Research Article

Rise and fall: Effects of tone and intonation on spoken word recognition in early childhood

Leher Singh *, Melissa Chee

Department of Psychology, National University of Singapore, Singapore

ARTICLE INFO

ABSTRACT

A crucial component of word learning is the ability to recognize words in spite of the varying forms they assume. This may be particularly challenging in tone languages as learners have to develop tone representations in the face of intonational variation in order to accurately recognize words. The effects of intonational variation on word recognition of tone-marked words in Mandarin Chinese were investigated in toddlers and preschoolers using a cross-sectional design. Participants were presented with known words where intonation (question/statement) and tone (rising/falling) were independently manipulated. Results demonstrated that word recognition in toddlers was heavily influenced by changes in the pitch contour of a tone due to intonational variation. In contrast, preschool were able to recognize tone-marked words regardless of simultaneous intonational variation, demonstrating a comparatively robust representation of lexical tone. Results chart an evolution in integrating pitch cues to tone and intonation over the first few years of life.

1. Introduction

The mastery of a human language is a feat often underestimated by the ease and automaticity with which early language acquisition seems to unfold. Children typically amass substantial vocabularies at an early age (Fenson et al., 1994) continuously adding new words to their vocabulary as they advance towards native language proficiency. The accretion of word knowledge necessarily entails harboring correct expectations of how words are defined in the learner's native language. This process is complicated by the fact that the speech signal is replete with variability along a multitude of dimensions, with each dimension varying in its linguistic and communicative significance in a context-dependent manner (Perkell & Klatt, 1986). For example, there are several sources of variation that do not alter the lexical identity of a word, but that influence the semantic interpretation of speech. Such sources of variation are often classified under paralinguistic or intonational sources of variation. These cues include affective prosody (Banse & Scherer, 1996), focus and phrase-level prominence (Fernald & Mazzie, 1991) and communicative intent, as exemplified by the question/statement distinction (Van Heuven & Haan, 2002).

Many forms of paralinguistic variation are driven by pitch movements (Ladd, 2008). However, a widely spoken class of languages – tone languages – also exploit pitch movements to draw lexical distinctions.[1] As a result, there is a relatively broad swath of functions ascribed to pitch movements in tone languages. This leaves it to the tone language learner to attend to pitch movements with due consideration to their specific function. There therefore exists a natural tension between generalizability and specificity of early lexical representations: memory for pitch movements has to be sufficiently precise such that learners distinguish words based on lexical contrast (i.e. attend to lexical tone changes); on the other hand, words also have to be flexibly defined with regards to pitch movements to enable successful word recognition in the face of intonational variation. In this way, tone languages serve as a potentially informative means to investigate the extent to which language learners negotiate the varying functions served by vocal pitch.

---

* Correspondence to: Department of Psychology, AS 4, 03-40, 9 Arts Link, Singapore 117570, Singapore. Tel.: +65 6516 7750; fax: +65 6773 1843.
   E-mail address: psyls@nus.edu.sg (L. Singh).
   [1] There appear to be some differences in the ways in which pitch is exploited to draw lexical versus paralinguistic distinctions. Lexical pitch (or tone) typically spans a much smaller fundamental frequency range than intonational variation (Xu, 1999) and elapse over a shorter time-course as compared with intonation (Lee, Tao, & Bond, 2008). Additionally, there are phonological rules that apply to the acoustic realization of tones that do not impact systematically upon intonation, such as *sandhi* (Chao, 1968). However, in the presence of conflict, corpus analyses have revealed that speakers of tone languages reduce the potential for ambiguity by maintaining tone constancy at the cost of intonational constancy when the two sources conflict (Karlsson, House, & Svantesson, 2012).

Languages that ascribe multi-tiered (i.e. lexical, intonational) functions to pitch movements are not linguistic obscurities: the majority of the world's language learners are tone language learners (Fromkin, 1978) and the majority of the world's languages are tone languages (Yip, 2002), yet the process by which learners differentiate the functions of pitch in tone languages remains unclear. This is perhaps due to the fact that the vast majority of psycholinguistic research in language acquisition is focused on non-tone (or intonation) languages, drawing predominantly from Romance and Germanic languages. As tone languages enter the fray of language acquisition research, it is becoming increasingly evident that sensitivity to tone may follow a different time course to sensitivity to consonants and vowels (Singh, Goh, & Wewalaarachchi, 2015; Yeung, Chen, & Werker, 2013). Moreover, the question of when tone language learners successfully negotiate tone-intonation relations (i.e. being able to control for the effects of intonation on the realization of tones) remains undetermined and represents the focus of the current study.

Variation in tone-intonation relations presents a potential challenge to language learners in that particular combinations of cues to tone and intonation can alter the precision with which lexical tones are specified. Based on whether pitch movements associated with lexical identity co-vary or conflict with those associated with intonational shifts, word recognition can be impacted. For example, Mandarin Chinese has 4 lexical tones (Tone 1: High; Tone 2: Rising; Tone 3: dipping; Tone 4: falling). In addition to tones, languages also use pitch to convey intonational contrasts (Bolinger, 1989; Gussenhoven, 2002) including Mandarin Chinese (Ho, 1977; Lee, 2005; Shih, 1988; Yuan, 2004). Furthermore, unlike in English, questions in Mandarin Chinese are not subject to syntactic movement (e.g. subject–auxiliary verb inversion), placing a greater reliance on pitch cues to the question/statement distinction in such cases. Canonical question forms in Mandarin Chinese are signaled by a global rise in pitch at the level of the sentence in addition to an expanded terminal rise on the final word of the syllable (Ho, 1977; Yuan, 2004, 2006; Zeng, Martin, & Boulakia, 2004). These pitch transformations can be applied to a declarative sentence structure (e.g. "This is a chair?").[2] In Mandarin Chinese, statement forms can therefore be syntactically and lexically identical to question forms, distinguished by a falling intonation contour on the final syllable of the word (Ho, 1977; Zeng et al., 2004). In such cases, common acoustic drivers of pitch variation in tone and intonation present a potential challenge to learners.[3] This raises the question of how native learners of languages such as Mandarin Chinese recover word meanings when cues to tone and intonation diverge.

Previous studies have investigated adults' abilities to reconcile tone-intonation ambiguity in native tone language speakers. At the outset, it should be noted that intonational variation does not dramatically change tone contours nor does it lead to the perception of a different tone (Ho, 1977; Shen, 1989). However, intonation cues can converge with or diverge from tone cues, altering the physical manifestation of a tone. In a study conducted by Yuan (2004) designed to investigate the consequence of such convergences and divergences between tone and intonation, Mandarin-speaking adults were asked to judge an intonational contour as a question or a statement. The tone of the final syllable for each question or statement was either a rising (Tone 2) or falling (Tone 4) tone. Native listeners encountered particular difficulty identifying question forms if the utterance ended with a word spoken in Tone 2 (rising intonation). Yuan accounts for these results by suggesting that tone identification may take priority over intonation identification, such that the perception of a rising tone may reduce the likelihood that the same stimulus is also perceived as belonging to a rising intonational category (i.e. a question). It should be noted that while this study did not focus on tone identification, but rather on intonational identification, Yuan's results demonstrate that judgments of intonation are compromised by tone identity under particular conditions.

While there have been few investigations for children's abilities to integrate tone and intonation in word recognition, there have been several studies to investigate children's abilities to perceive lexical tone contrasts. In the aggregate, these studies point to a strong and early sensitivity to lexical tone in infants raised in a tone language environment. For example, infants exposed to a native tone language demonstrate language-specific tone categories as early as four months of age as revealed by auditory discrimination of lexical tones (Yeung et al., 2013). In contrast, infants exposed to non-tone languages such as English show a decline in sensitivity to tone contrasts around 9 months of age (Mattock & Burnham, 2006; Yeung et al., 2013). In addition to studies on phoneme discrimination, studies on spoken word recognition have revealed that bilingual infants learning one tone language and one non-tone language differentiate words in Mandarin according to lexical tone but group words in English that are intonationally distinct by 11 months (Singh & Foong, 2012). Later at 18 months, when forming novel word-object mappings in the laboratory, tone-learning infants are able to integrate tone into novel words when establishing new word-meaning correspondences (Singh, Tam, Chan, & Golinkoff, 2014). Even later at 3 years, when listening to familiar words, toddlers appear highly sensitive to variation in lexical tone in comparison to vowel and consonant variation (Singh et al., 2015). In the aggregate, a series of studies therefore demonstrate an early and potent sensitivity to lexical tone in phoneme discrimination, word recognition and novel word learning over the first two to three years of life.

Investigations of tone production mirror these conclusions. Several studies have recorded children's productions of native tones and assessed the accuracy with which contrastive tones are produced by children. In a seminal study, Li and Thompson (1977) investigated tone production in infants over an 8-month period. They reported considerable maturation of tone productions prior to 3 years of age, concluding that tones were largely mastered prior to consonants and vowels at this stage. In a larger scale study, Hua and Dodd (2000) examined productive speech errors in consonant, vowel and tone productions, arriving at a similar conclusion that tones were less vulnerable to error than vowels and consonants. While Li and Thompson and Hua and Dodd's studies were conducted in Mandarin, consistent evidence for early acquisition of tones in production has been reported in Cantonese (see So & Dodd, 1995; To, Cheung, & McLeod, 2013). However, although tone categories appear to be acquired early in production, detailed

---

[2] It should be noted that questions can assume a declarative form in English too, although this structure is non-canonical.

[3] Even though questions in Mandarin can be optionally marked with the sentence-final particle "吗" ("ma"), Mandarin speakers can use unmarked questions in natural discourse (Zeng et al., 2004), although these forms are more strongly marked by intonation than particle questions. Unmarked questions and statements in Mandarin maintain the same word order and differ only by intonation (a construction that is also permissible in English).