Neural distance amplification of lexical tone in human auditory cortex

Xiaopeng Si, Wenjing Zhou and Bo Hong *, *IEEE Member*

***Abstract*—In tonal languages, like Chinese, lexical tone serves as a key feature to provide contrast in word meaning. Behavior studies suggest that Mandarin Chinese tone is categorically perceived. However, the neural mechanism underlying Mandarin tone perception is still poorly understood. In this study, an Oddball paradigm was designed by selecting two standard-deviant stimulus pairs with same physical distance but different category labels, among the synthesized tones with continuously varying pitch contours. Using electrocorticography (ECoG) recording over human auditory cortex, high temporal and spatial resolution cortical neural signals were used for the first time to investigate the cortical processing of lexical tone. Here, we found different neural responses to the two standard-deviant tone pairs, and the difference increased from low to high level along the hierarchy of human auditory cortex. In the two dimensional neural space, cross-category neural distance of lexical tones is selectively amplified on those high level electrodes. These findings support a hierarchical and categorical model of Mandarin tone perception, and favor the using of high-level electrodes for a better performance of lexical tone discrimination in speech brain computer interface.**

I. Introduction

In tonal languages, lexical tone serves as a unique phonetic feature to discriminate word meaning. Chinese, as an example of tonal language, takes vowel, consonant, and lexical tone as building block of its phoneme system, in which lexical tone belongs to supra-segmental feature [1]. Tones are pitch-varying patterns, which are generated by vocal fold vibration resulting in the fundamental frequency changing [2]. In Mandarin Chinese, there are 4 tones: level, rise, fall to rise and fall tone [2]. To investigate the neural mechanism of Chinese tone perception, previous neural imaging studies have mainly focused on hemisphere lateralization of tone processing by using PET, fMRI and scalp EEG, showing that the early pre-attentive stage of tone processing is right hemisphere dominance [3], but when combining lexical information left hemisphere dominates [1,4] . However, the neural mechanism underlying Mandarin tone perception is still poorly understood, especially on where and how tone is processed in human auditory cortex.

It has been shown that Mandarin tone is categorically perceived by both behavioral studies [5-6], and neural imaging studies using fMRI and scalp EEG [7-8]. However, there still lacks direct neurophysiological evidence. According to the hierarchical organization of human auditory cortex, low level auditory cortex such as superior temporal gyrus (STG) is sensitive to spectrotemporal features of sound, whereas high level cortex such as middle temporal gyrus (MTG) is coding higher level perceptual features such as speech objects, and its neural response shows invariance to acoustic physical features [9-10]. Here, we propose a hierarchical and categorical Mandarin tone perception model: neural distance of Mandarin tones that falls in the different category will be amplified as processing from low to high level along the hierarchy of human auditory cortex.

Directly recorded from the surface of human cortex, electrocorticography (ECoG) bears both high spatial resolution (~mm) and high temporal resolution (~ms). Especially, ECoG has been successfully used in the study of finding the encoding scheme of English phonetic features [11]. Here, for the first time, ECoG is employed to reveal the neural representation of lexical tones in Chinese, which may contribute to the speech brain computer interface development.

II. Materials & Methods

A. Subjects

The experiment was conducted with 4 patients (3 male, 1 female, from 20-50 years old) who suffered from intractable epilepsy and underwent temporary placement of ECoG electrode arrays to localize seizure foci prior to surgical resection. Prior to the implantation of electrodes, the patient gave written informed consent for his involvement in research. The experiments were carried out during stable interictal periods. No seizure had been observed 2 hour before or after the tests in any of the patients. Table I shows the detailed information of the subjects. The study was approved by the Research Ethics Committee of Tsinghua University and the affiliated Yuquan Hospital.

B. Behavior study & Stimulus selection

Pilot behavior study on rise-level-fall tone continuum is

TABLE I. CLINICAL DATA OF THE SUBJECTS

Subject No.	Age(yrs)	Sex	Hand	Grid
S1	48	Female	R	Rt
S2	32	Male	R	Rt, Rf
S3	27	Female	R	Lt, Lf, Lo
S4	24	Male	R	Rt

R=right; L=left; t=temporal; f = frontal; o=occipital

*This work was supported by National Program on Key Basic Research Projects of China (2011CB933204) and Ministry of Science and Technology Grant 2012BAI16B03.

B. Hong is with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China (e-mail: hongbo@tsinghua.edu.cn).

X. Si is with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China.

W. Zhou is with the Department of Neurology, Affiliated Yuquan Hospital, Tsinghua University, Beijing 100084, China.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

done to help us get the best stimuli for neural recording (Fig.1a, b). First, 13 tone tokens were synthesized with equal physical distance measured by equivalent rectangular bandwidth (ERB), which is an objective parameter in hearing study [12](Fig.1a). Then, tone identification and discrimination task were performed on 10 normal Chinese subjects to get the behavior performance. Finally, 3 tokens for neural recording were selected based on the identification curve, which formed 2 pairs（standard vs. cross -category deviant，standard vs. within -category deviant）with same physical distance(x axis in Fig.1b) but different perceptual distance (y axis in Fig.1b).

C. Paradigm for ECoG recording

We used the same physical distance but different perceptual category stimuli to perform the ECoG recording under a passive listening Oddball paradigm on epilepsy patient (Fig.1c). There are totally 550 trials, with 80% trials for standard and 10% trials for within-category deviant and 10% trials for cross-category deviant. Inter-stimulus interval was 1100ms with 5% jitter. The experiment program was implemented in Matlab (the Mathworks, USA) using Psychophysics Toolbox 3.0 extensions [13].

D. Data Collection

ECoG signals were recorded using a 96-channel g.USBamp amplifier/digitizer system (g.tec, Graz, Austria). The amplifier sampled at 1200 Hz with a high-pass filter of 0.1 Hz cutoff frequency and a notch filter at 50Hz to remove the power line noise. Four inactive epidural electrodes facing the skull were employed as the ground and the reference. All MRI data were acquired on a Philips Achieva 3.0T TX scanner before implantation. T1-weighted structural MRI images were acquired with an MPRAGE sequence. Imaging parameters were as follows: matrix size: 256×256, voxel size: $0.9\times0.9\times1mm^3$, 180 slices. Following the placement of the subdural grids, the 3D head CT images were obtained by the Siemens SOMATOM Sensation 64 CT for localizing the electrode positions. The CT images were automatically registered to the pre-surgical structural MRI images with Freesurfer [14], which evoked SPM implementation of the mutual information-based transform algorithm [15].

E. ECoG data processing

Auditory responsive electrodes for each subject were identified before further analysis (Fig.2a). First, ECoG data were band-pass filtered to 80-140Hz. Second, the filtered high gamma data were translated into power envelope by taking the absolute amplitude of the analytic signals through Hilbert transform. Third, the high gamma power envelopes were log-transformed so as to follow approximately normal distributions [16]. Then, by using linear regression between the post stimulus energy and pre-stimulus energy, we could get the t-statistics for each electrode. For the same task, the t-statistics were normalized by the maximum among all the electrodes to yield more prominent patterns (Fig.2b). Spectrogram was plotted by using short-time Fourier transform analysis, with a 200ms, 75% overlap Hamming-tapered moving window (t-statistic of time-frequency bins are shown in Fig.2c).

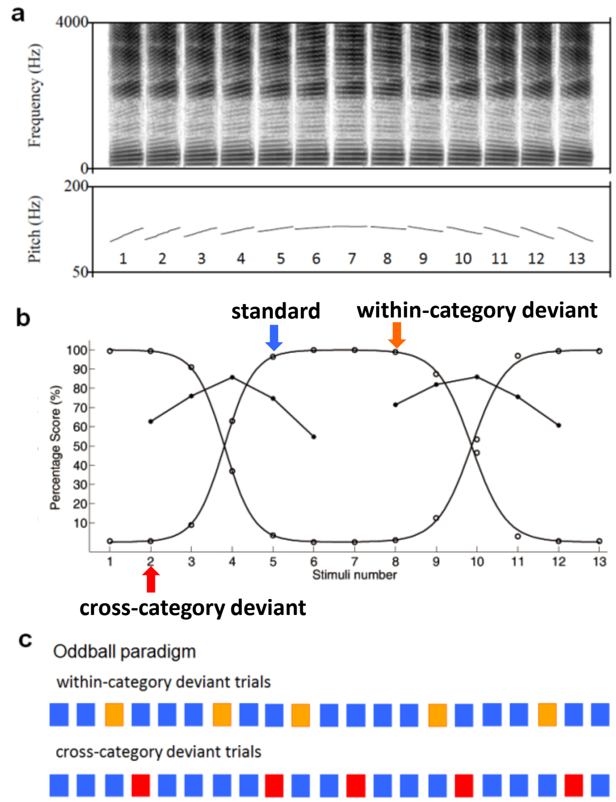

Fig.1. Behavior experiment and stimuli & paradigm design for neural recording. (a) Physcial property (spectrogram and pitch direction) of rise-level-fall tone continuum for behavior experiment. (b) Behavior performance (identification and discrimination task) of Mandarin tone continuum and neural recording stimuli selection. (c) Oddball paradigm for ECoG recording.

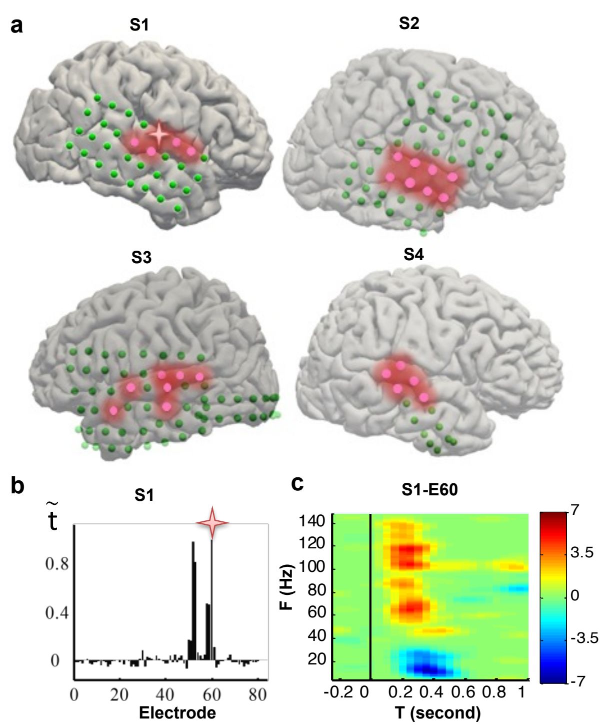

Fig.2. Functional auditory mapping results. (a) Auditory mapping results for 4 patients. (b) Normalized t value bar for subject S1. (c) Spectrogram result for electrode E60 of subject S1.

F. *Event-related high gamma analysis*

Based on the auditory responsive electrodes, event related high gamma changes were analyzed. The log-transformed power envelope of 0–1000ms post-stimulus duration was baseline corrected by subtracting the log-transformed mean power of the 250ms pre-stimulus duration（Fig.3b, c）. High gamma differences between different conditions were characterized by gray area ($p<0.05$) through two sample t-test.

G. *Classification*

Here, linear classifier was used to quantify the neural distance amplification between different stimuli pairs. First, low gamma (30-80Hz) and high gamma (80-140Hz) energy were extracted as two dimensional features (Fig.4c, d), to provide information corresponding to different known neural mechanisms [17]. Then, support vector machine (SVM) with linear kernel was applied for single trial classification. 60% of the data were used as training data set and 40% data as testing data set.

III. Results

A. *Auditory responsive brain areas*

Auditory responsive areas for 4 subjects are showed in Fig.2. These areas include superior temporal gyrus (STG) and middle temporal gyrus (MTG) on both hemispheres, which is consistent with previous results [11].

B. *High gamma envelope reveals categorical neural representation of Mandarin tone*

Fig.3 shows the different high gamma envelopes generated by standard, cross-category deviant and within-category deviant. Although, these two deviants had same physical distance from the standard stimulus, they generated different high gamma deviant responses (Fig.3b, c). Furthermore, difference wave of 'cross-category vs. standard' was significantly larger than that of 'within-category vs. standard)' (Fig.3d), which revealed categorical neural representation of Mandarin tone.

C. *Neural distance amplified as processing from low to high*

Fig.4 shows that neural distance was amplified as processing from low level electrode to high level electrode in the same subject. Fig.4c, d shows that the neural clusters of the high level electrode located in the middle temporal gyrus

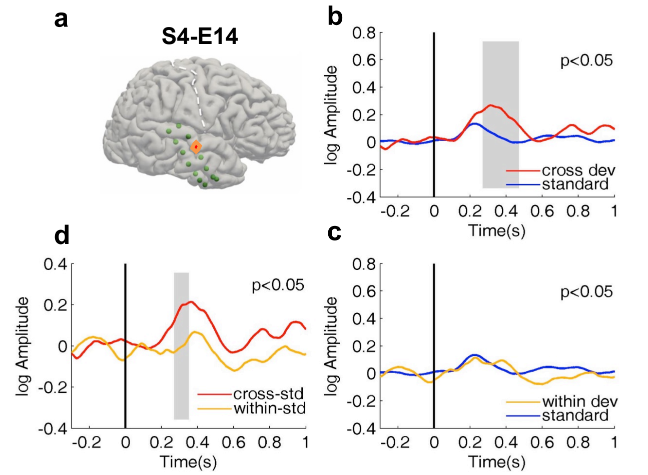

Fig.3. Event-related high gamma ECoG response. (a) Electrode location. (b,c) Averaged high gamma power envelope by standard stimulus (blue), cross-category deviant (red), and within-category deviant (yellow). Gray area indicated sigificiant difference ($p<0.05$). (d) Difference wave by subtracting standard stimulus from cross-category deviant (red) and difference wave by subtracting standard stimulus from within-category deviant (yellow).

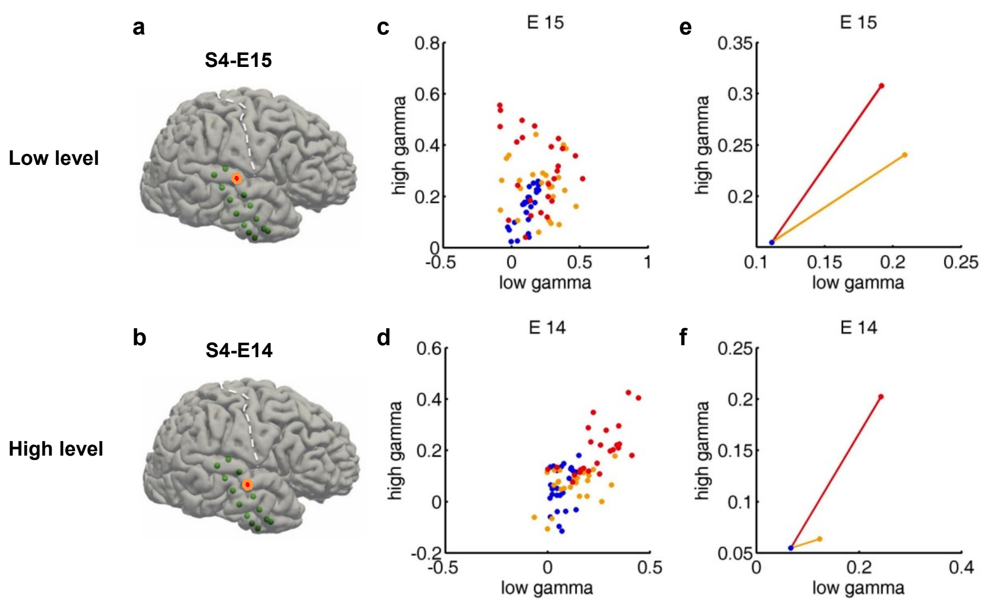

Fig.4. Neural distance amplified as processing from low to high level. (a) Low level electrode location. (b) High level electrode location. (c,d) Different neural clusters for standard (blue), cross-category deviant (red), and within-category deviant (yellow) in two dimensional neural features space. (e,f) Different neural distances resulted from computing Euclidean distance between different neural cluster center. Red is for cross-category vs. standard distance. Yellow is for within-category vs. standard distance.

TABLE II. PERFORMANCE OF CLASSIFICTION

Electrode pairs		Within-category vs. Standard	Cross-category vs. Standard	Difference	Amplification Ratio
S1	E59 ('Low level')	50.38%	49.39%	- 0.99%	0.98
	E60 ('High level')	48.34%	58.58%	+10.24%	1.21
S2	E28 ('Low level')	55.61%	55.65%	+ 0.04%	1.00
	E31 ('High level')	51.46%	66.82%	+15.35%	1.30
S3	E46 ('Low level')	48.74%	52.99%	+ 4.24%	1.09
	E4 ('High level')	54.38%	61.86%	+ 7.47%	1.14
S4	E15 ('Low level')	54.28%	65.81%	+11.53%	1.21
	E14 ('High level')	48.91%	66.00%	+17.09%	1.35

were more separable than that of the low level electrode located in the superior temporal gyrus. Furthermore, Fig.4e, f shows that the neural distances resulted from computing Euclidean distance between different neural cluster centers were also different: cross-category vs. standard distance (red line) was amplified compared to the within-category vs. standard distance (yellow line).

D. *Neural distance amplification revealed by classification*

In this study, linear SVM classifier was used to quantify the neural distance amplification between different stimuli pairs. Performance of classification for all subjects is shown in Table II. By comparing the difference and amplification ratio between low and high level electrodes, we can see clear amplification effect on high level electrodes (Fig.4 f).

IV. Discussion & Conclusion

In this study, by using ECoG recording and auditory functional mapping, human superior temporal gyrus (STG) and middle temporal gyrus (MTG) were found to be involved in speech processing, which is consistent with previous results [11]. High gamma power from the high level auditory cortex such as MTG vicinity was significantly larger for tone pairs that falls in the different category than that falls in the same category, which reveals the categorical neural representation of lexical tone. Furthermore, as processing from low to high along the hierarchy of human auditory cortex, the neural clusters for tone pairs from different category was getting more separable than that from the same category, despite their physical distances are the same. And the neural distance was amplified for cross-category tone pairs from low to high.

For the first time we use both high temporal and spatial resolution ECoG signals to reveal the neural representation of Chinese lexical tone, and these results suggest a hierarchical and categorical lexical tone perception model. This finding also favors the using of electrodes over higher processing areas, like MTG vicinity, for a better performance of lexical tone discrimination in speech brain computer interface.

Acknowledgment

The authors would like to thank Xiaofang Yang and Juan Huang for their help and suggestions on the behavior experiment design, and Chen Song for the comments on ECoG data analysis, and Hao Han, Yang Zhang, Le He and Yandong Zhu for MRI data collection.

References

[1] L. Hsieh, J. Gandour, D. Wong, and G. D. Hutchins, "Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience.," Brain Lang., vol. 76, no. 3, pp. 227–52, 2001.

[2] A. Jongman, C. Va, C. B. Moore, and J. A. Sereno, "Perception and Production of Mandarin Chinese Tones," Handb. Chinese Psycholinguist., pp. 1–21, 2001.

[3] H. Luo, J.-T. Ni, Z.-H. Li, X.-O. Li, D.-R. Zhang, F.-G. Zeng, and L. Chen, "Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants.," Proc. Natl. Acad. Sci. U. S. A., vol. 103, no. 51, pp. 19558–63, 2006.

[4] J. Gandour, Y. Tong, D. Wong, T. Talavage, M. Dzemidzic, Y. Xu, X. Li, and M. Lowe, "Hemispheric roles in the perception of speech prosody.," Neuroimage, vol. 23, no. 1, pp. 344–57, 2004.

[5] W. S.-Y. Wang, "Language Change" Ann. N. Y. Acad. Sci., vol. 228, no. 1 Origins and E, pp. 50–72, 1976.

[6] Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am., vol. 120, no. 2, p. 1063, 2006.

[7] L. Zhang, J. Xi, G. Xu, H. Shu, X. Wang, and P. Li, "Cortical dynamics of acoustic and phonological processing in speech perception.," PLoS One, vol. 6, no. 6, p. e20963, 2011.

[8] J. Xi, L. Zhang, H. Shu, Y. Zhang, and P. Li, "Categorical perception of lexical tones in Chinese revealed by mismatch negativity.," Neuroscience, vol. 170, no. 1, pp. 223–31, 2010.

[9] G. Hickok and D. Poeppel, "The cortical organization of speech processing," Nat. Rev. Neurosci., vol. 8, no. May, pp. 393–402, 2007.

[10] K. Okada, F. Rong, J. Venezia, W. Matchin, I.-H. Hsieh, K. Saberi, J. T. Serences, and G. Hickok, "Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech.," Cereb. Cortex, vol. 20, no. 10, pp. 2486–95, 2010.

[11] N. Mesgarani, C. Cheung, K. Johnson, and E. F. Chang, "Phonetic feature encoding in human superior temporal gyrus.," Science, vol. 343, no. 6174, pp. 1006–10, 2014.

[12] D. D. Greenwood, " Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. vol. 33, pp. 1344-1356, 1961.

[13] D. H. Brainard, "The psychophysics toolbox," Spatial Vision, vol.10, pp. 433-436, 1997.

[14] B. Fischl, "FreeSurfer," Neuroimage, vol. 62, no. 2, pp. 774-781, 2012.

[15] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information.," Med. Image Anal., vol. 1, no. 1, pp. 35–51, 1996.

[16] T. Gasser, P. Bächer, and J. Möcks, "Transformations towards the normal distribution of broad band spectral parameters of the EEG," Electroencephalogr. Clin. Neurophysiol., vol. 53, no. 1, pp. 119–124, 1982.

[17] S. Ray and J. H. R. Maunsell, "Different origins of gamma rhythm and high-gamma activity in macaque visual cortex.," PLoS Biol., vol. 9, no. 4, p. e1000610, 2011.