentation for the transcription task was randomized. Each sound stimulus was transcribed once by each transcriber.

First, for speech production measure, the disyllabic word was split into two syllables, which were transcribed separately. The 7,268 produced syllables were randomized and further transcribed using the International Phonetic Alphabet (IPA) by another five trained experts majoring in linguistics (mean age = 24.50 years). Especially for the tonal coding, the exact tonal categories were chosen from the following descriptions: high-level tone (T1), mid-rising tone (T2 or full sandhi: *T3 [35]), dipping tone (typical T3), high-falling tone (T4), low-falling tone (half-T3: *T3 [21]), and the neutral tone. Such stringent measures of phonetic transcription using IPA were aimed to assess the actual correct production of phonemes (Munro & Derwing, 1995) in a more fine-grained and precise manner. If none of the Mandarin phonology matched the transcription, the transcribers would log “none” instead. Each coder needed around 40 hours in total to complete the transcription of all the initials, finals, and tones. Three outcome measures were included to evaluate the speech production accuracy: Percent Initials Correct, Percent Finals Correct, and Percent Tones Correct, which were calculated with the number of correctly transcribed initials, finals, and lexical tones divided by the total number of syllables, respectively.

Second, in terms of word production measure, the 3,634 produced words were randomized and presented to five native speakers of Mandarin not majoring in linguistics (mean age = 21.38 years) with the E-Prime 2.0 program (Psychology Software Tools Inc., USA). The transcribers were not told about the target words before transcription. They were asked to write down each word they heard with two Chinese characters (each Chinese character representing a morpheme in the Mandarin word) one by one in a spreadsheet to evaluate word production intelligibility (i.e., how well the word is understood regardless of if all phonemes were accurately produced; transcribers were allowed to take their best guess). When none of the Chinese characters were suitable to be used to make a transcription, transcribers would log "none" instead. Each coder needed around 15 hours to complete the transcription of words. The number of correctly coded characters (morphemes) was divided by the total number of characters (morphemes) to yield "Percent Morphemes Correct".

To assess agreement among five raters, Kendall's Concordance Coefficient W was calculated in this study (Legendre, 2005). The interrater reliability with Kendall's coefficients of 0.802 for initial transcription, 0.793 for final transcription, 0.828 for tone transcription, and 0.884 for morpheme coding was reached, exhibiting relatively high inter-rater reliability.

2.3.2 User Experience

The user experience evaluation (Chen et al., 2019) was executed after the completion of all the treatment sessions, which was rated using a 5-point Likert scale based on the child's training performance (5 – the highest degree, 1 – the lowest degree). Each corresponding therapist rated the user experience for her own student. Such subjective observations evaluated the ways in which children with ASD approached different training methods. The user experience evaluation included five aspects: enjoyment, cooperation, consistency, interest, and motivation. Enjoyment refers to the degree of apparent pleasure in the learning process; cooperation means the degree of collaboration in learning a trial (whether the child could follow all the five steps within a trial); consistency indicates the continuity of the overall coordination throughout the learning process; interest refers to the degree of apparent interest in the training materials; motivation represents the degree of the apparent initiative before treatment sessions (whether the child appears to want to participate in training). The operational definitions of five aspects are shown in Supplemental Material S3.

2.4 Statistical Analyses

All the statistical analyses of outcome measures were performed in R (R Core Team, 2014). For the analyses of production accuracy, the generalized linear mixed-effects models (GLMMs) were created using the lme4 package (Bates et al., 2014). It is feasible for GLMM in R to include all the item responses transcribed from five different transcribers, and have them calculated within a single statistical model. In each GLMM, *treatment group* (MMLI vs. SRT), *test,* and their two-way interaction acting as fixed effects. When fitting GLMMs, *participant* and *item* were included as random effects. The R code for the full model: Accuracy ~ treatment group * test + (1+ test | participant: treatment group) + (1+ treatment group * test | item). By-participant and by-item random intercepts and random slopes for all possible fixed factors were included in the full model (Barr et al., 2013), which was compared with a simplified model that excluded a specific fixed factor using the ANOVA function in lmerTest package (Kuznetsova et al., 2017). Moreover, post-hoc pairwise comparisons were calculated with the lsmeans package (Lenth, 2016) with Tukey adjustment.

For the analyses of user experience, a generalized Poisson regression model (Consul & Famoye, 1992) was constructed in R using a glmer function (family = 'poisson'), with *treatment group* (MMLI vs. SRT), *aspect* (motivation, consistency, interest, cooperation, and enjoyment), and their two-way interaction acting as fixed effects. The generalized Poisson regression model is often a first-choice model for counts-based datasets, and has been found useful in fitting over-dispersed as well as under-dispersed count data. Given that the nonverbal and low-verbal participants in this study received different amounts of treatment sessions, their results were reported separately.

3 Results

3.1 Outcomes in Nonverbal Participants

There were six nonverbal participants with ASD (*G101, G102,* and *G103* in MMLI group; *G201, G202,* and *G203* in SRT group), who received 24 treatment sessions across four intervention phases in total (Figure 4). The probe assessment data were collected six times before, during, and after therapy. Only one participant with autism from the MMLI group (*G103*) began to acquire some initials, finals, lexical tones, as well as words in the trained items during the second-round training, while all the other five participants remained nonverbal even after 24 treatment sessions. Furthermore, none of the six nonverbal participants showed any improvement in the untrained items after training. For the evaluation of user experience, the subject from the MMLI group (*G103*) who showed gains in speech and word learning of trained items also obtained relatively higher scores of user experience, especially in the aspect of enjoyment.

3.2 Outcomes in Low-Verbal Participants

In total, there were 24 low-verbal participants with ASD (n = 12 in each treatment group) who received 12 treatment sessions across two intervention phases (Figure 4). Figure 5 shows the percentage of correct productions from two treatment groups in both trained and untrained items. The x-axis represents the probe assessment sessions and the y-axis stands for the percentage of correct initials, finals, lexical tones, and morphemes, respectively, from left to right.

[Figure 5 about here]

3.2.1 Production Accuracy of Initials

The GLMM was performed on the production accuracy of initials in trained stimuli, and the statistical results only showed a significant main effect of *test* (χ^2 (3) = 1143.70, $p < .001$). However, the GLMM did not reveal the significant main effect of *treatment group* (χ^2 (1) = 1.33, $p = .468$) nor the interaction effect of *treatment group* * *test* (χ^2 (3) = 3.35, $p = .341$). Further examination on the effect of *test* implied that the low-verbal participants from both groups showed significant improvement in producing initials in the trained items (all *p*s < .001), at Midtest, Immediate Posttest, as well as at Delayed Posttest. Furthermore, the results showed that the treatment methodology (MMLI vs. SRT) did not lead to outcome differences in the production accuracy of initials in the trained stimuli across all the probe assessments (Figure 5a & Table 4a).

[Table 4 about here]

Then, the GLMM on the accuracy of initials in the untrained items also merely showed a main effect of *test* (χ^2 (3) = 56.01, $p < .001$), while the main effect of *treatment group* (χ^2 (1) = 0.05, $p = .818$) and the interaction of *treatment group* × *test* (χ^2 (3) = 0.51, $p = .918$) were not significant. For both MMLI and SRT groups, as shown in Figure 5b, the number of correctly produced initials of untrained stimuli increased significantly after the whole 12 treatment sessions at Immediate Posttest (β = -0.39, SE = 0.07, $t = -6.03$, $p < .001$) and at follow-up assessment at Delayed Posttest (β = -0.34, SE = 0.07, $t = -5.22$, $p < .001$). Moreover, the two groups performed similarly in the production accuracy of initials in the untrained items across all the probe assessments (Figure 5b & Table 4b).

3.2.2 Production Accuracy of Finals

For the trained items in Set 1, the GLMM model on the accuracy of finals showed a significant main effect of *test* (χ^2 (3) = 1192.78, $p < .001$), while the main effect of *treatment group* (χ^2 (1) = 0.03, $p = .861$) was not significant. There was a significant two-way interaction of *treatment group* × *test* (χ^2 (3) = 17.61, $p < .001$). Compared with Pretest, both treatment groups made significant progress in the trained finals over the course of treatment (all *p*s < .001). Furthermore, the two training groups differed after two intervention phases (Figure 5a & Table 5a), with the MMLI group showing a higher production accuracy of trained finals than the matched control group at Immediate Posttest (β = 0.18, SE = 0.04, $t = 4.88$, $p < .001$), while no group differences were found at the other probe assessments (all *p*s > .05).

[Table 5 about here]

For the untrained items in Set 2 (Figure 5b), the GLMM on production accuracy of finals only reveal a main effect of *test* (χ^2 (3) = 79.11, $p < .001$). Neither the main effect of *treatment group* (χ^2 (1) = 0.18, $p = .675$) nor the interaction of *treatment group* × *test* (χ^2 (3) = 3.06, $p = .382$) was significant. The post hoc analyses demonstrated that both treatment groups showed significant progress in number of correctly produced finals in untrained items only at the immediate posttest (β = -0.59, SE = 0.07, $t = -7.97$, $p < .001$). Moreover, the two groups performed similarly in the production accuracy of finals in the untrained items (Figure 5b & Table 5b).

3.2.3 Production Accuracy of Tones

The GLMM on the accuracy of trained tones showed significant main effects of *treatment group* (χ^2 (1) = 334.59, p < .001) and *test* (χ^2 (3) = 1493.43, p < .001), as well as a significant two-way interaction of *treatment group* × *test* (χ^2 (3) = 112.87, p < .001). In comparison to the performance in Pretest (Figure 5a & Table 6a), both MMLI and SRT groups showed significant improvement at tone production of the trained items over the course of treatment (all ps < .001). Nevertheless, the growth rate was quite different, with participants who received MMLI training showing much higher production accuracies of trained tones compared with those receiving SRT, at Midtest (β = 0.22, SE = 0.04, t = 5.62, p < .001), Immediate Posttest (β = 0.61, SE = 0.04, t = 16.26, p < .001), as well as Delayed Posttest (β = 0.45, SE = 0.04, t = 12.11, p < .001).

[Table 6 about here]

Then, GLMM was performed on the accuracy of tones in untrained items, and the statistical results exhibited significant main effects of *treatment group* (χ^2 (1) = 40.85, p < .001) and *test* (χ^2 (3) = 85.20, p < .001), as well as a significant interaction of *treatment group* × *test* (χ^2 (3) = 14.27, p < .01). The low-verbal participants from the MMLI group had significant progress in production of lexical tones in untrained items, at Immediate Posttest and Delayed Posttest (ps < .001), while those from the SRT group showed no progress over the course of treatment (all *p*s > .05). In terms of group difference (Figure 5b & Table 6b), the experimental group of MMLI obtained much higher accuracy of tones in untrained stimuli after 12 treatment sessions at Immediate Posttest ($\beta = 0.42$, SE = 0.12, $t = 3.41$, $p < .001$) and two weeks later at Delayed Posttest ($\beta = 0.49$, SE = 0.13, $t = 3.90$, $p < .001$).

3.2.4 Word Production Intelligibility

First, for the trained items, the GLMM on word production intelligibility (Percent Morphemes Correct) showed significant main effects of *treatment group* (χ^2 (1) = 33.22, $p < .001$) and *test* (χ^2 (3) = 1611.81, $p < .001$), as well as a significant interaction of *treatment group* × *test* (χ^2 (3) = 38.65, $p < .001$), indicating that the two training groups showed different trajectories of word learning in the trained items. As shown in Figure 5a, compared to the baseline performance in Pretest, both MMLI and SRT groups showed noticeable improvements in trained word production (all *p*s < .001) when tested at Midtest, Immediate Posttest, and follow-up assessment two weeks later. In terms of group difference at different timepoints (Table 7a), the two treatment groups performed similarly on Percent Morphemes Correct in the Pretest ($\beta = -0.015$, SE = 0.038, $t = -0.40$, $p = .688$) and Midtest ($\beta = 0.011$, SE = 0.036, $t = 0.29$, $p = .772$), whereas the MMLI group produced higher accuracy in the trained morphemes than the control SRT group in the Immediate Posttest ($\beta = 0.247$, SE = 0.035, $t = 7.03$, $p < .001$) as well as Delayed Posttest ($\beta = 0.103$, SE = 0.035, $t = 2.94$, $p < .01$).

[Table 7 about here]

Second, for the untrained items, the GLMM on word production intelligibility (Percent Morphemes Correct) revealed a significant main effect of *test* (χ^2 (3) = 128.84, $p < .001$), but the main effect of *treatment group* did not reach significance (χ^2 (1) = 0.07, p = .789). There was a significant interaction of *treatment group* × *test* (χ^2 (3) = 19.09, p < .001). Post-hoc analysis showed that compared to the performance in Pretest, the MMLI group produced more untrained morphemes correctly in Midtest, Immediate Posttest, and Delayed Posttest (ps < .05). Over the same probe assessments, however, the SRT group only produced more untrained morphemes in the Immediate Posttest (β = -0.54, SE = 0.10, t = -5.37, p < .001) right after 12 treatment sessions (Figure 5b & Table 7b). For the between-group difference, across all the probe assessments, the MMLI group performed similarly to the control SRT group in terms of % Morphemes Correct in the untrained words (all ps > .05).

3.2.5 User experience

The generalized Poisson regression model on scores of user experience for the low-verbal children with ASD did not show significant main effects of *aspect* (χ^2 (4) = 8.22, p = .084) and *treatment group* (χ^2 (1) = 3.02, p = .082). Moreover, the Poisson regression model did not reveal the significant interaction of *treatment group* × *aspect* (χ^2 (4) = 0.24, p = .993). Thus, despite the trend, the experimental group of MMLI (M_{MMLI} = 3.42) did not receive higher scores across different aspects of user experience compared with the control group of SRT (M_{SRT} = 2.65).

4 Discussion

The current MMLI training app aimed to facilitate speech and word learning in tone-language-speaking children with ASD. For the low-verbal participants, while both MMLI and SRT groups were found to be effective in enhancing production skills, results suggested a faster rate of improvement in the production of finals, lexical tones, and words in the trained items for the experimental MMLI group. Moreover, the advantage of MMLI training transferred to the untrained items in terms of lexical tone production. For the six nonverbal participants, however, only one from the MMLI group responded to treatment in the trained items, while others from both training groups showed no progress and remained nonverbal even after 24 treatment sessions. We will discuss these findings hereunder.

4.1 Potential Mechanisms Responsible for the Training Efficacy of MMLI

MMLI resulted in greater improvements than SRT in most of the outcome measures for the Mandarin-speaking participants, which largely corroborated the efficacy of the AMMT-based training approach, with well-proven efficacy as a treatment for English-speaking children with autism (Chenausky et al., 2016, 2017; Sandiford et al., 2013; Wan et al., 2011). The data reported here further proved that the two key elements of MIT – intonation and hand tapping – added greatly to MMLI's effectiveness in children from another language system. When tapping on the virtual piano presented through the app, the nonspeech sounds with piano timbre would be generated in an effort to mimic the music-making activities. In contrast with previous studies (Chenausky et al., 2016, 2017; Sandiford et al., 2013; Wan et al., 2011), our MMLI system did not utilize real musical notes, but was modified to match various pitch variations of lexical tones in Mandarin phonology. Hoelzley (1993) proposes that the unique timbre of musical instruments may increase motivation and attention in children with autism. The MMLI group seemed to enjoy the treatment more than the SRT group based on the clinical observation, but it was not supported by the statistics.

Furthermore, music making through bimanual tapping on the tuned piano icons is a multimodal activity that not only captures the attention in children with autism, but also possibly primes and integrates the bilateral sensorimotor networks with shared motor, auditory, and visual neural representations of the articulatory/hand movements (Bangert et al., 2006; Lahav et al., 2007). In particular, it is speculative that the arcuate fasciculus (AF), a fiber bundle that connects the auditory perceptual regions in the temporal lobe with the motor-related regions in the frontal lobe (Catani et al., 2005), might be developed or reconstructed through intonation and music-making activities (Wan et al., 2010, 2011). The AF was thought to play an important role in auditory-motor mapping (Chenausky et al., 2017), and might be responsible for the bidirectional mapping between speech articulation and acoustics (Leclercq et al., 2010), as well as facilitating new word learning especially in the left bundle (López-Barroso et al., 2013). Another neural substrate likely to be engaged during music making is the putative mirror neuron system (MNS). It has been suggested that dysfunctional MNS underlies some of the speech and language deficits in individuals with ASD (Iacoboni & Dapretto, 2006). The elements in MMLI training, such as imitation, hand tapping, and synchronization, might activate brain regions that overlap with MNS, thus highlighting the potential benefits of such sensorimotor training to facilitate expressive language in developmental disorder such as autism (Overy & Molnar-Szakacs, 2009).

4.2 Training Efficacy for Low-Verbal Children with ASD

While MMLI holds promise for improving speech learning in Mandarin-speaking children with ASD in general, the effectiveness of MMLI was unbalanced among different components of syllables (i.e., initials, finals, and tones). A shown in Figure 5a, low-verbal participants with ASD receiving MMLI started to show superiority over those receiving SRT in their ability to correctly articulate Mandarin lexical tones in trained items as early as Midtest, and such advantage further expanded after 12 treatment sessions at Immediate Posttest and was maintained at Delayed Posttest. For speech production of Mandarin finals which use vowel(s) as the whole final or as the nucleus, the low-verbal MMLI participants only experienced comparatively greater improvement than the SRT participants after 12 treatment sessions, and such advantage did not persist at the follow-up assessment. In terms of the speech production of Mandarin initials that were composed of consonants, the two treatment groups performed similarly in the trained items over all the probe assessments. Furthermore, for the untrained items (Figure 5b), MMLI produced significantly greater gains merely in the lexical tone learning in low-verbal children with ASD compared to the control therapy, SRT. Such generalization skills would be greatly beneficial to children with ASD, who show difficulty in transferring learned knowledge to a new context (Church et al., 2015; Happé & Frith, 2006). In a short conclusion, as observed from the current data, the efficacy of MMLI was much higher in the training of lexical tones, followed by vowels, and then consonants.

The greater improvement on lexical tone acquisition should not be surprising given that relative to SRT, MMLI presented participants with additional information of pitch contours embedded in piano-timbre nonspeech. Individuals with ASD have often demonstrated pitch or melodic processing superiority in various musical and nonspeech stimuli (e.g. Gomot et al., 2002; Heaton, 2005; O'Riordan & Passetti, 2006). Accumulating evidence pointed to a two-way transferability of pitch expertise across domains of music and speech (as lexical tones) in neuro-typical children and adults (Bidelman et al., 2013; Nan et al., 2018; Wong et al., 2007). By targeting the clinical population, the current data provided the first empirical evidence of using the relative strength of music, a ubiquitous nonspeech form, to compensate for the relative weakness of speech sounds, especially lexical tone acquisition for tone-language-speaking children with ASD. One recent study (Nan et al., 2018) demonstrated that the six months of piano training not only enhanced the lexical tone discrimination, but also improved vowel and consonant discrimination in 4- to 5-year-old Mandarin-speaking TD children, suggesting strengthened common sound processing across domains underlying the benefits of musical training. In this training study, however, we failed to detect the benefits of music-supported MMLI training on the acquisition of initials (consonants). On the one hand, the shorter, weaker, and more aperiodic consonants in speech sounds are likely to be impacted more in a co-occurring nonspeech background than the stable, periodic components of tones and vowels. On the other hand, since the syllable-initial consonants were acquired later than the vowels and tones in Mandarin-speaking TD children (Hua & Dodd, 2000), the relatively short-term training duration in this study may be another potential factor leading to the failure of transfer effects on the late-acquired consonant production.

4.3 Training Efficacy for Nonverbal Children with ASD

The nonverbal participants in our study belong to the subgroup of the autism spectrum with severe language/cognitive impairment. They were completely nonverbal, despite having received extensive speech therapy (four to 16 months) prior to recruitment. In the current study, except for one nonverbal child with ASD responding to the MMLI treatment, the other five nonverbal participants could not correctly produce even one trained word after 24 treatment sessions, and meanwhile, they received lower scores on user experience. It is unlikely that this is due to the stringent measure of phonetic transcription, since these five nonverbal participants did not even spontaneously produce any verbal attempts during probe assessments. In contrast, as reported in Wan et al. (2011), the English-speaking nonverbal participants with autism who received similar AMMT training began to vocalize some “word approximations” after 10-15 treatment sessions. It should be noted that the speech samples produced by the nonverbal participants in Wan et al.’s (2011) study were imitations rather than spontaneous productions, and should inform the degree and nature of progress. In our study, however, the spontaneous speech samples were collected from participants in a picture naming task without cueing or demonstration. Another possibility is that the nonverbal participants in our study were more severely impaired in terms of language and cognitive capacity compared to those in Wan et al.'s (2011) study. Given the extreme challenges these participants face, more treatment sessions or some other training approaches should be delivered to nonverbal children with ASD. In some cases, both parents and therapists have observed an increase in speechlike vocalizations during vocal play in daily life, which might be an anecdotal evidence for speech development in these nonverbal participants (Rvachew & Brosseau-Lapré, 2012).

4.4 Limitations and Clinical Implications

This study has several limitations. First, as with many other studies of autism, a limitation of the current training study is its small sample size, and replication in larger-scale randomized studies will be necessary. Second, more treatment sessions or some other training approaches should be applied to help nonverbal children with ASD improve their speech production skills. Third, considering the substantial heterogeneity of the autism spectrum, there are various types of speech disorder in ASD, such as motor speech disorder (dysarthria or apraxia), speech delay, or combination of these (Chenausky et al., 2019). More in-depth investigation of speech therapy in different subtypes would help determine whether MMLI is effective for all children with autism or whether it only works well for certain subtypes. Understanding these mechanisms will help tailor the interventions, select the most appropriate personalized treatment, and make predictions about prognosis. Fourth, applying the current MMLI training to some other tonal language speakers with ASD would be an important next step.

Taken together, the AMMT-based training program of MMLI, notwithstanding its limitations, provided an effective training approach in accelerating the rate of speech sound (especially lexical tone) and word production for Mandarin-speaking children with ASD. The languages of the world exhibit great diversity. Some estimates suggest that around 60–70% of the world's languages are tonal (Yip, 2002), and more than half of the world's population speak a tone language (Fromkin, 1978). Thus, there is a high demand for MMLI, which could be modified and applied to help some other tone-language-speaking children with autism beyond Mandarin-speaking ones to better acquire the phonological category of lexical tones. With respect to practical significance, the current MMLI approach is realized in the smartphone/iPad app, which is easily accessible and has the potential to be utilized remotely in the home environment as implemented by a parent or family member. This is important for speech therapy for children with autism from counties with a shortage of speech-language pathologists. Finally, the success of AMMT-based MMLI also lends support to the positive effects of music-supported treatments in individuals with ASD (James et al., 2015; Reschke-Hernández, 2011; Salomon-Gimmon & Elefant, 2019; Sharda et al., 2018).

5 Conclusion

This study compared the efficacy of MMLI, an AMMT-based treatment, with the control therapy in eliciting spoken language for tone-language-speaking children with autism. Relative to the control treatment, there was greater improvement in Mandarin-speaking children with ASD after they received MMLI training, in terms of lexical tone, final, and word learning in the trained items. Such enhanced training efficacy on lexical tone production persisted at two weeks post-therapy, and even generalized to untrained items that were not practiced. The results hold promise for the efficacy of MMLI to improve speech production in tone-language-speaking children with autism. Because the nonverbal and low-verbal children with autism had a very limited repertoire of speech sounds prior to treatment, the acquisition of speech sounds and words through MMLI is an important gain that provides a foundation for subsequent speech and language rehabilitation. More importantly, this study offers the first empirical evidence of the advantages of utilizing musical elements to facilitate lexical tone acquisition in the clinical population of ASD, which adds a new clinical perspective to our understanding of the close relationship between music and speech.

6 Acknowledgments

This work was in part supported by the Major Program of National Social Science Foundation of China (18ZDA293), the Natural Science Foundation of China overseas collaboration grant (31728009), the General Research Fund of the Research Grants Council of Hong Kong (15610321), and a key project from *Cangzhou Normal University* (xnjjw1907). We sincerely thank all the students from *College of Qiyue Communication, Cangzhou Normal University*, for their research assistance. We sincerely thank all the child participants from *Cangzhou Research Centre for Child Language Rehabilitation* and *Tuofu Kindergarten* in Cangzhou, and their parents, for their participation and cooperation.

References

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, *29*(2), 130–131. https://doi.org/10.1001/archneur.1973.00490260074018

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington, VA : American Psychiatric Publishing.

Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 57(12), 1424-1433.

Baldwin, C. L. (2012). *Auditory Cognition and Human Performance: Research and Applications*. CRC Press.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H.-J., & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*(3), 917–926. https://doi.org/10.1016/j.neuroimage.2005.10.044

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLOS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0060676

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (Version 6.0. 14) [Computer program]*. http://www.praat.org/

Buday, E. M. (1995). The effects of signed and spoken words taught with music on sign and speech imitation by children with autism. *Journal of Music Therapy*, *32*(3), 189–202. https://doi.org/10.1093/jmt/32.3.189

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8–16. https://doi.org/10.1002/ana.20319

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

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. https://doi.org/10.1017/S0305000916000581

Chen, F., Wang, L., Peng, G., Yan, N., & Pan, X. (2019). Development and evaluation of a 3-D virtual pronunciation tutor for children with autism spectrum disorders. *PLOS ONE*, *14*(1), e0210858. https://doi.org/10.1371/journal.pone.0210858

Chen, F., Yan, N., Pan, X., Yang, F., Ji, Z., Wang, L., & Peng, G. (2016). *Impaired categorical perception of Mandarin tones and its relationship to language ability in autism spectrum disorders*. Proc. Interspeech 2016, 233-237. https://doi.org/10.21437/Interspeech.2016-1133

Chenausky, K., Brignell, A., Morgan, A., & Tager-Flusberg, H. (2019). Motor speech impairment predicts expressive language in minimally verbal, but not low verbal, individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, *4*, 2396941519856333. https://doi.org/10.1177/2396941519856333

Chenausky, K., Norton, A., Tager-Flusberg, H., & Schlaug, G. (2016). Auditory-Motor Mapping Training: Comparing the Effects of a Novel Speech Treatment to a Control Treatment for Minimally Verbal Children with Autism. *PLOS ONE*, *11*(11), e0164930. https://doi.org/10.1371/journal.pone.0164930

Chenausky, K. V., Norton, A. C., & Schlaug, G. (2017). Auditory-motor mapping training in a more verbal child with autism. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00426

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch perception in tone language-speaking adults with and without autism spectrum disorders. *I-Perception*, 8(3), 2041669517711200. https://doi.org/10.1177/2041669517711200

Church, B. A., Rice, C. L., Dovgopoly, A., Lopata, C. J., Thomeer, M. L., Nelson, A., & Mercado, E. (2015). Learning, plasticity, and atypical generalization in children with autism. *Psychonomic Bulletin & Review*, *22*(5), 1342–1348. https://doi.org/10.3758/s13423-014-0797-9

Consul, P. C., & Famoye, F. (1992). Generalized poisson regression model. *Communications in Statistics – Theory and Methods*, *21*(1), 89–109. https://doi.org/10.1080/03610929208830766

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, *28*(6), 479–485. https://doi.org/10.1023/A:1026043926488

Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary test of nonverbal intelligence*. Austin, TX: Pro-Ed.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language*, 51, 419–439.

Fromkin, V. (1978). *Tone: A linguistic survey*. New York: Academic Press.

Gilliam, J. E. (2006). *Gilliam Autism Rating Scale: GARS 2*. Austin, TX: PRO-ED.

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: Electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(5), 577–584. https://doi.org/10.1017/S0048577202394058

Hairston, M. (1990). Analyses of responses of mentally retarded autistic and mentally retarded nonautistic children to art therapy and music therapy. *Journal of Music Therapy*, *27*(3), 137–150. https://doi.org/10.1093/jmt/27.3.137

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25.

Heaton, P. (2005). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 787. https://doi.org/10.1007/s10803-005-0024-7

Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception: An Interdisciplinary Journal*, *15*(3), 291–305. https://doi.org/10.2307/40285769

Hoelzley, P. D. (1993). Communication potentiating sounds: Developing channels of communication with autistic children through psychobiological responses to novel sound stimuli. *Canadian Journal of Music Therapy*, *1*, 54–76.

Hsieh, H. I. (1972). Lexical diffusion: Evidence from child language acquisition. *Glossa*, 6(1), 89–104.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern standard Chinese). *Journal of Child Language*, 27(1), 3–42.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–951.

James, R., Sigafoos, J., Green, V. A., Lancioni, G. E., O'Reilly, M. F., Lang, R., Davis, T., Carnett, A., Achmadi, D., Gevarter, C., & Marschik, P. B. (2015). Music therapy for individuals with autism spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders*, *2*(1), 39–54. https://doi.org/10.1007/s40489-014-0035-4

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. Journal of Consulting and Clinical Psychology, 76(1), 125–137. https://doi.org/10.1037/0022-006X.76.1.125

King, M. L., Takeguchi, K., Barry, S. E., Rehfeldt, R. A., Boyer, V. E., & Mathews, T. L. (2014). Evaluation of the iPad in the acquisition of requesting skills for children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *8*(9), 1107–1120. https://doi.org/10.1016/j.rasd.2014.05.011

Klinger, L., Dawson, G., & Renner, P. (2002). Autistic disorder. In *Child Psychopathology* (2nd ed., pp. 409–454). New York: Guilford Press.

Koegel, R. L., Shirotova, L., & Koegel, L. K. (2009). Brief report: Using individualized orienting cues to facilitate first-word acquisition in non-responders with autism. *Journal of Autism and Developmental Disorders*, *39*(11), 1587–1592. https://doi.org/10.1007/s10803-009-0765-9

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science*, *8*(1), F1–F12. https://doi.org/10.1111/j.1467-7687.2004.00384.x

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

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*(2), 308–314. https://doi.org/10.1523/JNEUROSCI.4822-06.2007

Lau, J. C., To, C. K., Kwan, J. S., Kang, X., Losh, M., & Wong, P. C. (2020). Lifelong Tone Language Experience does not Eliminate Deficits in Neural Encoding of Pitch in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 1-20. https://doi.org/10.1007/s10803-020-04796-7

Leclercq, D., Duffau, H., Delmaire, C., Capelle, L., Gatignol, P., Ducros, M., Chiras, J., & Lehéricy, S. (2010). Comparison of diffusion tensor imaging tractography of language tracts and intraoperative subcortical stimulations: Clinical article. *Journal of Neurosurgery*, *112*(3), 503–511. https://doi.org/10.3171/2009.8.JNS09558

Legendre, P. (2005). Species associations: The Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, *10*(2), 226. https://doi.org/10.1198/108571105X46642

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

Liu, Y., Li, J., Zheng, Q., Zaroff, C. M., Hall, B. J., Li, X., & Hao, Y. (2016). Knowledge, attitudes, and perceptions of autism spectrum disorder in a stratified sampling of preschool teachers in China. *BMC Psychiatry*, *16*(1), 142. https://doi.org/10.1186/s12888-016-0845-2

López-Barroso, D., Catani, M., Ripollés, P., Dell'Acqua, F., Rodríguez-Fornells, A., & Diego-Balaguer, R. de. (2013). Word learning is mediated by the left arcuate fasciculus. *Proceedings of the National Academy of Sciences*, *110*(32), 13168–13173. https://doi.org/10.1073/pnas.1301696110

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry*, *63*(6), 694–701. https://doi.org/10.1001/archpsyc.63.6.694

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule: ADOS–2*. Los Angeles, CA: Western Psychological Services.

Miller, S. B., & Toca, J. M. (1979). Adapted melodic intonation therapy: A case study of an experimental language program for an autistic child. *The Journal of Clinical Psychiatry*, *40*(4), 201–203.

Millman, R. E., & Mattys, S. L. (2017). Auditory verbal working memory as a predictor of speech perception in modulated maskers in listeners with normal hearing. *Journal of Speech, Language, and Hearing Research*, *60*(5), 1236–1245. https://doi.org/10.1044/2017_JSLHR-S-16-0105

Mok, P., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Speech Prosody 2008*, 423–426.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D. E., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*(28), E6630–E6639. https://doi.org/10.1073/pnas.1808412115

Ning, C. Y. (2013). *Test of language ability in Mandarin-speaking preschoolers*. Institute of Linguistics, Tianjin Normal University: Tianjin University Press.

Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., Gillberg, C., & Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, *19*(8), 934–943. https://doi.org/10.1177/1362361314556782

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy. *Annals of the New York Academy of Sciences*, *1169*(1), 431–436. https://doi.org/10.1111/j.1749-6632.2009.04859.x

O'Riordan, M., & Passetti, F. (2006). Discrimination in autism within different sensory modalities. *Journal of Autism and Developmental Disorders*, *36*(5), 665–675. https://doi.org/10.1007/s10803-006-0106-1

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*(5), 489–504. https://doi.org/10.1525/mp.2009.26.5.489

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140090. https://doi.org/10.1098/rstb.2014.0090

Plesa-Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with ASD. *Autism*, 20, 591–604.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

Reschke-Hernández, A. E. (2011). History of music therapy treatment interventions for children with autism. *Journal of Music Therapy*, *48*(2), 169–207. https://doi.org/10.1093/jmt/48.2.169

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, *60*(5), 464–477. https://doi.org/10.1111/jir.12284

Rvachew, S., Brosseau-Lapré, F. (2012). Developmental Phonological Disorders: Foundations of Clinical Practice. San Diego, CA: Plural Publishing.

Salomon-Gimmon, M., & Elefant, C. (2019). Development of vocal communication in children with autism spectrum disorder during improvisational music therapy. *Nordic Journal of Music Therapy*, *28*(3), 174–192. https://doi.org/10.1080/08098131.2018.1529698

Sandiford, G. A., Mainess, K. J., & Daher, N. S. (2013). A pilot study on the efficacy of melodic based communication therapy for eliciting speech in nonverbal children with autism. *Journal of Autism and Developmental Disorders*, *43*(6), 1298–1307. https://doi.org/10.1007/s10803-012-1672-z

Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., Tan, M., Nadig, A., & Hyde, K. (2018). Music improves social communication and auditory–motor connectivity in children with autism. *Translational Psychiatry*, *8*(1), 1–13. https://doi.org/10.1038/s41398-018-0287-3

Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, *10*(4), 303–316. https://doi.org/10.1016/S0010-9452(74)80024-9

Sun, X., Allison, C., Matthews, F. E., Sharp, S. J., Auyeung, B., Baron-Cohen, S., & Brayne, C. (2013). Prevalence of autism in mainland China, Hong Kong and Taiwan: a systematic review and metaanalysis. *Molecular Autism*, 4(1), 7. https://doi.org/10.1186/2040-2392-4-7

Sun, X., Allison, C., Wei, L., Matthews, F. E., Auyeung, B., Wu, Y. Y., Griffiths, S., Zhang, J., Baron-Cohen, S., & Brayne, C. (2019). Autism prevalence in China is comparable to Western prevalence. *Molecular Autism*, *10*(1), 7. https://doi.org/10.1186/s13229-018-0246-0

Tager‐Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Turner, L. M., Stone, W. L., Pozdol, S. L., & Coonrod, E. E. (2006). Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism*, *10*(3), 243–265. https://doi.org/10.1177/1362361306063296

Wan, C. Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., Norton, A., & Schlaug, G. (2011). Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. *PLOS ONE*, *6*(9), e25505.

Wan, C. Y., Demaine, K., Zipse, L., Norton, A., & Schlaug, G. (2010). From music making to speaking: Engaging the mirror neuron system in autism. *Brain Research Bulletin*, *82*(3), 161–168. https://doi.org/10.1016/j.brainresbull.2010.04.010

Wan, C. Y., Rüüber, T., Hohmann, A., & Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Perception: An Interdisciplinary Journal*, *27*(4), 287–295. https://doi.org/10.1525/mp.2010.27.4.287

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

Wang, W. S. Y. (1978). The three scales of diachrony. In B. B. Kachru (Ed.), *Linguistics in the Seventies: Directions and Prospects* (pp. 63–75). University of Illinois: Urbana, IL.

Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, *7*, 43254. https://doi.org/10.1038/srep43254

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872

Wu, H., Lu, F., Yu, B., & Liu, Q. (2020). Phonological acquisition and development in Putonghua-speaking children with autism spectrum disorders. *Clinical Linguistics & Phonetics*, 34(9), 844-860.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yoder P. & Stone W. L. (2006). A Randomized Comparison of the Effect of Two Prelinguistic Communication Interventions on the Acquisition of Spoken Communication in Preschoolers With ASD. *Journal of Speech, Language, and Hearing Research*, 49(4), 698–711. https://doi.org/10.1044/1092-4388

Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S., & Zhang, Y. (2015). Pitch processing in tonal-language-speaking children with autism: An event-related potential study. *Journal of Autism and Developmental Disorders*, *45*(11), 3656–3667. https://doi.org/10.1007/s10803-015-2510-x

Yu, L., Stronach, S., & Harrison, A. J. (2020). Public knowledge and stigma of autism spectrum disorder: Comparing China with the United States. *Autism*, 24(6) 1531–1545. https://doi.org/10.1177/1362361319900839

Tables

Table 1

Characteristics of children with ASD and age-matched TD children

	ASD (*n* = 30)		TD (*n* = 30)		*t*	*p*
	M	*SD*	*M*	*SD*		
Age (in months)	67.80	15.06	67.57	14.12	0.56	.578
Language Ability	31.67	26.84	88.07	8.81	-11.81	<.001
Nonverbal IQ	59.53	12.32	105.83	17.34	-13.78	<.001
Working Memory	3.90	4.39	12.23	3.65	-12.00	<.001
% Initials Correct	18.69	16.71	69.74	13.81	-14.23	<.001
% Finals Correct	17.60	15.59	66.31	13.74	-14.05	<.001
% Tones Correct	17.59	15.71	68.38	14.74	-14.24	<.001
% Morphemes Correct	22.68	20.99	73.88	14.73	-12.30	<.001

Table 2

Characteristics of participants with ASD in two treatment groups

	MMLI (*n* = 15)		SRT (*n* = 15)		*t*	*p*
	M	*SD*	*M*	*SD*		
Age (in months)	66.13	15.32	69.47	15.14	-0.63	.542
Language Ability	33.07	25.35	30.27	29.08	0.47	.643
Nonverbal IQ	60.53	13.86	58.53	10.97	0.65	.528
Working Memory	3.93	4.38	3.87	4.55	0.05	.962
% Initials Correct	18.49	17.81	18.90	16.17	-0.28	.785
% Finals Correct	17.88	17.05	17.32	14.60	0.22	.830
% Tones Correct	17.90	17.71	17.28	14.05	0.16	.872
% Morphemes Correct	22.44	23.48	22.92	19.02	-0.15	.887

Table 3

The warm-up stage at the beginning of each treatment session and the five-step structure of an MMLI trial vs. an SRT trial

	MMLI (Experimental Group)	**SRT (Control Group)**
Warming Up	Musical melodies without lyrics are played, and musical toys such as shaking maracas are introduced to facilitate their movements. Moreover, a rhythmic tapping of the foot is also used in time to music.	Playing checkboards without verbal or with minimally verbal instruction.
Steps	**MMLI Trial**	**SRT Trial**
1. Word Introduction	Therapist introduces the target word by showing a word picture (such as “tiger”) on the phone/iPad app and then **intoning (singing)** the word “[lɑu35 xu214]” by **tapping** the piano icons 1× per syllable.	Therapist introduces the target word by showing a word picture (such as “tiger”) on the phone/iPad screen and then **speaking** the word “[lɑu35 xu214]” without finger tapping.
2. Synchronous Production	Therapist produces target with the child. Therapist **intones and taps** “Let’s sing it together” and in synchrony with child “[lɑu35 xu214]”.	Therapist produces target with the child. Therapist **speaks** “Let’s speak it together” and in synchrony with child “[lɑu35 xu214]”.
3. Unison with Fading	Therapist and participant begin to **intone and tap** the target word together, but after the first syllable, the therapist stops while the child continues to intone and tap the next syllable. “[lɑu35] ____”.	Therapist and participant begin to **speak** the target word together, but after the first syllable, the therapist stops while the child continues to produce the next syllable. “[lɑu35] ____”.
4. Immediate Imitation	Therapist firstly **intones and taps** the target word alone. Afterwards, participant imitates the word, and therapist remains silent. “My turn first: [lɑu35 xu214]. Now your turn: _______”.	Therapist firstly **speaks** the target word alone. Afterwards, participant imitates the the word, and therapist remains silent. “My turn first: [lɑu35 xu214]. Now your turn: _______”.
5. Independent Production	The child is further encouraged to independently **intone and tap** the target word once again. “__________”	The child is further encouraged to independently **speak** the target word once again. “__________”

Table 4

Means (and standard deviations) of % Initials Correct for low-verbal participants by treatment group (MMLI and SRT) and probe assessment in (a) Trained Items, and (b) Untrained Items.

	Group	Pretest		Midtest		Immediate Posttest		Delayed Posttest	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
(a) Trained	MMLI	23.41	11.83	26.93	11.35	38.48	10.45	36.80	10.98
	SRT	24.03	13.27	27.92	12.90	38.03	11.51	38.34	12.63
(b) Untrained	MMLI	21.66	13.95	22.85	17.64	28.17	18.11	27.82	14.97
	SRT	21.94	14.47	22.45	18.62	28.42	15.18	26.91	12.55

Table 5

Means (and standard deviations) of % Finals Correct for low-verbal participants by treatment group (MMLI and SRT) and probe assessment in (a) Trained Items, and (b) Untrained Items.

	Group	Pretest		Midtest		Immediate Posttest		Delayed Posttest	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
(a) Trained	MMLI	22.93	12.92	25.54	14.02	37.98	12.61	35.60	12.04
	SRT	22.42	14.28	26.25	13.37	34.21	13.55	33.78	14.02
(b) Untrained	MMLI	19.54	14.31	20.36	14.62	27.71	17.58	23.51	15.46
	SRT	18.03	16.93	18.57	15.18	24.89	14.37	19.08	13.89

Table 6

Means (and standard deviations) of % Tones Correct for low-verbal participants by treatment group (MMLI and SRT) and probe assessment in (a) Trained Items, and (b) Untrained Items.

	Group	Pretest		Midtest		Immediate Posttest		Delayed Posttest	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
(a)	MMLI	22.43	9.83	30.32	7.44	45.01	5.47	42.02	5.67
Trained	SRT	21.57	11.54	26.37	10.55	32.85	11.70	33.03	13.42
(b)	MMLI	22.53	15.06	26.03	14.08	34.78	14.95	34.03	14.90
Untrained	SRT	22.42	14.84	22.19	17.45	27.27	15.27	25.57	15.14

Table 7

Means (and standard deviations) of % Morphemes Correct for low-verbal participants by treatment group (MMLI and SRT) and probe assessment in (a) Trained Items, and (b) Untrained Items.

	Group	Pretest		Midtest		Immediate Posttest		Delayed Posttest	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
(a)	MMLI	28.71	14.03	36.17	8.22	50.87	8.00	47.72	9.18
Trained	SRT	29.05	12.64	35.83	13.20	45.32	13.30	44.35	14.82
(b)	MMLI	24.96	17.97	28.62	13.82	37.62	14.10	34.41	14.73
Untrained	SRT	27.02	15.24	26.04	19.45	34.48	17.95	29.07	17.33

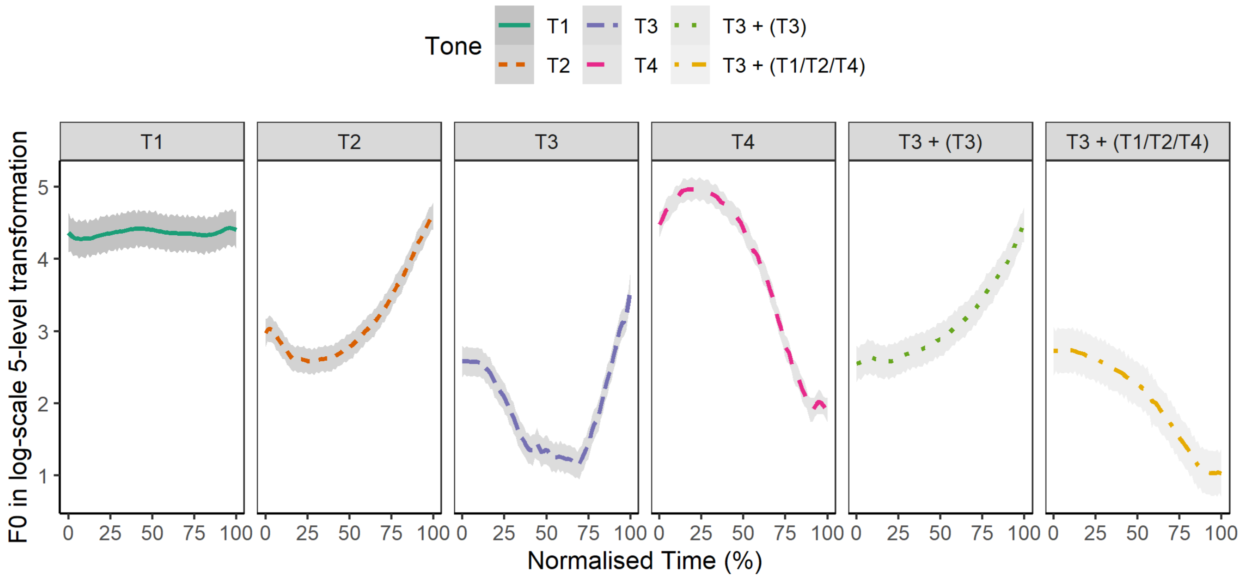

Figure 1. Pitch contours of four Mandarin citation tones (T1 [55], T2 [35], T3 [214], and T4 [51]) as well as two allophonic variants of T3 due to tone sandhi (full sandhi: *T3 [35] occurring before another T3; half-T3: *T3 [21] when before T1/T2/T4). The five-level digits in square brackets are used to transcribe tones in Chao's tone letters (Chao, 1930). Grey shades indicate standard error.

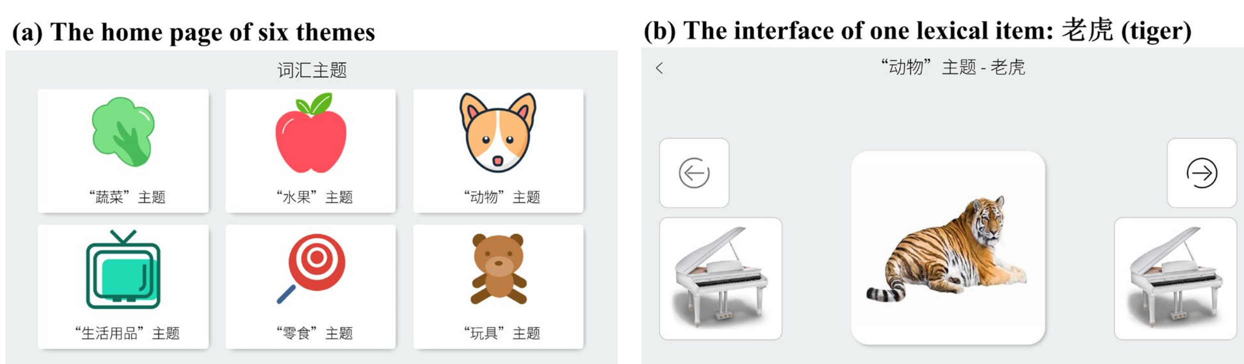

Figure 2. The user interface of the app: (a) the home page of six themes, (b) the interface of one lexical item.

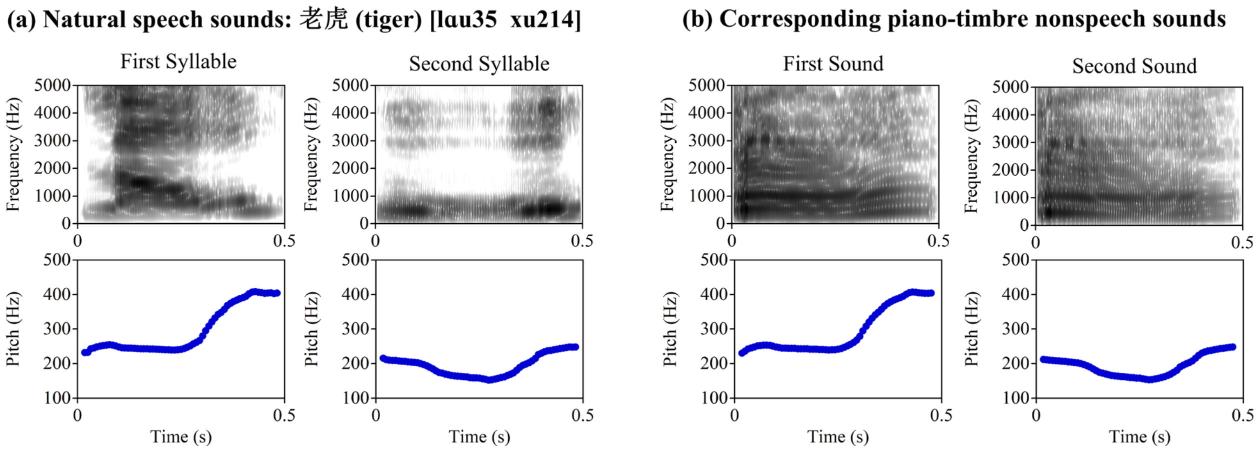

Figure 3. The spectrograms (the upper row) and pitch contours (the bottom row) of the lexical item "老虎" (tiger) [lɑu35 xu214] in (a) natural speech sounds, and (b) piano-timbre nonspeech sounds. The two types of sounds share exactly the same pitch contours with blue curves.

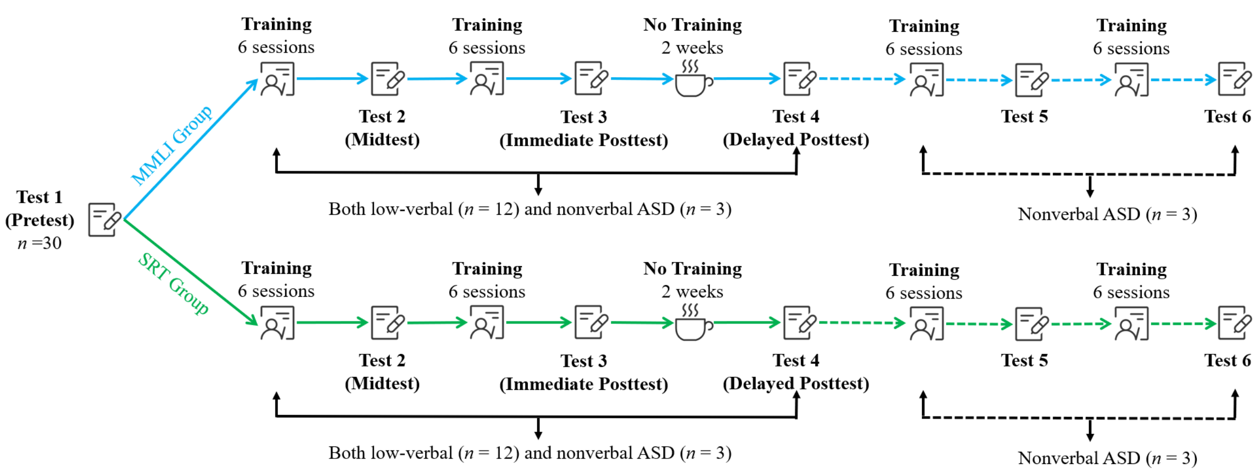

Figure 4. The probe assessments and training procedure for two treatment groups (MMLI and SRT). Tests 1–6 represent the Pretest, Midtest, Immediate Posttest, and Delayed Posttest, second-round Midtest, second-round Immediate Posttest, respectively.

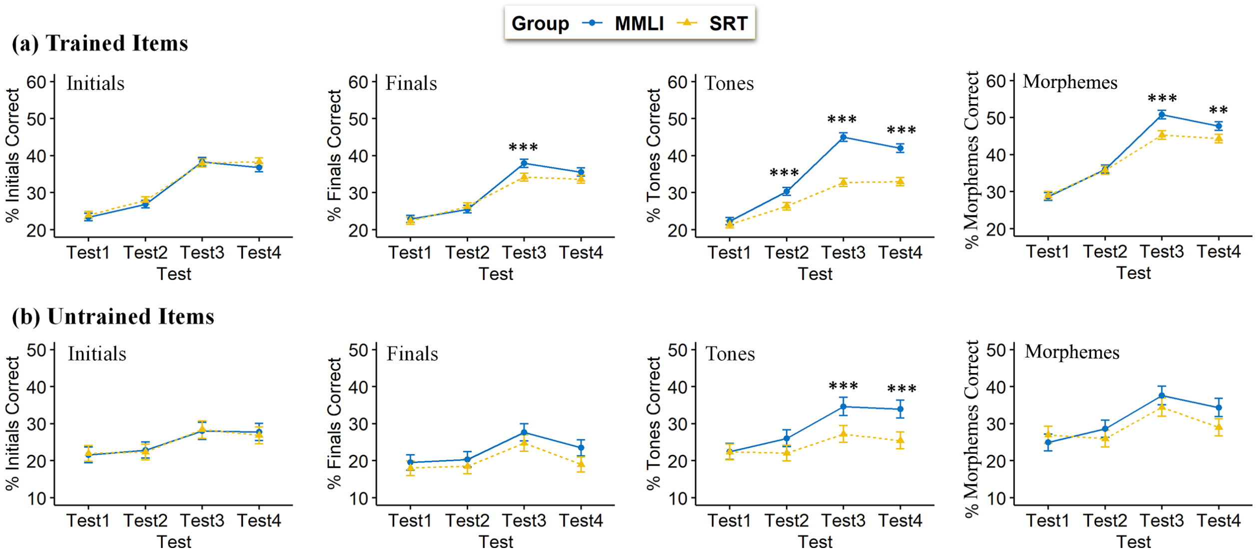

Figure 5. The production accuracy of initials, finals, tones, and morphemes for low-verbal participants by treatment group (MMLI and SRT) and probe assessment in the (a) Trained Items, and (b) Untrained Items. Tests 1–4 represent the Pretest, Midtest, Immediate Posttest, and Delayed Posttest respectively. ****p* < .001; ***p* < .01 after Tukey adjustment for the comparison of MMLI vs. SRT. Error bars: +/- 1 Confidence Interval.