Zhang, Y. (2016). Categorical Perception. In Sybesma, R., Behr, W., Gu, Y., Handel, Z., Huang, J., Myers, J. (Eds.), Encyclopedia of Chinese Language and Linguistics , Leiden, Netherlands: Brill.

Available online: http://dx.doi.org/10.1163/2210-7363_ecll_COM_000071

Categorical Perception

Yang Zhang
Department of Speech-Language-Hearing Sciences
Center for Neurobehavioral Development
University of Minnesota
Minneapolis, MN, 55455
USA
Correspondence: zhanglab@umn.edu

Abstract

Categorical perception provides an account for how human symbolic thinking is grounded in perception and action. The study of Chinese lexical tones offers a unique venue to investigate the origin and development of CP in the brain and its importance in language and cognition.

Key words

Categorical Perception; Lexical Tones; Neuroimaging; Language Acquisition

The acoustic realization of consonants and vowels in a language is highly variable depending on the speaker, the phonological context, and the speaking rate. Categorical perception (CP) is a mechanism to explain how the speech signal with infinite variability is perceptually mapped onto a finite set of discrete linguistic symbols. Although the initial studies only tested English material and the research field is dominated by English-based studies, CP is considered to be a universal feature of speech perception that applies to all languages, including Chinese with its distinctive lexical tone system.

The discovery of CP was made possible by the invention of a formant speech synthesizer known as the Pattern Playback at Haskins Laboratories, and it involved two behavioral tasks (Liberman, Harris, Hoffman and Griffith, 1957). The identification task presents sounds from a synthetic speech continuum and asks listeners to label them. The discrimination task tests sounds in pairs or triplets to see whether listeners perceive differences in the sounds. The typical CP results are characterized by a phonetic boundary effect, which shows not only a labeling function with sudden membership change along the continuum, but also a discrimination function with a distinct peak at the category boundary (Fig. 1). The essence of the CP phenomenon is that within-category discrimination is very poor whereas across-category discrimination is significantly better despite the fact that the physical differences are synthetically controlled to be equal. In other words, listeners perceptually assimilate stimuli from the same speech category, making them much harder to distinguish than stimuli from different categories. This pattern of auditory discontinuity is opposed to continuous perception, where discrimination performance along the continuum is a linear function of the physical differences involved and is independent of identification.

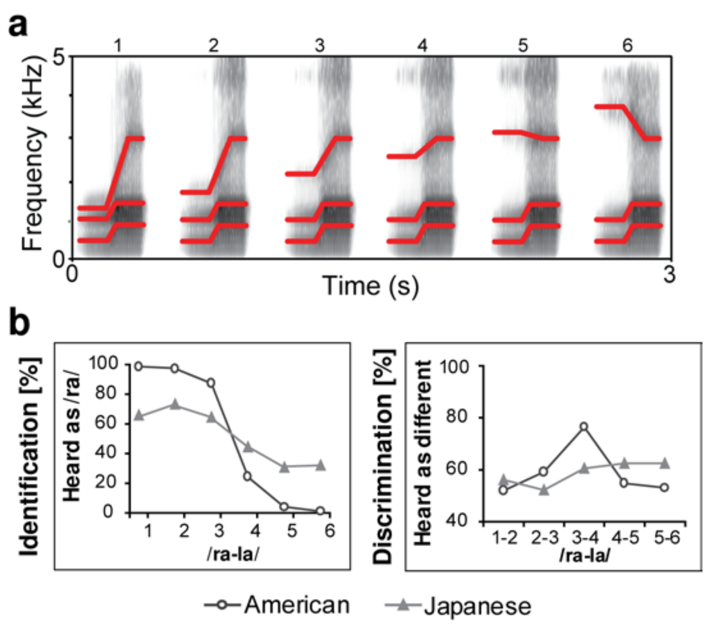

Figure 1. Results from an adult cross-language CP experiment (Zhang et al., 2005). (a) Spectrogram for the /ra-la/ continuum that contains systematic manipulation of the third formant in equal steps moving from /ra/ to /la/. (b) Identification and discrimination functions showing the sharp crossover "phonetic boundary effect" for CP in the American subjects but not in the Japanese subjects.

The CP evidence allows theorists to account for how human symbolic thinking is grounded in perception and action via the interplay between higher-level conceptual systems and lower-level perceptual systems. Insights as well as controversies about the origin and nature of CP come from investigations of human and nonhuman subjects beyond the scope of speech and language, including modern brain imaging and computational modeling techniques. For example, animal research indicates that CP may partly reflect the existence of natural acoustic partitioning boundaries in the mammalian auditory system (Kuhl and Miller, 1978; Steinschneider, Schroeder, Arezzo and Vaughan, 1995). Early developmental work shows that infants as young as 1 month old can categorically perceive the /ba-pa/ continuum (Eimas, Siqueland, Jusczyk and Vigorito, 1971).However, vowel discrimination appears to be less categorical in both children and adults (Fry, Abramson, Eimas and Liberman, 1962; Pisoni, 1973; Swoboda, Morse and Leavitt, 1976), which may be attributed to the differences in short-term memory of rapid formant transition for consonants and steady formant structure for vowels. More recent studies further indicate that infants' discrimination of consonants may not be as strictly categorical as previously thought (Maye, Werker and Gerken, 2002; McMurray and Aslin, 2005), suggesting a maturational trajectory of CP that is shaped by the ambient language

An important aspect of CP is how perceptual warping of equal physical distances in the synthetic speech continuum is influenced by the listener's linguistic knowledge and experience. For instance, auditory deprivation may result in complete absence of CP behavior in deaf animals as well as in humans with severe hearing loss early in life (Périer, Alegria, Buyse, D'Alimonte, Gilson and Serniclaes, 1984). Cross-language studies suggest the profound effects of language experience on CP. Although 8~10-month-old American and Japanese infants show similar discrimination for the phonemic /r-l/ contrast that exists in English but not in Japanese (Kuhl, Stevens, Hayashi, Deguchi, Kiritani and Iverson, 2006), only adult American listeners demonstrate CP for the /ra-la/ continuum (Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura, 1975; Zhang, Kuhl, Imada, Kotani and Tohkura, 2005) (Fig. 1). It has also been demonstrated that early bilinguals, late bilinguals, and monolinguals have distinct patterns of speech discrimination for the same set of stop sounds (Sundara and Polka, 2008).

The experience-driven perceptual warping in CP is closely related to another phenomenon known as the Perceptual Magnet Effect (PME) (Kuhl, 1991). PME shows that listeners develop senses of category goodness (or prototypicality) about individual speech tokens, which lead to reduced sensitivity to acoustic differences near the best instances (prototypes) and increased sensitivity in the vicinity of poor exemplars. Given that the endpoints of a synthetic speech continuum are typically classified as the prototypes for the given stimulus set, both CP and PME theories can demonstrate reduced discriminability near the prototypical sounds and enhanced discriminability near phonetic boundary even though the mechanisms have been argued to be separate (Iverson and Kuhl, 2000). Speech scientists have recently offered a unifying Bayesian statistical model to address some discrepancies in the empirical data(Feldman, Griffiths and Morgan, 2009). The Bayesian model attempts to explain how people infer causes from effects and update their rationalization with new evidence. For instance, how would we learn the probability of a future event occurring if we only know how many times it occurred or did not occur in the past? In the Bayesian approach, listeners are ideal observers, who consistently treat perception as a probabilistic solution on the basis of prior knowledge and a likelihood function for the data to be observed. Perceptual warping in CP and PME would arise as a consequence of finding the optimal solution to the statistical problem of speech perception in “noise.” Bayesian simulations show consistent results with the empirical data from various studies, demonstrating that listener’s perception is biased toward prior category knowledge for identifying and discriminating the stimuli and that their sensitivity to phonetic details is modulated by the joint statistical properties of priors and sensory noise.

Recent renewed interest in CP research has been fueled by the desire to delineate the neural basis for CP with modern brain imaging techniques. A cortical network of superior temporal gyrus (including the transverse temporal gyrus also known as the primary auditory cortex and the posterior superior temporal gyrus also known as the second auditory cortex), superior temporal sulcus, middle temporal gyrus, inferior parietal gyrus, inferior frontal gyrus and precentral gyrus (also known as the primary motor cortex) has been identified to separately contribute to continuous coding of the physical variation in stimuli and the discrete partitioning of the continuum into distinct categories (Chang, Rieger, Johnson, Berger, Barbaro and Knight, 2010; Hutchison, Blumstein and Myers, 2008; Joanisse, Zevin and McCandliss, 2007; Liebenthal, Binder, Spitzer, Possing and Medler, 2005; Möttönen and Watkins, 2009) (Fig. 2). Training data in adulthood further demonstrate substantial brain plasticity in various cortical regions, but the reported failure to achieve native-like CP behavior with enhanced across-category discrimination relative to within-category discrimination indicates the difficulties in second language learning under the constraints of the adult perceptual system that is optimally tuned for and neurally committed to the detection of auditory patterns of the native language (Callan, Tajima, Callan, Kubo, Masaki and Akahane-Yamada, 2003; Zhang, Kuhl, Imada, Iverson, Pruitt, Stevens, Kawakatsu, Tohkura and Nemoto, 2009).

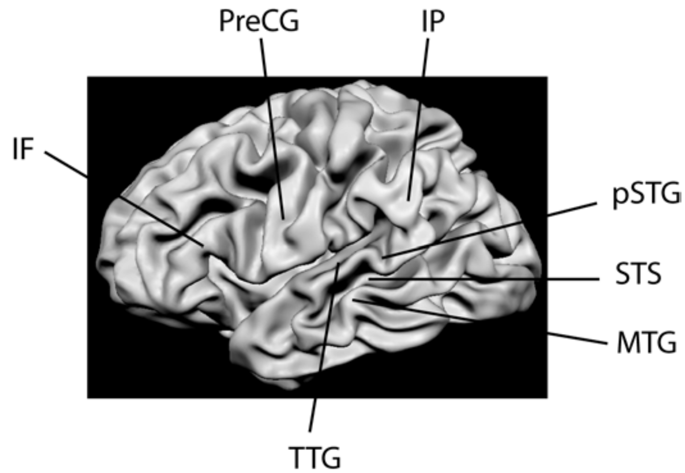

Figure 2. Cortical network of brain regions supporting CP: pSTG = posterior superior temporal gyrus, TTG = transverse temporal gyrus, STS = superior temporal sulcus, MTG = middle temporal gyrus, IP = inferior parietal gyrus, IF = inferior frontal gyrus, preCG = precentral gyrus (primary motor cortex). The TTG region (i.e., the primary auditory cortex) primarily reflects the acoustic mode of continuous perception, and the other cortical regions have been found to contribute to the phonetic boundary effect that is characteristic of CP. Lateralization in favor of the left hemisphere is found for CP except the inferior frontal cortex.

Collectively, the English-based empirical data indicate that although there is an innate basis for the CP behavior, language environment and learning experience fundamentally shapes the way the brain represents and processes the category structures of speech sounds. In this regard, studies using different language samples are necessary to test the universality of CP as a fundamental property of the speech perception system. The lexical tones in Chinese offer a unique venue to investigate CP at the suprasegmental level (Chan, Chuang and Wang, 1975; Hallé, Chang and Best, 2004; Huang and Johnson, 2010; Peng, Zheng, Gong, Yang, Kong and Wang, 2010; Xi, Zhang, Shu, Zhang and Li, 2010; Xu, Gandour and Francis, 2006; Zheng, Minett, Peng and Wang, 2012). Although some studies were able to replicate the CP behavior for stop consonants in normal native Chinese adults as well as children (Cheung, Chung, Wong, McBride-Chang, Penney and Ho, 2009; Liu, Shu and Yang, 2009; Xi, Jiang, Zhang and Shu, 2009), the primary goal of all the published studies was focused on lexical tones. This narrow interest may be a consequence of the pressure to meet the novelty criterion for journal publication, and it may also reflect the technical difficulty in creating the Chinese-based synthetic continuum suitable for a CP study.

There are a number of important theoretical issues concerning lexical tone perception that have not been previously addressed in the English-based CP studies. These issues include: (1) to what extent processing of lexical tones involves language-specific or general auditory mechanisms for pitch information, (2) how tonal knowledge is influenced by different experiential factors such as language learning, musical training and pathological condition, (3) how tonal knowledge is acquired in the developing brain and mature brain, and (4) whether lexical tone processing is predictive of higher-level language skills such as reading.

The general consensus of the Chinese studies is that CP is observed in continua involving contour tones and, to a lesser degree, for level tones. This perceptual pattern is consistent with the differential CP test results for English consonants and vowels. The level tones contain long and stable fundamental frequencies (F_0), which are analogous to the steady state formants in simple vowels. In contrast, the contour tones show fast-changing dynamic F_0 analogous to the formant transition pattern in consonants. The steady state acoustic cues allow an "acoustic mode" of listening for the steady F_0 information and thereby reduce the phonetic boundary effect for CP. The transient acoustic cues may force listeners to rely on a perceptual labeling strategy due to difficulties in keeping track of the transient and detailed F_0 changes in short-term auditory memory.

As in the previous English-based studies, the experiential factors play an important role in determining the presence/absence of CP as well as the exact location of the phonetic boundary in the synthetic continuum of contour tones. For instance, native Korean (a non-tonal language) speakers did not show the CP behavior for contour tones (Xi et al., 2010). Similarly, Mandarin speakers and Cantonese speakers show differences in their CP responses due to the different lexical tonal systems in the two dialects (Peng et al., 2010; Zheng et al., 2012). From a developmental perspective, Chinese studies have revealed immaturity of CP for lexical tones before the age of 6 (Liu et al., 2009). Research on clinical populations further demonstrates impaired CP for lexical tones in Chinese-speaking dyslexic children, suggesting that higher-level linguistic skills such as reading may be built on basic phonological processing skills (Zhang, Zhang, Shu, Xi, Wu, Zhang and Li, 2012).

A third important finding from the Chinese studies is that the CP response pattern for the contour tones appears to transfer to pitch processing for intonations as well as nonspeech stimuli. For instance, native Chinese speakers demonstrated CP behavior for intonation contour patterns just like lexical tone contrasts (Liu and Rodriguez, 2012). They also showed enhanced neural sensitivity to across-category change detection relative to within-category discrmination for both speech and nonspeech stimuli (Xi et al., 2010). Moreover, Chinese-speaking congenital amusics, also known as people who are "tone-deaf" from birth with pitch discrimination substantially below the normal range, did not show CP for lexical tones (Jiang, Hamm, Lim, Kirk and Yang, 2012). These data all point to shared roots for processing pitch information in both linguistic and non-linguistic domains.

Given the limited scope of lexical tones, it is desirable to design new CP studies using other important phonemic contrasts in Chinese. Further research is needed to map out the developmental trajectory of CP and understand the associated brain mechanisms such as hemispheric specialization and regional specificity for the range of gradient responses between canonical CP and non-CP behaviors in individual subjects. It is also important to address similarities and differences in linguistic and nonlinguistic processing of pitch information and how behavioral performance is constrained by various experiential factors as a function of age, language experience, and pathological condition.

References

Callan, Daniel E., Keiichi Tajima, Akiko M. Callan, Rieko Kubo, Shinobu Masaki and Reiko Akahane-Yamada, "Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast", *Neuroimage* 19, 2003, 113-124.

Chan, Stephen W., Chiu-Kuang Chuang and William S. Y. Wang, "Cross-linguistic study of categorical perception for lexical tone", *Journal of the Acoustical Society of America* 58, 1975, S119.

Chang, Edward F., Jochem W. Rieger, Keith Johnson, Mitchel S. Berger, Nicholas M. Barbaro and Robert T. Knight, "Categorical speech representation in human superior temporal gyrus", *Nature Neuroscience* 13, 2010, 1428-U1169.

Cheung, Him, Kevin K. H. Chung, Simpson W. L. Wong, Catherine McBride-Chang, Trevor B. Penney and Connie S. H. Ho, "Perception of tone and aspiration contrasts in Chinese children with dyslexia", *Journal of Child Psychology and Psychiatry* 50, 2009, 726-733.

Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk and James Vigorito, "Speech perception in infants", *Science* 171, 1971, 303-306.

Feldman, Naomi H., Thomas L. Griffiths and James L. Morgan, "The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference", *Psychological Review* 116, 2009, 752-782.

Fry, Dennis B., Arthur S. Abramson, Peter D. Eimas and Alvin M. Liberman, "The Identification and Discrimination of Synthetic Vowels", *Language and Speech* 5, 1962, 171-189.

Hallé, Pierre A., Yueh C. Chang and Catherine T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners", *Journal of Phonetics* 32, 2004, 395-421.

Huang, Tsan and Keith Johnson, "Language specificity in speech perception: Perception of mandarin tones by native and nonnative listeners", *Phonetica* 67, 2010, 243-267.

Hutchison, Emmette R., Sheila E. Blumstein and Emily B. Myers, "An event-related fMRI investigation of voice-onset time discrimination", *Neuroimage* 40, 2008, 342-352.

Iverson, Paul and Patricia K. Kuhl, "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?", *Perception & Psychophysics* 62, 2000, 874-886.

Jiang, Cunmei, Jeff P. Hamm, Vanessa K. Lim, Ian J. Kirk and Yufang Yang, "Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics", *Memory and Cognition*, 2012.

Joanisse, Marc F., Jason D. Zevin and Bruce D. McCandliss, "Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm", *Cerebral Cortex* 17, 2007, 2084-2093.

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

Kuhl, Patricia K. and James D. Miller, "Speech perception by the chinchilla: identification functions for synthetic VOT stimuli", *Journal of the Acoustical Society of Amerca* 63, 1978, 905-917.

Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani and Paul Iverson, "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months", *Developmental Science* 9, 2006, F13-F21.

Liberman, Alvin M., Katherine S. Harris, Howard S. Hoffman and Belver C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries", *Journal of Experimental Psychology* 54, 1957, 358-368.

Liebenthal, Einat, Jeffrey R. Binder, Stephanie M. Spitzer, Edward T. Possing and David A. Medler, "Neural substrates of phonemic perception", *Cerebral Cortex* 15, 2005, 1621-1631.

Liu, Chang and Amanda Rodriguez, "Categorical perception of intonation contrasts: effects of listeners' language background", *Journal of the Acoustical Society of America* 131, 2012, EL427-433.

Liu, Wenli, Hua Shu and Yufang Yang, "Speech perception deficits by Chinese children with phonological dyslexia", *Journal of Experimental Child Psychology* 103, 2009, 338-354.

Maye, Jessica, Janet F. Werker and LouAnn Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination", *Cognition* 82, 2002, B101-111.

McMurray, Bob and Richard N. Aslin, "Infants are sensitive to within-category variation in speech perception", *Cognition* 95, 2005, B15-26.

Miyawaki, Kuniko, Winifred Strange, Robert Verbrugge, Alvin M. Liberman, James Jenkins and Osamu Fujimura, "An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English", *Perception & Psychophysics* 18, 1975, 331-340.

Möttönen, Riikka and Kate E. Watkins, "Motor Representations of Articulators Contribute to Categorical Perception of Speech Sounds", *The Journal of Neuroscience* 29, 2009, 9819-9825.

Peng, Gang, Hong-Ying Zheng, Tao Gong, Ruo-Xiao Yang, Jiang-Ping Kong and William S. Y. Wang, "The influence of language experience on categorical perception of pitch contours", *Journal of Phonetics* 38, 2010, 616-624.

Périer, Olivier, Jesus Alegria, Monique Buyse, Graziella D'Alimonte, Didier Gilson and Willy Serniclaes, "Consequences of auditory deprivation in animals and humans", *Acta Otolaryngologica Supplement* 412, 1984, 60-70.

Pisoni, David B. , "Auditory and phonetic memory codes in the discrimination of consonants and vowels", *Perception & Psychophysics* 13, 1973, 253-260.

Steinschneider, Mitchel, Charles E. Schroeder, Joseph C. Arezzo and Herbert G. Vaughan, Jr., "Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns", *Brain and Language* 48, 1995, 326-340.

Sundara, Megha and Linda Polka, "Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction", *Cognition* 106, 2008, 234-258.

Swoboda, Philip J., Philip A. Morse and Lewis A. Leavitt, "Continuous vowel discrimination in normal and at risk infants", *Child Development* 47, 1976, 459-465.

Xi, Jie, Wei Jiang, Lin-Jun Zhang and Hua Shu, "Categorical Perception of VOT and Lexical Tones in Chinese and the Developmental Course", *Acta Psychologica Sinica* 41, 2009, 572.

Xi, Jie, Linjun Zhang, Hua Shu, Yang Zhang and Ping Li, "Categorical perception of lexical tones in Chinese revealed by mismatch negativity", *Neuroscience* 170, 2010, 223-231.

Xu, Yisheng, Jackson T. Gandour and Alexander L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction", *Journal of the Acoustical Society of America* 120, 2006, 1063-1074.

Zhang, Yajing, Linjun Zhang, Hua Shu, Jie Xi, Han Wu, Yang Zhang and Ping Li, "Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones", *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 53, 2012, 874-882.

Zhang, Yang, Patricia K. Kuhl, Toshiaki Imada, Makoto Kotani and Yohichi Tohkura, "Effects of language experience: neural commitment to language-specific auditory patterns", *Neuroimage* 26, 2005, 703-720.

Zhang, Yang, Patricia K. Kuhl, Toshiaki Imada, Paul Iverson, John Pruitt, Erica Stevens, Masaki Kawakatsu, Yohichi Tohkura and Iku Nemoto, "Neural signatures of phonetic learning in adulthood: A magnetoencephalography study", *Neuroimage* 46, 2009, 226-240

Zheng, Hong-Ying, James W. Minett, Gang Peng and William S. Y. Wang, "The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study", *Language and Cognitive Processes* 27, 2012, 184-209.