Lexical Representation of Consonant, Vowels and Tones in Early Childhood

Hwee Hwee Goh1 Charlene Hu1, Kheng Hui Yeo2 & Leher Singh1

1 National University of Singapore
2 A*Star

a0071548@nus.edu.sg, charlene.hu@gmail.com, icedwater@gmail.com, psyls@nus.edu.sg

Abstract
Several previous studies have investigated the extent to which segmental detail is represented in the developing lexicon. However, the majority of previous studies have focused on consonant and vowel representation, with little attention to the representation of lexicon tone in spite of its predominance in languages of the world. The current research provides a direct comparison of vowels, consonants and lexical tone representation using a mispronunciation paradigm in children at two ages (2.5 to 3.5 years and 4 to 5 years). Results point to asynchronous emergence of sensitivity to vowel, consonant and tone variation. Tone sensitivity appeared to be high in the younger cohort relative to vowels and consonants. By contrast, vowel and consonant sensitivity was strong in the older age group, with tone sensitivity appearing relatively weak. Findings point to a close coupling in vowel and consonant sensitivity in the preschool years and a dissociation with lexical tone.

Index Terms: language acquisition, phonological development

1. Introduction
To successfully understand language, a child’s mental representation of a word must be phonologically precise. Sound changes that can drive meaning changes (i.e. phonemic shifts) must be detected and differentiated from sound changes that do not alter meaning (i.e. non-phonemic variation). Vocabulary development is largely dependent on a learner’s accuracy in linking sound and meaning.

In previous studies, sound-meaning links in language comprehension have been studied using a variety of laboratory methods. A widespread methodology is the mispronunciation paradigm. In this paradigm, children are exposed to two pictures, the name for one picture being early-acquired and likely known to the child (e.g. apple) and the name for the other being late-acquired and likely unfamiliar to the child (e.g. a manual typewriter). In some trials, the familiar object is correctly pronounced (e.g. “apple”), while in others, it is mispronounced (e.g. “opple”). Fixation to the target upon hearing the correct form or a mispronounced variant is interpreted as an indicator of the extent to which lexically representations are phonetically well-specified in the developing lexicon. Mispronunciation studies have revealed much about the developmental trajectory of vowel and consonant sensitivity, showing that vowels and consonants are initially represented with varying degrees of levels of relative strength in the early lexicon depending on the methodological paradigm in use [1] [2] [3].

Lexical tone is a highly frequent segment type in human languages and indeed, most languages incorporate tone as a phonetic segment [4]. Most of the world’s language learners learn a tone language natively [5] and as such, it is a type of segment with which most children contend. However, there has been relatively little research on how tone is represented in the early lexicon and furthermore, on whether tone is encoded with relative precision as compared with consonants and vowels. The current study directly compares the effects of vowel, consonant and tone variation on language comprehension.

There are several reasons to posit that tone representation may be distinct from other types of phonemic segments such as vowel and consonants. First, lexical tone is compositionally distinct from vowels and consonants, the latter defined primarily by acoustic variation in the first and second formants whereas tone is defined principally by fundamental frequency variation. Second, theoretical descriptions of linguistic phonology ascribe to tone an independent role from vowels and consonants [6], leading to the emergence of auto-segmental phonology established primarily to account for lexical tone as a phonetic segment. Finally, the distinction between phonemic and non-phonemic detail is arguably less clear in the case of lexical tone, which capitalizes on fundamental frequency variation. Fundamental frequency variation is widely observed in tone and non-tone languages alike to communicate a wide variety of non-lexical functions, such as question/statement distinctions, grammatical packaging, emotional variation [7]. It is incumbent upon the tone language learner to appreciate the many functions of fundamental frequency variation in addition to effecting lexical changes. By contrast, vowel and consonant variation are primarily recruited to signal lexical contrast. While non-phonemic vowel and consonant variation is rampant (e.g. regional accents), for the most part, vowels and consonants are not purposefully manipulated along a continuous dimension to serve an independent communicative or pragmatic function. As such, lexical tone, consonants and vowels are functionally and structurally distinct, inviting the possibility of asynchronous acquisition of each type of phonetic segment. The current study aims to chart sensitivity to consonantal, vocalic and tonal detail in early word recognition.

2. Methodology

2.1. Participants
Forty-nine native language speakers of Mandarin Chinese participated in the current study. The participant sample comprised 24 toddlers (mean age: 36 months; range: 30 to 42 months, 12 boys) and 25 preschoolers (mean age: 50 months; range: 45 to 57 months, 13 boys). Data from 3 participants were excluded based on failure or refusal to complete the testing session. All participants were typically developing children and were performing at grade level.
2.2 Experimental design

Participants were tested using the preferential looking paradigm, widely used to obtain on-line measures of word recognition and word learning. Previous instantiations of the method to investigate phonological specificity of vowel and consonants have relied on children’s responses to mispronunciations, which serves as the paradigm for the current study [1][2][3].

The mispronunciation paradigm utilizes a split screen display with a familiar object on the left/right side and an unfamiliar object on the other side. Familiarity and side of screen were counterbalanced across trials. During each trial, children view a familiar and unfamiliar object on screen for 2500 msec. (salience phase). Following this, they hear a verbal label for the familiar object (naming phase). The proportion of time spent fixating the target during the salience phase is computed and compared with the proportion of time spent looking at the target during the naming phase. Word recognition is inferred when proportion of total looking to the target increases significantly from salience phase to the naming phase. Other possible responses include no increase in fixation to target between the salience and naming phase, suggesting that the participant is uncertain about the referent for the verbal label and fixates equally on the target and unfamiliar object (distractor). A final alternative is a significant reduction in fixation to target during the salience and naming phase, suggesting that the participant has associated the verbal label with the unfamiliar object. The trial structure is displayed in Figure 1.

![Figure 1: Schematic diagram of test trial structure](image)

Figure 1: Schematic diagram of test trial structure

The experiment comprised 21 trials all of which were produced in Mandarin Chinese. At the start of each experiment, there were three practice trials consisting of familiar objects. Following the practice trials, participants were presented with 18 test trials. Each test trial consisted of a familiar object (target) and an unfamiliar object (distractor). Each test trial also consisted of a salience phase, designed to measure the attentional draw of each stimulus followed by a naming phase. For half of the 18 test trials, the target was correctly labeled (correct pronunciation trials). For the other half of the test trials, the target was mispronounced. For 3 trials, the target was mispronounced by inserting a vowel substitution. Each vowel substitution was a single-feature substitution, with one trial involving a change in vowel backness, one trial involving a change in roundedness and one trial involving a change in height. For 3 of the test trials, the target word was mispronounced by virtue of a consonant substitution (1 place of articulation change, 1 manner of articulation change and 1 aspiration change). For the last three test trials, mispronunciations were induced via a tone change (Tone 1 -> Tone 2; Tone 2 -> Tone 4; Tone 1 -> Tone 4). Tone 3 was omitted from the design because it is subject to tone sandhi and may be more complex for young children to process than Tones 1, 2, and 4.

Proportion of fixation to the target was computed for each phase by coding participants’ eye movements frame-by-frame (every 33 msecs) for the duration of the test session. Test trials for which participants did not fixate the target and distractor during the pre-naming phase were excluded. In addition, test trials where attention to the screen was .2 or below were excluded. Following the experiment, participants were given a vocabulary test where they were tested on their productive knowledge of each of the targets. Trials for which participants did not know the name of the target word were excluded. This resulted in 26% of trials being excluded from analysis.

3. Results

The primary dependent variable was proportion of total looking to target during the salience and the naming phase, with a significant increase between the two phases indicating that participants had mapped the verbal label on to the target.

A 4 x 2 x 2 analysis of variance was performed with age (younger/older), pronunciation type (vowel mispronunciation, consonant mispronunciation, tone mispronunciation, and correct pronunciation) and phase (salience and naming phases) as factors. Results revealed a marginally significant three-way interaction of factors, F(3, 141) = 2.61, p=.05. Simple comparisons were then conducted to examine naming effects for each pronunciation type within older and younger children. Within the younger cohort, there was a significant increase in proportion of fixation to target during the naming phase relative to the salience phase when targets were correctly pronounced (t(23) = -2.97, p=.01) but not when targets were mispronounced due to a vowel change (t(23) = -1.07, p=.30), or due to a consonant change (t(23) = -1.07, p=.30) or due to a tone change (t(23) = 1.35, p=.19). Within the older cohort, there was a significant increase in proportion of fixation to target during the naming phase relative to the salience phase when targets were correctly pronounced (t(24) = -5.16, p=0.0) but there was a significant decrease in fixation to target when targets were mispronounced due to a consonant change (t(24) = 2.22, p = .04), but no significant difference when targets were mispronounced due to a vowel change (t(24) = 1.20, p = .24), nor when due to a tone change (t(24) = -.084, p = .93).

Previous studies have compared the difference in proportion of total looking to the target during salience and naming phases for different trial types to investigate the strength of the mispronunciation effect (i.e. the degree of sensitivity to the mispronunciation) for different types of mispronunciations (see Figure 2). To this end, naming effects were computed by subtracting proportion total looking to the target during the naming phase from those during the salience phase. Traditionally, higher values indicate an insensitivity to mispronunciations and lower values indicate a high sensitivity to mispronunciations [1][2][3]. To investigate the relative...
strength of mispronunciation effects for different pronunciation types (vowel mispronunciation, consonant mispronunciation, correct pronunciation) across age, a 4 x 2 (pronunciation type x age) repeated-measures ANOVA was computed with naming effects as the dependent variable. Results demonstrated a main effect of pronunciation type (F(3,141) = 3.61, p=.02), a marginal interaction of pronunciation type and age (F(3,141) = 2.60, p =.06) and there was no main effect of age (F(1, 47) = 1.44, p = .24). Individual comparisons were conducted within age group to compare the strength of vowel, consonant and tone mispronunciation effects. Mispronunciation effects for vowels, consonants and tones were compared to naming effects for correct pronunciations. Within the younger cohort, vowel mispronunciation effects did not differ from naming effects for correct pronunciation trials, t(23) = -2.00, p =.06). However, for tone mispronunciation effects, effects were significantly stronger than for correct pronunciations, t(24) = -2.88, p =.008. Within the older cohort, vowel mispronunciations effects were significantly stronger than for the correct pronunciations (t(24) = -3.148, p = .004); similarly, consonant mispronunciation effects were stronger than the correct pronunciation trials (t(24) = -4.542. However, the naming effects for the tone mispronunciation trials were only marginally stronger than that of the correct trials (t(24) = -2.00, p = .06).

![Figure 2: Naming Effects for Different Pronunciation Types](image)

### 4. Discussion

The objective of the current research was to investigate the phonological specificity with which tones, vowels and consonants are represented in the developing lexicon. Using a mispronunciation paradigm, recognition of familiar words was investigated when labels were correctly and mis-pronounced. Results demonstrate that at both age groups (older and younger), participants recognized targets when they were correctly pronounced and correctly reject labels that are mispronounced via tone, vowel or consonant substitutions. However, mispronunciation effects were not equal in size across consonants, vowels and tones nor were they equivalent across age groups. In younger children (2.5 to 3.5 years of age), consonant and vowel mispronunciations were associated with relatively weak effects and tone mispronunciations were associated with relatively strong effects. In older children (4 to 5 years of age), mispronunciation effects caused by consonants and vowels were relatively strong – although they remained comparable in magnitude – and those associated with tone changes were relatively weak.

### 5. Conclusions

Previous research has focused almost exclusively on the extent to which vowels and consonants are represented in the mental lexica of young children. By contrast, lexical tone – a highly frequent type of phonetic segment – has received little attention. The current study aims to directly compare the effects of vowel, consonant and tone mispronunciations to determine the extent to which each of these sources of phonemic variation are represented in children’s lexica.

Previous research investigating phonological specificity in vowels and consonants has revealed evidence for equivalent sensitivity to vowels and consonants in early learners using the mispronunciation paradigm employed herein [1][2]. While vowels and consonants serve similar functions in human languages, they are structurally quite distinct. Vowels are typically longer than consonants and in most languages, vowels form a more limited inventory than consonants. Vowels are not contrasted by voicing, like consonants. Vowels are steady-state constituents of the speech signal whereas consonants are acoustically transient by comparison. Furthermore, one might argue for greater similarity between vowels and tones in that vowels and tones are convolved in the speech signal. Both tones and vowels are extended in duration relative to consonants and are defined by steady-state acoustic energy. Finally, there is an association between tone and vowel length suggesting they are coupled in natural discourse [8]. The current study points to a dissociation between vowels/consonants and lexical tone at two points in development and to a close coupling of vowel and consonants in early lexical representations. Furthermore, vowel and consonant mispronunciation effects appear weak relative to tone early in development, a pattern which appears to reverse at older age groups. Findings point to an asynchronous course of sensitivity to vowels, consonants and lexical tone in child language acquisition.

### 6. References