Sensitivity to Vowel, Consonant and Tone Variation in Early Childhood

Thilanga D. Wewalaarachchi and Leher Singh

National University of Singapore

thilanga.d.w@u.nus.edu, psyls@nus.edu.sg

Abstract

Children have to possess robust phonological representations in order to become proficient language users. Mandarin Chinese learners' spoken word recognition abilities were investigated via an eye-tracking paradigm. Kindergarteners (5-6 years) were presented with correct pronunciations and mispronunciations involving a vowel, consonant or tone substitution. A robust ability to process correct pronunciations relative to mispronunciations was found. However, sensitivities to vowel, consonant and tone mispronunciations were not comparable. Results point to a divergence in the phonological sensitivities of segments and tones. Findings are discussed in terms of the properties and function of these sources of variation in tone languages.

Index Terms: Language Acquisition, Phonological Development, Lexical Tone

1. Introduction

A significant challenge in early language learning is the development of language-specific phonological sensitivities. Children have to learn to reject incorrect pronunciations of words and only accept pronunciations that fall within the prescribed boundaries of each phonetic category represented in the word. To date, most of the prior work pertaining to childrens' phonological development has been with intonation language learners such as learners of English [e.g. 1, 2]. Intonation languages employ two sources of phonemic variation in distinguishing lexical meaning: vowels and consonants. However, most children learn languages that employ an additional source of phonemic variation: lexical tone [3]. A disproportionate focus in research is problematic as theories pertaining to the development of phonological sensitivities are then limited to findings obtained from studying language typologies that form a statistical minority. The present study aims to compare vowel, consonant and lexical tone sensitivity in a crucial area of language development: spoken word recognition.

One way to investigate phonological development is to observe children's visual responses to correct pronunciations and incorrect pronunciations of familiar words (mispronunciation paradigm). In this way, the cost of word recognition that is associated with the substitution of each source of phonemic variation can be directly compared. It is not yet clear whether vowels, tones and consonants are equally well specified in the early childhood lexicon, which serves as the purpose of this study. In the recent years, work with adults have highlighted that the relative constraints that vowels, consonants and lexical tone place on word recognition depend on the type of language in question [4, 5]. In some intonation languages, it has been suggested that consonants are more important than vowels at the level of the lexicon [5]. However, a bias to prioritize consonant information is not language universal. This is not the case in tone languages, where vowel information constrains word recognition more than consonants [4]. Likewise, patterns of phonological development for children learning intonation versus tone languages are likely to be unique.

A possible reason that vowels, consonants and tones exert different effects on word recognition is that they are compositionally distinct [6]. Lexical tones differ from vowels and consonants in that tone results in syllable-level or suprasegmental changes, while vowels and consonants result in segmental changes [6]. Other types of suprasegmental cues in language include lexical stress and intonation. In addition, lexical tone is defined primarily by fundamental frequency or pitch, while vowels and consonants can be defined in terms of formant properties [6]. Vowel features are determined by tongue position and lip rounding while consonant features are characterized by phonation, manner and place of articulation. Although the acoustics of segmental cues and tonal cues differ, recent evidence suggests that tonal information is not processed at a protracted rate relative to segmental information [7].

Prior research demonstrates that toddlers (2 year-olds) are initially equally sensitive to mispronunciations involving vowels, consonants and tones [8]. However, at a later stage, younger pre-schoolers (2.5-3 year-olds) are more sensitive to tone mispronunciations and less so to mispronunciations involving vowels and consonants [9]. Interestingly, older pre-schoolers (4-5 year-olds) are more sensitive to mispronunciations involving vowels and consonants and less so to tone mispronunciations [9]. However, given that Mandarin learning pre-schoolers only begin to reconcile intonation functions of tone (question and statement distinction) between the ages of 4 and 5 [10], it is possible that mispronunciation effects for tone temporarily weaken because older pre-schoolers are in the midst of negotiating the complex differentiation of tonal and intonational functions during this stage of development. The preschool years are a period of aggressive language development [1]. It is possible that after a temporary period of attenuation to tone due to the functional differentiation of pitch during the preschool years, tone sensitivity may strengthen in school-aged children. The goal of the present study is to build upon prior research [9] to determine whether the previously observed attenuation in tone sensitivity extends through the next year. The present study investigates the relative impact of vowel, consonant and tone identity on lexical disambiguation in Mandarin Chinese learning kindergarteners.

2. Method

2.1. Participants

Eighteen native learners of Mandarin Chinese were sampled for this study (nine boys). Participants were between the ages of five and six (*M*= 70.94months, *SD*= 3.26months), and had no known disabilities or developmental delays.

2.2. Stimuli

Eighteen concrete, imageable words that were judged to be familiar to children aged 5 to 6 were chosen as test stimuli. A post-experimental vocabulary test was conducted to ensure that these early-acquired words were indeed familiar to individual participants. During this receptive vocabulary test, test stimuli were paired, and participants were asked "哪一个是 [target]?" (English translation: which is the [target]?). Children performed at 100% accuracy.

Test stimuli belonged to four trial types: correct pronunciations, vowel, consonant, and tone mispronunciations (please see Table 1 for sample stimuli). The higher number of correct pronunciation trials relative to the number of individual mispronunciation trials served to sustain participants' attention to the experimental task [9]. All mispronunciations resulted in non-words.

Trial Type		Test Stimuli (Translation)
Practice		Flower Tree Door
Correct Pronunciation		Pig Shoe Egg Ruler Deer T-shirt Watch Fork Dress
Tone Mispronunciation	2 to 1 4 to 2 4 to 1	Cow Rice Noodles
Vowel Mispronunciation	Backness Height Roundedness	Paper Trousers Chicken
Consonant Mispronunciation	Place Manner Aspiration	Pen Car Ball

Table 1. *Sample Stimuli List*

During the experiment, test stimuli were presented in sentence final position, with the carrier phrase "你看，那是 [target]" (English translation: Look, that is a [target]). A female adult native Mandarin Chinese speaker recorded all auditory stimuli in a child-directed manner.

Visual stimuli comprised of photographed target objects (test stimuli) and distractor objects. Novel distractors were chosen to serve as a potential alternative referent in mispronunciation trials [9]. Positions of familiar target and novel distractor were randomized and counterbalanced.

2.3. Procedure

A preferential looking paradigm was employed to investigate spoken word recognition. All participants were tested in a quiet room with a caregiver present. A 17-inch Macintosh computer was placed at participants' eye-level at a distance of 30cm. Auditory stimuli were played via external speakers at a conversational level of 70dB. The experiment comprised of 3 practice trials and 18 test trials. The purpose of these practice trials was to initiate children to the paradigm and so data from the first three trials were not analyzed.

Each trial comprised of two phases of equal duration. During the pre-naming phase, participants heard the neutral directive "Look, that is a". During the post-naming phase, participants heard the test stimulus, which was synchronized to begin at the 2500ms mark. The static visual display (familiar target and novel distractor) remained on screen for the entire duration of the trial (please see Figure 1 for sample trial structure).

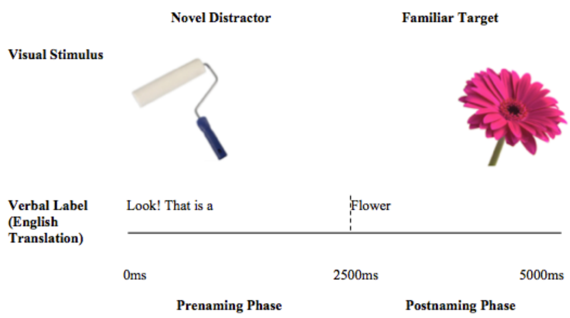

Figure 1. *A Schematic Diagram of Trial Structure*

3. Results

Participants' eye movements were coded offline, frame-by-frame at a rate of 30 frames per second. For each frame, the coder determined if the participant was looking left, right or elsewhere. As per convention, the time window used for statistical analysis was from 200ms to 1200ms after the onset of the test stimulus [2]. Figure 2 depicts the time course of word recognition for this duration.

Analyses comprise of investigations pertaining to the time course of spoken word recognition. Unlike analyses of naming effects (e.g. [9]), time course analyses provide insight on the temporal dynamics of lexical disambiguation, reflecting efficiency of online processing that occurs in real-time as the sentence unfolds. Time course analyses have the potential to reveal a fine-grained view into the constraints on spoken word recognition [11]. This is particularly relevant in the present study to guide conclusions about how tonal information is processed relative to segmental information (vowels and consonants).

To allow for time required to program an eye-movement as a result of hearing the auditory label, a 200ms time allowance was afforded [2]. Should the auditory label correspond to a correct pronunciation, a sustained fixation to the target object is expected. On the other hand, should the auditory label correspond to a mispronunciation, a decrease in fixation to the target object and a sustained fixation to the distractor object is

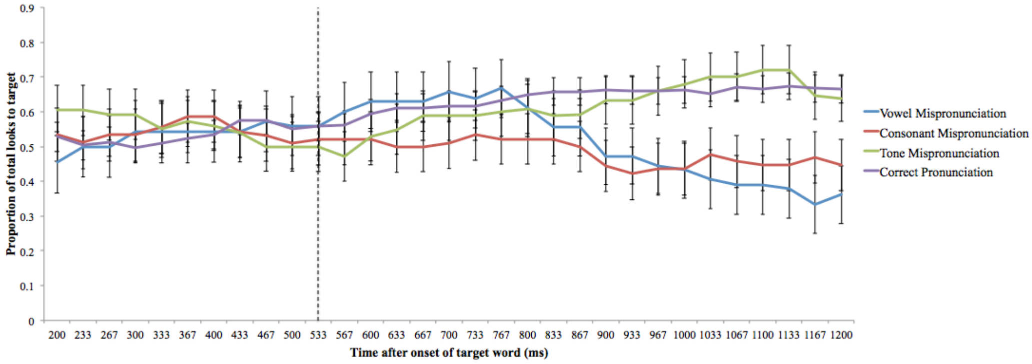

Figure 2. *Proportion of total looks to target during post-naming (Error bars reflect SEM). Dashed line indicates average offset of test stimuli.*

expected. For mispronunciation trials, higher rates of 'false-alarm' indicated by greater fixation to the target object are indicative of less efficient processing. To allow for analysis of fixation patterns by epoch, gaze data was bundled into five 200ms time intervals e.g., 200-400ms, 400-600ms … 1000-1200ms [2].

A repeated-measures ANOVA with trial type as independent factor, time interval as between subject factors, and PTL as dependent measure was conducted. Trial type comprised of four levels: correct pronunciation, vowel, consonant, and tone mispronunciation. There was a main effect of trial type, $F(1,25) = 133.47$, $p < .0001$ (partial $eta^2 = .84$), revealing that time course of word recognition statistically differed depending on trial type. In addition, there was a main effect of time interval, $F(4,25) = 12.13$, $p < .0001$ (partial $eta^2 = .1$), revealing that time course of word recognition statistically differed depending on time interval. Lastly, there was a significant two-way interaction between trial type and time interval, $F(4,25) = 60.61$, $p < .0001$ (partial $eta^2 = .91$), revealing that the differential time course patterns observed for each trial type varied systematically depending on time interval.

Post-hoc comparisons were computed to statistically examine these effects. For each time interval, post hoc comparisons were computed between trial types. Pairwise comparisons (with Bonferroni corrections) revealed greater fixations to target for correct pronunciations relative to tone mispronunciations for zero continuous time intervals. In contrast, pairwise comparisons (with Bonferroni corrections) revealed greater fixations to target for correct pronunciations relative to vowel mispronunciations for four continuous time intervals (400-600ms, 600-800ms, 800-1000ms and 1000-1200ms). Likewise, pairwise comparisons (with Bonferroni corrections) revealed greater fixations to target for correct pronunciations relative to consonant mispronunciations for four continuous time intervals (400-600ms, 600-800ms, 800-1000ms and 1000-1200ms). Contrasting both types of segmental mispronunciations, pairwise comparisons (with Bonferroni corrections) revealed neither vowel nor consonant mispronunciation had greater fixation to target for any continuous time intervals.

In summary, time course analyses revealed that sources of phonemic variation exert differential constrains on spoken word recognition. Tone information was found to affect word recognition to a lesser degree as compared to segmental information. In addition, vowels and consonants affected word recognition to the same degree.

4. Discussion

In the present study, kindergarteners who were native speakers of Mandarin Chinese were tested on their sensitivity to vowel, consonant and tone substitutions within a spoken word recognition task. Participants demonstrated a robust ability to process correct pronunciations relative to mispronunciations. When processing correct pronunciations, participants showed steady and sustained fixations toward the target object. Fewer fixations toward the target object were observed when processing vowel, consonant and tone mispronunciations. However, sensitivities to vowel, consonant and tone mispronunciations were not comparable. There was a striking divergence in the time course of spoken word recognition when processing of segmental information versus tone information. Sensitivities to vowels and consonants were comparable: vowels did not constrain word recognition more than consonants. However, sensitivities to tones were weaker than sensitivities to segments.

The dissociation observed between segmental information and tonal information complements prior work with children [9] and adult Mandarin Chinese speakers [12]. One postulate that has been put forth to explain this finding is the information load hypothesis [4, 12]. In essence, the higher the likelihood that a linguistic signal appears in speech the lower the information value it confers [13]. This probability is related to inventory size. In Mandarin Chinese, the size of the tone inventory is very small relative to the size of the segment inventories [12]. Therefore, it is possible that tones due to its low information value are assigned lower priority in conditions of limited resources. A thorough statistical account contrasting various tone languages that have a different ratio of tones to segments awaits further investigation.

It should be noted that work with adult tone language speakers has typically observed two distinct patterns of tone sensitivity depending on the type of task and nature of stimuli used. Under experimental conditions that allow for pre-activation of the target word: the use of pictures [7] or highly predictable sequence of words [14], tone sensitivity is comparable to segmental sensitivity. On the other hand, under experimental conditions that are more ambiguous and/or challenging such as lexical reconstruction [4] and speeded classification [12], tone sensitivity is weaker relative to segmental sensitivity. It could be that during the period of early childhood, children are still trying to reconcile how much priority ought to be attributed to tone information.

Taken together with findings from [9], the present study suggests that weak sensitivity to tone information relative to segments is not temporary. Rather, it appears that native learners of Mandarin Chinese go through a period of prolonged weak sensitivity to tone mispronunciations through the preschool and kindergarten years. The observation of a sustained weak mispronunciation effect for a source of phonemic variation is surprising given maturation effects previously observed with intonation language learners processing segments in word recognition [1].

The present study provides further evidence to the difficulty of generalizing findings obtained from intonation language learners to all language learners and attest to the promise of investigating sensitivity to lexical tone – in conjunction with vowels and consonants – to advance a comprehensive view of the developing mental lexicon. It appears that the developmental trajectory of lexical tone acquisition is distinct to that of segmental acquisition.

5. Funding

Ministry of Education Tier 1 Academic Research Fund grant (FY2013-FRC2-009) and a start-up grant from the National University of Singapore to LS.

6. Acknowledgements

We are grateful to Felicia Woo for assistance with stimulus recording and Hwee Hwee Goh for assistance with experimental design.

7. References

[1] Havy, M., Bertoncini, J., & Nazzi, T. (2011). Word learning and phonetic processing in preschool-age children. *Journal of experimental child psychology, 108*(1), 25-43.

[2] Creel, S. C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science, 15*(5), 697-713.

[3] Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

[4] Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and speech*, *59*(1), 59-82.

[5] Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, *24*(5), 573-583.

[6] Singh, L., & Fu, C. S. (2016). A new view of language development: the acquisition of lexical tone. *Child development, 87*(3), 834-854.

[7] Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420.

[8] Wong, L. H, & Singh, L. (2015) Sensitivity to phonological detail in bilingual and monolingual infants. Poster presented at Workshop on Infant Speech Perception: WISP. Sydney, NSW

[9] Singh, L., Goh, H. H., & Wewalaarachchi, T. D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition*, 142, 1-11.

[10] Singh, L., & Chee, M. (2016). Rise and fall: Effects of tone and intonation on spoken word recognition in early childhood. *Journal of Phonetics*, *55*, 109-118.

[11] Swingley, D. (2011). The looking-while-listening procedure. *Guide to Research Methods in Child Language*.

[12] Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*(5), 689-708.

[13] Garner, W. R. (1988). The contribution of information theory to psychology. In W. Hirst (Ed.), *The making of cognitive science: Essays in honour of George A. Miller*, 19-35. Cambridge, UK: Cambridge University Press.

[14] Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes*, *22*(4), 566-594.