Brain and Language 183 (2018) 21–31

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Unattended processing of hierarchical pitch variations in spoken sentences

Xiaoqing Li[a,*], Yiya Chen[b,*]

[a] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[b] Leiden University Center for Linguistics (LUCL) & Leiden Institute for Brain and Cognition (LIBC), Leiden, The Netherlands

ARTICLE INFO

Keywords:
Unattended processing
Hierarchical processing
Neural processing
Lexical tone
Sentence-level pitch
MMN

ABSTRACT

An auditory oddball paradigm was employed to examine the unattended processing of pitch variation which functions to signal hierarchically different levels of meaning contrasts. Four oddball conditions were constructed by varying the pitch contour of critical words embedded in a Mandarin Chinese sentence. Two conditions included lexical-level word meaning contrasts (i.e. TONE condition) and the other two sentence-level information-status contrasts (i.e. ACCENTUATION condition). Both included stimuli with early vs. late acoustic cue divergence points. Results showed that the two early-cue conditions elicited earlier Mismatch Negativities, regardless of their functional hierarchy. The deviant stimuli induced theta-band power increases in the TONE condition but beta-band power decreases in the ACCENTUATIION condition, regardless of the timing of their acoustic cues. These results suggest that, in an unattentive state, the human brain can functionally disentangle hierarchically different levels of pitch variation, and the brain responses to these pitch variations are time-locked to the presence of the acoustic cues.

1. Introduction

To comprehend speech, the human brain must integrate incoming information across multiple timescales. For example, one must not only identify the speech sounds of a single word but also at the same time integrate a sequence of words to derive sentence-level representation and meaning. What remains a puzzle is how and when these hierarchical levels of information are processed to arrive at comprehension efficiently. There has been accumulating evidence that the anatomical structure of the human brain recapitulates the temporal hierarchy that is inherent in the dynamics of sound or speech (Chaudhuri, Knoblauch, Gariel, Kennedy, & Wang, 2015; Kiebel, Daunizeau, & Friston, 2008; Lerner, Honey, Silbert, & Hasson, 2011). Speech processing studies suggest further that neural oscillatory activities not only can dynamically track the acoustic features at syllabic rate (Luo and Poeppel, 2007; Kubanek et al., 2013; Golumbic et al., 2013; Nourski, Reale, Oya, Kawasaki, and Kovach, 2009; Peelle, Gross, and Davis, 2012; Pasley et al., 2012) but also can be synchronized to internally constructed hierarchical linguistic structures (Ding, Melloni, Zhang, Tian, & Poeppel, 2016).

Existing studies on the processing of hierarchical structures of speech have mainly focused on attended speech. Attention state is indeed an important modulating factor on speech processing. An increasing number of studies have demonstrated that brain activities preferentially track attended speech streams, relative to unattended ones, with neural responses to unattended speech fluctuations being reduced or even disappearing. This attention-modulation effect on speech processing has been observed at both the high-level language brain areas and the low-level acoustic brain areas (Ding & Simon, 2012a, b; Kerlin, Shahin, & Miller, 2010; Mesgarani & Chang, 2012). What remains unclear is whether and how hierarchically different speech information is processed by the brain in an unattentive state.

Detecting changes in the functional meaning of unattended speech can be crucial for everyday life and sometimes even crucial for the safety of the listener. This is because in everyday situations, important information conveyed in the speech signal often occurs outside the center of listeners' attention. Furthermore, besides syllable-related general acoustic features which have been the main focus of existing research on hierarchical speech processing, supra-segmental feature is also a very important source of speech signal, which signals a variety of linguistic and paralinguistic functions. Such signals unfold over different hierarchical time scales. Until now, it is still unknown whether and how the human brain can distinguish the hierarchical organization of supra-segmental features in an unattentive state. This study focused mainly on one aspect of supra-segmental features, namely pitch variations. Our specific interest is in how the human brain processes word- and sentence- level pitch variations in unattended speech.

Pitch in speech can be modulated at different time scales to signal

* Corresponding authors at: CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, No. 16 Lincui Road, Chaoyang District, Beijing 100101, China (X. Li). Leiden University Centre for Linguistics, Leiden University, Postbus 9515, Leiden, 2300 RA, The Netherlands (Y. Chen).
E-mail addresses: lixq@psych.ac.cn (X. Li), yiya.chen@hum.leidenuniv.nl (Y. Chen).

https://doi.org/10.1016/j.bandl.2018.05.004
Received 19 September 2017; Received in revised form 11 April 2018; Accepted 8 May 2018
0093-934X/

meanings at different levels of linguistic organization. In about 60–70% of the languages of the world, pitch can vary at a syllable rate, with variations at this time scale distinguishing lexical-semantic meanings (Van Lancker and Fromkin, 1973). They are known as tone languages. For example, in Standard Chinese, a prototypical lexical tone language, the same syllable *ma* can mean 'mother', 'hemp', 'horse', and 'to scold' when it is produced with different pitch contours, known as different lexical tones (high level tone (T1), high rising tone (T2), low dipping tone (T3), and high falling tone (T4)).

In all languages of the world, pitch can fluctuate at the time scale of sentence to indicate sentence-level pragmatic-semantic meanings, such as signaling new vs. given (or focused vs. background) information. For example, in languages like English, the relative prominence of a particular part in a sentence can be signaled via pitch changes, typically referred to as accentuation, to highlight, e.g., the importance of the information (Ladd & Jun, 2008). In Standard Chinese, the new or focused information in a sentence context can also be highlighted with pitch variation (hereafter also referred to as accentuation), which, different from languages like English, is mainly realized as the distinctive realization of the lexical tonal contours within an expanded pitch range (Chen, 2006; Chen & Gussenhoven, 2008). Standard Chinese therefore provides an interesting test ground to explore whether and how the same supra-segmental feature (namely, pitch) can be processed differently by the human brain when it fluctuates at different time scales, namely sentence-level accentuation-induced pitch variation versus syllable-level lexical tone-induced pitch variation.

In the field of psycholinguistics, considerable studies have shown that during attended speech comprehension, the human brain can process multiple timescales of pitch variations for different meanings. At the lexical level, it has been shown that when participants are asked to perform semantic processing, the time course and amplitude of the N400 effect are comparable for tonal violation and consonant/vowel violation, suggesting that lexical tone is processed, just like segments, as phonemic information (Brown-Schmidt & Canseco-Gonzalez, 2004; Schirmer, Tang, Penney, Gunter, & Chen, 2005). For pitch variation at the sentence level, there has also been studies, both in tonal and non-tonal languages, which show that accentuation is immediately processed to comprehend the incoming utterance, as indicated by an enhanced N400 or P300 evoked in the unmatched condition (Hruska et al, 2001; Johnson, Breen, Clifton, & Morris, 2003; Magne et al, 2005; Li and Yang, 2013a, 2013b; Li, Yang, & Hagoort, 2008; Li, Chen, & Yang, 2011). These results suggest that listeners can immediately detect the mismatch between the information state of a word (e.g., the focus of the sentence) signaled by the presence/absence of an accentuation and that indicated by the preceding discourse/sentence context.

How are lexical tone and accentuation processed in an unattentive state? Variants of passive auditory oddball paradigm have been used to tap into unattended speech processing. In this paradigm, MMNs (mismatch negativities) are usually elicited by low-probability deviant stimuli embedded in high-probability standard stimuli, which reflect the brain's automatic responses to the deviant auditory stimuli (Näätänen, Gaillard, and Mäntysalo, 1978; Näätänen & Alho, 1997; Näätänen, 2001; Cowan, Nugent, Elliott, Ponomarev, and Saults (1999)). Studies employing this paradigm have examined the between- and within-category processing of lexical tones and found that the left hemisphere is more sensitive to the between-category processing of lexical tones while the right hemisphere is more sensitive to the low-level acoustic processing of pitch variation within the same lexical tone (e.g., Luo et al., 2006; Ren, Yang, & Li, 2009; Wang, Wang, & Chen, 2013; Xi, Zhang, Shu, Zhang, & Li, 2010). Furthermore, tonal pairs with an early acoustic divergence detection point has been found to elicit earlier brain responses (indicated by earlier MMN peak latency) than tonal pairs with a relatively later acoustic divergence detection point (Chandrasekaran, Krishnan, & Gandour, 2007; Li & Chen, 2015). Conjointly, these MMN studies suggest that, in an unattentive state, the human brain can distinguish lexical tonal categories indicted mainly by the pitch signal, and the speed of recognition is determined by the timing of the corresponding acoustic divergence points.

As mentioned earlier, pitch contours vary hierarchically at multiple timescales, which can be used to convey lexical semantics or sentence-level pragmatic-semantic meanings. Although there is a sufficient amount of experimental evidence for unattended categorical or acoustic processing of lexical tone pitch contours, no existing MMN studies have tapped directly into the question of whether and how in an unattentive state, the human brain is able to disentangle hierarchically different levels of pitch variation (i.e. lexical-level tone vs. sentence-level accentuation). There are two possible outcomes based on the existing literature. One possibility is that unattended processing of pitch variations occurs only at the lower acoustic level, and therefore, we should only observe processing of pitch variations as different acoustic categories but not as functionally different linguistic categories that vary at hierarchical timescales, such as pitch variation due to lexical tones vs. that due to accentuation. An alternative possibility, however, is that the human brain undergoes a relatively higher functional processing of pitch variations in the acoustic stream even outside of focused attention, and therefore, neural activities differ for the processing of pitch variations that function at different timescales (namely, syllable-level lexical tone vs. sentence-level accentuation). In other words, even out of focused attention, the human brain can distinguish hierarchically different pitch variations as different linguistic/functional categories (e.g., different phonological categories that have different linguistic functions).

In line with the first possibility, an early MMN study found that MMNs elicited by deviant pitch variations consistently showed the same hemisphere dominance (namely, the right hemisphere dominance as revealed by the source localization of the MMNs using a source estimation technique 'LORETA'), regardless of their linguistic functions (tone vs. intonation) or speech context (speech vs. non-speech) (Ren et al., 2009), which suggests that unattended pitch variations are mainly processed at the lower acoustic level. The second possibility is supported by studies which showed that, relative to within-category lexical-tone deviants, the across-category lexical-tone deviants elicited larger MMN in the left-hemisphere electrodes, although they involved exactly the same degree of acoustic differences, which reflects long-term phonemic traces of lexical tones (Xi et al., 2010). The inconsistent results revealed by the two studies are likely to be related to the different experimental designs and analysis techniques. Hemisphere lateralization/dominance was examined by using the source estimation technique 'LORETA' in Ren and colleagues' study, but by analyzing the surface-electrode topography in Xi and colleagues' study. Moreover, although Ren et al. (2009) investigated the potential differences between lexical-level (tone, T3-declarative/T4-declarative for standard/deviant stimuli) and sentence-level (intonation, T4-declarative/T4-interrogative) pitch variations, they didn't' strictly control the acoustic differences between these two conditions. As for Xi et al. (2010), although the acoustic differences between the within- and across-category lexical-tone conditions were strictly controlled, they didn't examine the processing of pitch variations at different timescales. Therefore, it's still unclear whether and how the human brain can distinguish hierarchically different pitch variations as different linguistic/functional categories.

Another important dimension of pitch processing is the timing characteristics that underlie the processing of different timescales of pitch variations. Li et al. (2008) have tapped into this issue by comparing the time course of tone and accentuation processing during attended speech comprehension. Their results showed that lexical tonal information is identified about 90 ms earlier than sentence-level accentuation, reflected in the latency of the N400 effects elicited by tone and accentuation violations. However, no study thus far has tapped into the time course of unattended processing of hierarchically different pitch variations.

We will again entertain two possibilities. The first is that unattended