中国语言学会语音学分会 主办
PAC Phonetic Association of China

第2辑

Chinese Journal of Phonetics

中国语音学报

主编 鲍怀翘

商务印书馆
创于1897 The Commercial Press

中国语言学会语音学分会　主办

中国语音学报

第2辑

主　编：鲍怀翘

副主编：孔江平

李爱军

陶建华

2009年·北京

图书在版编目（CIP）数据

中国语音学报. 第2辑/中国语言学会语音学分会主办，
鲍怀翘主编. 北京：商务印书馆，2009
ISBN 978-7-100-06718-8

I. 中… II. 中… III. 汉语-语音学-文集 IV. H11-53

中国版本图书馆CIP数据核字(2009)第124385号

ZHŌNGGUÓ YǓYĪN XUÉBÀO
中国语音学报
第2辑
中国语言学会语音学分会　主办
鲍怀翘　主编

商务印书馆出版
（北京王府井大街36号　邮政编码 100710）
商务印书馆发行
北京瑞古冠中印刷厂印刷
ISBN 978-7-100-06718-8

2009年12月第1版　　开本 787×1092　1/16
2009年12月北京第1次印刷　　印张 11¾
定价：24.00元

《中国语音学报》编辑委员会

目　录

CONTEXTUAL VARIABILITY AND INFANTS' PERCEPTION OF TONAL CATEGORIES

SHI Rushen

Abstract: This paper reviews some of the major findings in infants' perception of phonetic categories, including consonants, vowels and lexical tones. We discuss empirical evidence suggesting a development from language general to language specific perception of phonetic categories. We focus on the problem of variability, especially the type of variability resulting from phonetic contexts. We then present results demonstrating that the ability to filter out contextual variability is a gradual learning process, unlike certain other perceptual abilities that are available at the initial state. The findings of this study will be discussed in the context of the mechanisms of early language development.

Keywords: lexical tones, phonetic categories, contextual variability, infant speech perception

1. BACKGROUND

Research in the past four decades has revealed remarkable speech perception abilities in preverbal infants. At birth infants' perceptual sensitivity appears to be language general. During the second half of the first year of life infants gradually focus on the sound structure of the ambient language. This was manifested in experiments in which infants shortly after birth showed acute perception of both native and non-native phonetic contrasts but gradually declined in their discrimination of non-native contrasts. In the first demonstration of this development (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984) 6- and 8-month-old English- and Hindi-learning infants as well as Hindi-speaking adults discriminated the Hindi dental and retroflex stops, but by 10 to 12 months of age, only Hindi infants continued to discriminate the sounds whereas English-learning infants were no longer able to discriminate this non-native contrast. The same pattern of development was shown for vowel perception. The decline in discriminating non-native vowel contrasts begins to emerge around six months of age (Polka & Werker, 1994), earlier than in the case of consonants. In addition, Kuhl et al (Kuhl, 1991; Kuhl, et al., 1992) showed that by six months of age infants' vowel perception reflects the internal structure of their native language vowel categories: only native vowels exhibit a perceptual magnet effect related to their prototypical exemplars. Thus, it is not the case that infants begin phonetic category learning from scratch and gradually establish representations. But rather, phonetic category perception appears to begin with rich initial sensitivity and is gradually attuned to and shaped in accordance with the native language structure, a process of reorganization.

In a recent study on the perception of English /r/ and /l/ by English- and Japanese-learning infants, Kuhl and colleagues (Kuhl, et al., 2006) replicated findings of earlier studies, showing that Japanese infants, who discriminated this non-native contrast at a young age, declined in performance by 10-12 months of age. Furthermore, while English-learning infants discriminated this contrast at six to eight months of age, their performance at 10-12 months improved relative to that at the younger age. These studies suggest that although the perceptual sensitivity is initially rich, phonetic category development is not a simple process of attenuating non-native categories. Improved sensitivity to native categories occurs while the perception of non-native categories declines. The improvement in the perception of native language contrasts continues into later childhood. Recent work showed that native language experience facilitates the perception of the English voiced d-ð contrast in four-year-old English-learning children (Sundara, Polka & Genesee, 2006).

Besides consonants and vowels, tones constitute another type of phonetic categories. Many languages use tones to contrast lexical meaning. The primary acoustic cue for tones is the fundamental frequency. For example, Mandarin, in which the tonal bearing unit is the vowel or the syllable (depending on theories), has four lexical tones – high, low, rise and fall. High and low tones are static tones while rise and fall are dynamic tones. Thus, the meaning of /ma/ changes depending on the tones, e.g., ma-high (mother), ma-rise (hemp), ma-low (horse), ma-fall (to curse). Little research exists on preverbal infants' perception of lexical tones. Although it has been shown in music perception studies that five- to 10-month-old infants can distinguish between tunes of different pitch contours (e.g., Trehub, Bull, & Thorpe, 1984), such results may represent processes that are entirely different from those for language. Given that preverbal infants' speech perception moves from language general to language specific, not only in their discrimination of segments, but also in prosodic perception (for example, stress based prosody: Bijeljac-Babic & Nazzi, 2007; Jusczyk, Cutler, & Redanz, 1993; Skoruppa, et al., 2007), it follows that the same pattern of perceptual reorganization may occur for lexical tones.

In recent work Mattock and colleagues (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008) tested the discrimination of Thai tones by tone- versus non-tone-language learners. At four and six months of age, infants of both tone- and non-tone-language environment showed successful tonal discrimination. By nine months of age, only tone-language infants maintained the discrimination. Infants learning English and French declined in their discrimination of lexical tones by this age, i.e., becoming "tone deaf". Interestingly, Mattock and colleagues found that non-tone-language infants continued to distinguish between musical analogs of these lexical tones, consistent with evidence from musical perception studies (e.g., Trehub, Bull, & Thorpe, 1984). This suggests that the perceptual system processes lexical tones in the language mode, distinctly from other modes of auditory processing.

Studies on infants' perception of phonetic categories such as those discussed above typically used syllable-size exemplars produced in citation forms, usually one exemplar per category. The results are thus not revealing about infants' ability to categorize variable tokens of the same sound in multi-word utterances. In experiments involving citation syllables of certain continuous acoustic dimensions, infants did show categorical perception. For instance, babies can categorically discriminate stop consonant tokens that differ in voice onset time (VOT): they distinguished exemplars across a category boundary while failing to do so for tokens varying equally in VOT within the boundary (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). In experiments testing tones varying along the F_0 continuum (Harrison, 2000), Yoruba-learning six- to eight-month-old infants showed evidence of categorical perception of tonal categories, whereas English-learning infants perceived within- and cross-boundary exemplars of these tonal categories in a non-categorical fashion.

It is important to note that variability in natural speech is far more complex than the types of continua used in the above experiments. The acoustic realization of any given phonetic category is highly variable, affected by many factors including speech rate, emotion, speaker differences, phonetic contexts, etc. Perceiving variable forms of the same speech category can be challenging for infants. Limited work has been conducted to examine some of these aspects. Eimas and Miller (1980) showed that two-to-four-month-old infants are able to discriminate stop consonants even when different speech rates affecting VOT values were introduced. Infants aged six months can also ignore cross-speaker variations when perceiving vowel categories (Kuhl 1979). It is unknown whether the ability to handle these sources of variability in phonetic categorization is part of the initial perceptual biases (as in the case of speech discrimination shown in previous studies) or requires gradual learning from language experience.

Research is lacking with regards to infants' ability to handle contextual variability during phonetic categorization. Speech sounds are encoded in a parallel fashion, not strung one after another as discrete acoustic elements (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). There is a high degree of coarticulation between abstract phonetic units, leading to great amount of changes in the realization of each phonetic category and overlap between neighboring segments. Input speech to infants contains about 90% multi-word utterances (e.g., van de Weijer, 1998; Shi, Morgan, & Allopenna, 1998). The acoustic variability of phonetic categories is thus substantial. Indirect data on infants' processing of such contextual variability can be found in word segmentation studies, in which infants are first familiarized with a word target, and then tested with sentences containing the target in various phonetic contexts as opposed to other sentences not containing the target. Using this technique, Jusczyk and Aslin (1995) showed that infants recognized single-syllable targets in sentences at 7.5 months of age, but not at six months of age. Other studies (e.g., Bortfeld, et al., 2005; Hoehle & Weissenborn, 2003; Shi, Cutler, Werker, & Cruickshank, 2006; Shi & Lepage, 2008) subsequently confirmed that segmentation of monosyllabic words occurs from the second half of the first year of life. Hence, on the one hand, the remarkable discrimination and categorization of phonetic categories (involving citation tokens without contextual variation) shown in very young babies reflect certain innately available perceptual biases. On the other hand, infants' categorization of phonetic categories in the face of contextual variations, as shown in these indirect studies, appears to require a period of learning.

No study has examined infants' categorization of tonal categories in variable phonetic contexts. The experiments by Mattock and colleagues (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008) employed one token for each category, and the Harrison study (2000) used multiple tokens. The stimuli of both sets of studies were devoid of any phonetic context. The pitch realizations for tones in connected speech, however, are greatly altered by the tones of the neighboring syllables (e.g., Gandour, Potisuk, & Dechongkit, 1994; Xu, 1997). For example, when a high level tone is preceded by a low tone, the high tone is realized with a rising pitch. How does the infant learn that the underlying representation of this tone is a static high? A solution to this problem was proposed in recent studies (Gauthier, Shi, & Xu, 2007a, 2007b). In those studies unsupervised neural networks received as training input syllable-size raw F_0 curves versus velocity profiles of the F_0 curves for the four tones in Mandarin. The tones were produced by multiple speakers in sentential contexts with all tones as the neighboring contexts and with different focus conditions (production data from Xu, 1997, 1999). The networks categorized the test tokens successfully, and the performance was nearly perfect for networks trained with velocity profile. These results indicate that the naïve learning system (as in the case of infants) can learn to categorize tones by normalizing variability resulted from varying tonal contexts and speakers.

The present study aimed at investigating Mandarin-learning infants' categorization of tones in variable tonal contexts. Given the findings of previous segmentation studies, we hypothesized that categorizing tones that occur in different tonal contexts is a gradual learning process. We tested Mandarin-learning infants aged four to eight months. It was expected that tonal categorization should be observed in older infants, but not in younger ones.

2. THE EXPERIMENT

2.1 Methods

2.1.1 Participants

Participants were 18 monolingual Mandarin-learning infants, half in the younger age group (four to six months), and half in the older age group (eight to 11 months). The families were recent immigrants to Montreal from China. The only language used in these families was standard Mandarin.

2.1.2 Speech stimuli

A minimal tonal pair of the syllable "*mi*" (/mi/) was determined as the targets of the experiments, *mi*-rise and *mi*-fall. Each of these two syllables is a homophone for a number of morphemes. The *mi*-rise meanings include "confusion", "rotten", "riddle", etc., and the *mi*-fall can mean "secrecy", "search", etc. These are unlikely to be the kind of vocabulary used in parental speech directed to preverbal infants. In addition to these two tones in citation syllables, nonsense utterances were constructed, each of which contained a target "*mi*" preceded by another syllable "*fu*" (/fu/) bearing one of the four Mandarin tones. There were a total of eight nonsense combinations, four for *mi*-rise utterances and four for *mi*-fall utterances: *fu*-high *mi*-rise, *fu*-rise *mi*-rise, *fu*-low *mi*-rise, *fu*-fall *mi*-rise, *fu*-high *mi*-fall, *fu*-rise *mi*-fall, *fu*-low *mi*-fall, *fu*-fall *mi*-fall. The use of such novel utterances made it possible to test infants' generalization ability in tonal categorization. In addition, we avoided the use of longer utterances or sentences, since they would make word segmentation more difficult, especially for younger infants. The simple utterances thus enabled us to minimize the segmentation issue so that we could better determine infants' initial ability to categorize tones in contexts. Only preceding tones were used as it has been shown that preceding tones exert most of the contextual influence on lexical tones (e.g., Xu, 1997).

A female Mandarin native speaker recorded multiple tokens of the tonal targets *mi*-rise and *mi*-fall, and then tokens of the eight nonsense *fu*+*mi* tonal combinations. The recording took place in an IAC acoustic chamber. Recorded speech was then prepared into auditory stimuli files to be used by the infant experimental software.

2.1.3 Design

There were two phases in the experiment. During the first phase, the Familiarization phase, the infant was presented with trials containing a target tone in citation syllabic forms, *mi*-rise or *mi*-fall. The maximum length for each trial was 14 seconds. The trials played three citation productions of a tonal target repeatedly during the Familiarization phase until the infant accumulated 30 seconds of attentive listening time (measured by looking time to a visual display, i.e., a black-and-white checkerboard, on a central monitor at the sound source). Upon reaching this pre-established familiarization looking criterion, the experiment automatically moved into the Test phase. The Test phase served to assess the recognition of the target tone in variable tonal contexts, where the pitch realizations of the target tone could be significantly altered by the preceding tone, therefore they were much different from the citation forms heard during the Familiarization phase. The Test phase contained two types of trials: disyllabic utterances containing *mi*-rise (four *fu*+*mi*-rise utterances) versus disyllabic utterances containing *mi*-fall (four *fu*+*mi*-fall utterances). The maximum length was 14 seconds for each trial, which presented the four utterances repeatedly in a quasi-random order (with the same exemplar never repeated in adjacent positions). The two trial types were presented alternatingly and repeatedly for five times per type, thus a total of 10 Test trials. The average duration was 675.33 milliseconds for *mi*-rise citation tokens and 683.73 milliseconds for *mi*-fall tokens for the Familiarization trials. The durations for *mi*-rise and *mi*-fall syllables were shorter in *fu*+*mi* utterances during Test trials, with 389.31 milliseconds for *mi*-rise and 397.9 milliseconds for *mi*-fall. The inter-stimulus interval was 500 milliseconds in all Familiarization and Test trials.

In addition to the Familiarization and Test phases, the experiment began with a pre-trial and ended with a post-test trial, during which the same checkerboard was presented together with water bubble sounds. The pre-trial served to acquaint the infant with the experimental procedure, and the post-test trial enabled us to verify if the infant was on task throughout the experiment. If she or he was on task, looking time should recover during the post-test trial relative to the last Test trial, as the auditory stimuli of the post-test trial were clearly noticeable. The bubble sounds in the pre-trial and post-test trial were distinct from the speech stimuli used in the Familiarization and Test phases, so they would not interfere with the experiment.

Infants were randomly assigned to one of the two target tones, half hearing *mi*-rise and half *mi*-fall in citation forms during Familiarization. All infants heard the two trial types during Test. Hence, the target Test trial type (i.e., utterances containing the familiarized target tone) for one group of infants was the non-target for the other group of infants. The first Test trial was either the target trial or non-target trial, counter-balanced across infants.

2.1.4 Procedure and apparatus

A visual fixation procedure was used for this experiment. Infants were tested individually in an IAC acoustic chamber. The infant sat on the parent's lap, about two meters in front of a central monitor. The parent wore noise cancellation headphones and heard masking music. The parent was instructed not to interact with his or her infant during the experiment. Auditory stimuli were presented from two loud-speakers located at the same position as the monitor. The experiments were run with the software Habit (Cohen, Atkinson, & Chaput, 2000) from a MacIntosh computer, which presented each trial and recorded all looks online during each trial. The software also calculated online total looks during the Familiarization phase and advanced the experiment to the Test phase automatically. In the adjacent control room the researcher, who was blind to the order of trial types, observed the infant's eye responses from a closed-circuit TV and pressed down a key on the MacIntosh computer whenever the infant looked towards the monitor. The key presses were recorded automatically into the experimental software. The researcher could not hear the experimental sound, nor see the visual display.

This was a totally infant-controlled experiment. Each trial was controlled by the infant's looks: a trial was initiated contingent upon the infant's look to the central monitor, and a trial was terminated when the infant ceased looking at the monitor. Auditory and visual stimuli were presented in synchrony so that listening preferences could be interpreted based on looking responses. Between trials a red light flashed repeatedly in concert with a cricket sound in order to attract the infant's attention to the monitor. As soon as the infant looked towards the monitor, the attention-getting red light and cricket sound terminated and a trial started.

2.1.5 Predictions

The task of this experiment resembles to some degree target detection/identification tasks in adult speech perception studies. Adults are typically presented with a target (such as a vowel, a tone, a syllable, etc) and are asked to identify it in larger speech materials by pressing a computer key. Such a method is difficult to implement with preverbal infants. Our task is an adaptation of the adult method for infants. The tonal target in citation is presented during the Familiarization phase. Immediately following this phase, the target tokens appear in variable tonal contexts. If infants recognize the same target in contexts, there should be a priming effect. Based on many previous infant word recognition studies (e.g., Jusczyk & Aslin, 1995; Nazzi, et al., 2006), infants typically produce a longer listening time (measured by looking times) for utterances containing the target than they do for utterances not containing the target.

Therefore, we expected that if infants in the present experiment identified the target tone embedded in variable tonal contexts during Test, they should produce a longer looking time while listening to such utterances, as opposed to utterances containing the other tone of the minimal pair. Recall that citation forms of a tone are much different from the pitch patterns of the same tone embedded in varying preceding tonal contexts (e.g., Xu, 1997). Hence, successful tonal categorization across tokens in the present experiment would indicate that infants have an abstract level of tonal representations.

In addition, the Test phase, which contained a total of 10 Test trials, was designed to also assess infants' recognition efficiency. Recognition efficiency in infants has been studied (e.g., Hallé, Durand, & de Boysson-Bardies, 2008; Shi, Werker, & Cutler, 2006; Vihman, Nakai, DePaolis, & Halle, 2004). Following the method of these previous studies, we analyzed total looking times across all Test trials, as well as first and second parts of the Test phase. If infants immediately identified the target tone, looking time differences should be observed already early in the Test phase. But if infants' ability to categorize tones in variable tonal context is not yet robust, they should exhibit a looking time difference only later in the experiment.

2.2 Results

In the first analysis each infant's total looking times for trials containing the familiarized tonal target and his or her total looking time for trials containing the non-familiarized contrasting tone (i.e., the other of the tonal minimal pair) during the Test phase were computed respectively. An analysis of variance was conducted with Tonal Type (target versus non-target) as the within-subject factor, and Age (four-to-six months versus eight-to-ten months) as the between subject factor. Results revealed neither a main effect of Tonal type, nor Age, nor a Tonal Type x Age interaction. Therefore, total looking times across all Test trials showed no evidence of tonal categorization in variable phonetic contexts.

However, total looking times across all Test trials may not be sensitive enough to reveal subtle categorization abilities that infants may be developing. If infants were able to categorize tones in contexts but were slow in the processing, a differential looking pattern may emerge only towards the later part of the Test phase. Analyses were then conducted on the initial six trials and the final four trials of the Test phase respectively. For the younger infants, listening times (indicated by looking times) to target sequences and to non-target sequences were neither different for the initial six trials ($t(8) = .124, p = .904$), nor for the final four trials ($t(8) = .4, p = .699$). For the older infants, listening time for target sequences was significantly longer than that for non-target sequences for the last four trials of the Test phase ($t(8) = 2.332, p = .048$), although these infants' looking times to the two types of sequences did not differ during the initial six Test trials ($t(8) = .24; p = .816$), as shown in Figure 1 and 2.

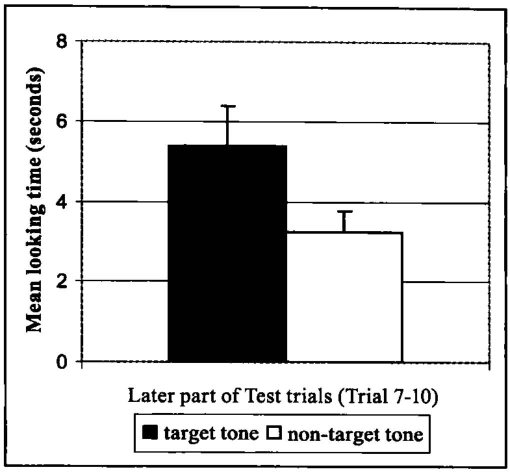


Figure 1: Mean looking times (and standard errors) produced by older infants during the later part of the Test phase (Trials 7 to 10) for utterances containing the target tone and those containing the non-target tone of the minimal pair. Infants showed a preference for trials containing the target.

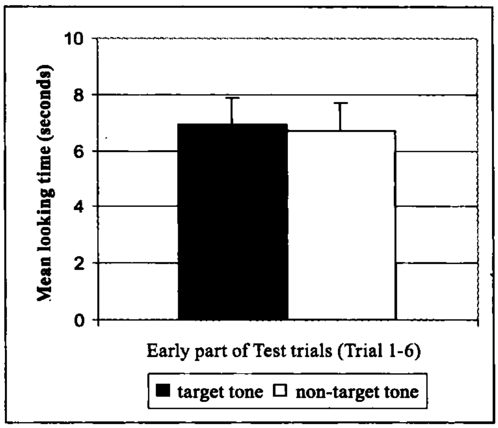


Figure 2: Mean looking times (and standard errors) produced by older infants during the early part of the Test phase (Trials 1 to 6) for utterances containing the target tone and those containing the non-target tone of the minimal pair. No difference was present.

This suggests that infants of the two age groups responded differently. At a younger age infants could not categorize tones in variable tonal contexts. This ability seems to emerge by eight to 11 months of age even though the performance at this age is not yet efficient. Infants needed to listen to several Test trials and compare the target tone with the contrasting tone of the minimal pair before beginning to exhibit a preferential response to the utterances containing the familiarized target tone. These findings support our hypothesis that categorization of tones in variable tonal contexts is a gradual learning process.

3. DISCUSSION AND CONCLUSION

Understanding the variability problem is one of the major challenges in the studies of adult speech perception and infant speech development. The problem is particularly hard for infants, who receive no parental instruction and have limited upper level linguistic knowledge to help. On the one hand, infants must learn to treat variations within a phonetic category as unimportant, whether such variations are acoustically small or large. On the other hand, they must learn to treat variations across phonetic categories as important even though the difference may be less or equal to within-category variations. Certain types of variability may be context independent such as inherent VOT variations for stop consonants, which are due to the inability for the vocal apparatus to be positioned in exactly the same way with exactly the same timing control for all productions of a phonetic category. Natural differences between speakers, which are context independent, can lead to large variability for phonetic categories such as vowels and tones. The present study concerns context dependent variability, i.e., contextual influence on the acoustic forms of a category. Due to the parallel nature of phonetic encoding (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), extensive variability can be observed in the realization of a given phonetic category. Aside from recent simulation studies (Gauthier, Shi & Xu, 2007a, 2007b), no studies known to us have directly tested infants' phonetic category perception in variable contexts.

Using lexical tones as a testing case, the present experiment showed that Mandarin-learning infants at eight to 11 months of age, but not at four to six months of age, begin to categorize dynamic rise and fall tones in Mandarin in variable tonal contexts. They treat tokens of a tone in contexts as the same class as citation tokens of the same tone. This indicates some level of abstract tonal representations in infants. Whereas the tokens in citation and in contexts are quite different in surface fundamental frequency patterns (Xu, 1997), one common property for all these tokens lies in the velocity profile of the fundamental frequency, i.e., the movement information that reflects the speaker's intended target (Gauthier, Shi, & Xu, 2007a). Infants in the present experiment were likely to have used the velocity profile information to normalize contextual variability.

The finding that only older infants showed evidence of categorization suggests that the ability to form phonetic categories in variable coarticulated contexts requires a period of learning. Evidence of a gradual learning process was also observed in the results of earlier versus later part of the testing phase in older infants. While infants did categorize the target tones, this ability was only evident in the later part of the experiment, indicating that the ability is not yet robust at this age. We predicted that efficiency should improve with additional learning experience.

The findings of this study have relevance to the mechanisms of language acquisition. There has been a long debate over the role of nature versus nurture in language development. Nativist approaches (e.g., Chomsky & Halle, 1968) maintain that linguistic structures are innately endowed to humans, and that language acquisition is a process of mapping the input speech onto these innate representations rather than creating representations from induction. Based on empiricist views, however, linguistic representations can be induced and constructed from the input. Findings of initial language general phonetic perception to later language specific perception (e.g., Werker &

Tees, 1984) have been interpreted by some researchers as evidence supporting nativist approaches. But it can be argued that these findings are equally compatible with the view that there are certain innately biased general auditory processes, rather than innate linguistic representations unique to humans (see Aslin, Werker, & Morgan 2002 for a detailed discussion). For example, speech perception abilities observed in newborn infants and adults was also shown in chinchillas (Kuhl & Miller, 1975).

The present study demonstrates that the ability to categorize lexical tones in variable tonal context comes after many months of language exposure. Whether this development reflects a process of inductive learning as viewed by empiricists or a process of mapping input structure to existing linguistic categories needs to be further investigated. Testing inductive learning versus mapping is known to be an experimentally difficult challenge. The data of this study do not offer a definitive test of this question. But they do suggest that the type of biases that allow newborn infants to categorize context independent phonetic classes is not present at the onset of language development.

4. ACKNOWLEDGMENT

This research was supported by Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada, and Canadian Foundation for Innovation. The author thanks Bruno Gauthier for his assistance in stimuli preparation and subject testing.

5 REFERENCES

[1] Aslin, R. N., Werker, J. F., & Morgan, J. L. 2002. Innate phonetic boundaries revisited. *Journal of the Acoustical Society of America, 112*(4), 1257-1260.

[2] Bijeljac-Babic, R., & Nazzi, T. 2007. French-learning infants' sensitivity to lexical stress. Presentation at the 32nd Boston University Conference on Language Development.

[3] Bortfeld, H., Morgan, J., Golinkoff, R., & Rathbun, K. 2005. Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298-304.

[4] Chomsky, N. A., & Halle, M. 1968. *The sound pattern of English.* New York, NY: Harper & Row.

[5] Cohen, L. B., Atkinson, D. J., & Chaput, H. H. 2000. Habit 2000: A new program for testing infant perception and cognition. [Computer software]. Austin, the University of Texas.

[6] Eimas, P. D., & Miller, J. L. 1980. Contextual effects in infant speech perception. *Science, 209*, 1140-1141.

[7] Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. 1971. Speech perception in infants. *Science, 171*(3968), 303-306.

[8] Gandour, J., Potisuk, S., & Dechongkit, S. 1994. Tonal coarticulation in Thai. Journal of Phonetics, 22, 477-492.

[9] Gauthier, B., Shi, R., & Xu, Y. 2007a. Learning phonetic categories by tracking movements. *Cognition, 103*(1), 80-106.

[10] Gauthier, B., Shi, R., & Xu, Y. 2007b. Simulating the acquisition of lexical tones from continuous dynamic input. *Journal of the Acoustical Society of America, 121*(5), EL190-EL195.

[11] Hallé, P., Durand, C., & de Boysson-Bardies, B. 2008. Do 11-month-old French infants process articles? *Language and Speech,*, *51*, 23-44.

[12] Harrison, P. 2000. Acquiring the phonology of lexical tone in infancy. *Lingua, 110*, 581-616.

[13] Höhle, B., & Weissenborn, J. 2003. German learning infants' ability to detect unstressed closed class elements in continuous speech. *Developmental Science, 6*, 122-127.

[14] Jusczyk, P., & Aslin, R. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

[15] Jusczyk, P. W., Cutler, A., & Redanz, N. J. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

[16] Kuhl, P. K. 1979. Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66*, 1668-1679.

[17] Kuhl, P. K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93-107.

[18] Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

[19] Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*, F13-F21.

[20] Kuhl, P. K., & Miller, J. D. 1975. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69-72.

[21] Liberman, A., F. Cooper, D. Shankweiler & M. Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review 74*(6), 431-611.

[22] Mattock, K., & Burnham, D. 2006. Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3). 241-264.

[23] Mattock, K., Molnar, M., Polka, L., & Burnham, D. 2008. The developmental course of lexical tone perception in the first year of life. *Cognition, 106*.

[24] Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. 2006. Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*, 283-299.

[25] Polka, L., & Werker, J. F. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435.

[26] Shi, R., Morgan, J. L., & Allopenna, P. 1998. Phonological and acoustic bases for early grammatical assignment: A cross-linguistic perspective. *Journal of Child Language, 25*, 169-201.

[27] Shi, R., Cutler, A., Werker, J., & Cruickshank, M. 2006. Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America, 119*, EL61-EL66.

[28] Shi, R., & Lepage, M. 2008. The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science, 11*(3), 407-413.

[29] Shi, R., Werker, J., & Cutler, A. 2006. Recognition and representation of function words in English-learning infants. *Infancy, 10*, 187-198.

[30] Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E. Sebastian-Galles, N., Limissuri, R.A., & Peperkamp, S. 2007. Presentation at the 32nd Boston University Conference on Language Development.

[31] Sundara, M., Polka, L., & Genesee, F. 2006. Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition, 100*(2), 369-388.

[32] Trehub, S. E., Bull, D., & Thorpe, L. A. 1984. Infants' perception of melodies: The role of melodic contour. *Child Development, 55*, 821-830.

[33] Vihman, M. M., Nakai, S., DePaolis, R. A., & Halle, P. 2004. The role of accentual pattern in early lexical representation. *Journal of Memory and Language, 50*, 336-353.

[34] Weijer, J. van de 1998. *Language input for word discovery.* Unpublished doctoral dissertation, University of Nijmegen, Nijmegen, The Netherlands.

[35] Werker, J.F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. 1981. Developmental aspects of cross-language speech perception. *Child Development, 52*, 349-355.

[36] Werker, J. F. & Tees, R.C. 1984. Cross-language speech perception: Evidence for perceptual reorganization in the first year of life. *Infant Behavior and Development, 7*, 49-63.

[37] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics 25*, 61-83.

[38] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics, 27*, 55-105.

SHI Rushen, Département de psychologie, Université du Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3P8.

Email：shi.rushen@uqam.ca

PERCEPTUAL TRANSCRIPTION AND ACOUSTIC DATA: THE EXAMPLE OF /i/ IN YONGNING NA (TIBETO-BURMAN)*

Alexis MICHAUD, Jacqueline VAISSIÈRE

Abstract: On the basis of fieldwork on a Tibeto-Burman language, Yongning Na, some reflections are offered on three interrelated issues: (i) To what extent does perceived allophonic variation correspond to articulatory/acoustic reality, to what extent does it merely reflect the investigator's perceptual expectations? (ii) How can acoustic data help select International Phonetic Alphabet symbols for the phonemic units brought out by distributional analysis? (iii) How can acoustic data help characterize the vowels and consonants encountered in fieldwork, both in a structural (language-internal) perspective, and in a cross-language perspective?
Keywords: phonemics, acoustics, allophony, perception, Na/Naxi/Mosuo

1. INTRODUCTION: USEFULNESS OF ACOUSTIC DESCRIPTION

The selection of International Phonetic Alphabet symbols for the phonemes of a language is essentially based on structural arguments and expert listening, rather than on acoustic analysis and modelling. This is true both for the most documented languages, because the descriptions that serve as standards are usually based on a tradition that predates acoustic phonetics, and for newly described languages, because acoustic analysis is seldom a priority in fieldwork. IPA notations, which are oriented towards the transcription of phonemic oppositions, offer some indications on articulatory characteristics, but they clearly do not encapsulate enough information to reproduce the sounds under study.

By contrast, the acoustic theory of speech production [2] allows for a representation of acoustico-perceptual characteristics of speech sounds by reference to the resonantial properties of the vocal tract: the F-pattern. Unlike IPA notation, a characterisation in terms of F-pattern is sufficiently detailed to serve as input for speech synthesis, via articulatory modelling [8]. (To be quite precise: the F-pattern needs to be supplemented by information on nasality, if any, and on phonation type.) In the belief that speech sounds can be characterised very accurately in terms of a target F-pattern and of the articulatory configuration used to attain this pattern, the second author of this article is currently developing a notation [11] to describe the phonemes of individual languages by reference to certain fixed acoustical properties. This project is reminiscent of that of Daniel Jones in defining the cardinal vowels, insofar as the aim consists in providing a stable reference point for linguistic description. To our knowledge, the cardinal vowels are defined in articulatory

* Many thanks to Latami Dashi (拉他咪·达石) for his help and advice during fieldwork, and to the language consultants for their hospitality and their patience. Our thanks to Liberty Lidz for useful discussions about the phenomena that she observed in Luòshuǐ and elsewhere. The authors alone are responsible for any errors.

terms, and the audio renderings proposed to illustrate them fluctuate somewhat, e.g. some of the cardinal vowels recorded by Peter Ladefoged are somewhat different from those of Daniel Jones [11]. It appears worth investigating in which ways an acoustic definition can supplement the IPA notations for a given language, and whether this facilitates cross-language comparisons.

As a tentative step towards implementing the research agenda outlined above, the present article exemplifies how acoustic observations can help overcome difficulties encountered in fieldwork. The object of study is the Na language (Tibeto-Burman family) as spoken in Yongning, China (Chinese coordinates: 云南省丽江市宁蒗彝族自治县永宁乡平静村). This language is also known as Mosuo (摩梭话) or Eastern Naxi (纳西语东部方言).

The tonal system of Yongning Na is currently under analysis. It appears to be based on two simple tones, H(igh) and L(ow), and two contours, L-to-M(id) and M-to-H. These four patterns, combined with a 'zero' value (absence of tonal specification), yield five possibilities for monosyllables, illustrated respectively by the words /ʐwæ/ 'horse', /ŋv̩/ 'silver', /bu/ 'pig', /ɬv̩/ 'brains', and /lɑ/ 'tiger'. On disyllables, these five patterns are also attested; in addition, there also exist six more patterns, four of which arise through compounding (word-final L; word-final H; L+MH; L+H; LM+L) whereas the status of the last two remains unclear. A detailed description of this tonal system will be set out elsewhere. In the present article, tone is indicated in superscript before the word.

The acoustic analysis bears on the close, front vowel of Yongning Na, transcribed as /i/.

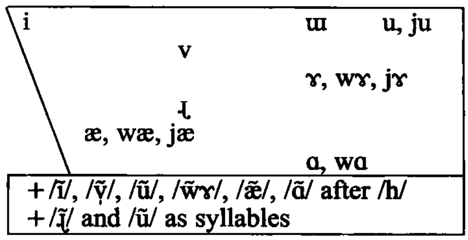
2. THE ISSUE: COARTICULATION, OR CATEGORICAL ALLOPHONY?

One of the issues encountered in fieldwork on Yongning Na was the notation of high vowels. The first, tentative notations included [e] as well as [i], [o] as well as [u]. Looking back at these transcriptions, it appears that close vowels [i] and [u] are in complementary distribution with close-mid vowels [e] and [o]. Table 1 shows their distribution. Table 2 presents an inventory of rhymes.

Table 1: Distribution of close and mid-close vowels in the first field notations. [1]

rhyme	[e], [o]	[i], [u]
initial consonant	retroflexes (all modes); dental fricatives and affricates	labials; dental stops; laterals; alveolo-palatals; velars; nasals; glottal; no initial

Table 2: A preliminary inventory of the rhymes of Yongning Na.



The standard way of reporting on such observations is to describe the phoneme at issue as having two allophones, and to make a reasoned choice of one of these allophones to represent the phoneme: for instance, one could choose to transcribe /i/ for [i, e], and /u/ for [u, o]. Analysis as two allophones amounts to saying that the difference between the phonetic realisation of these phonemes in the two sets of consonantal contexts listed in table 1 goes beyond what can be predicted from coarticulatory properties of the preceding consonant: that the two contextual variants are phonetically different sounds.

This raises an important issue: does our classification of the realisations of one phoneme into two phonetic categories – e.g. transcribing [i] in [ᴹi] 'spot, pimple' vs. [e] in [ᴴʈʂe] 'earth' – reflect an acoustic difference between these two realisations in Yongning Na, or does it merely reflect our perception as non-

[1] Note that [7] does have a contrast between /i/ and /e/: L. Lidz analyses the 'fricative vowels' [ɿ] and [ʅ] as allophones of /i/ (e.g. [dzɿ], as in the verb 'to eat', is analysed as /dzi/, whereas we analyse it as /dzɯ/); the syllables where we have /dzi/, e.g. /dzi.lɯ/ 'wheat', are transcribed by [7] with an /e/.