Temporal Relationship of Top-down and Bottom-up Processing in Tone Perception

Lan Shuai 1, Tao Gong 2

1 Department of Chinese, Translation and Linguistics, City University of Hong Kong, Hong Kong
2 Department of Linguistics, University of Hong Kong, Hong Kong
Susan.Shuai@gmail.com, Gtojty@gmail.com

Abstract
Both the top-down and bottom-up processing are identified in language comprehension: the top-down processing depends on previous language experience and the bottom-up processing depends on instant language inputs. In speech perception, both types of processing are needed while perceiving pitch variations as lexical tones. In this paper, based on three ERP experiments, we propose a temporal model unifying both types of processing in a parallel way. These two types of processing, existing in both the early and late time windows, modulate the degrees of the general left lateralization of different ERP components (N1, P2, and N400).

Index Terms: tone perception, lateralization, ERP, top-down processing, bottom-up processing

1. Introduction
In speech perception, both the top-down and bottom-up processing are crucial for perceiving speech signals. The top-down processing refers to the interpretation of input signals as meaningful speech sounds based on previous experience, whereas the bottom-up processing refers to the acoustic or phonetic processing of speech signals. A comprehensive model of speech perception should consider both types of processing. However, most available models (e.g. the TRACE model [1]) are based solely on alphabetic languages, leaving out tones.

In the phonological system of tone languages, tone categories convey meanings together with speech segments such as consonants and vowels. The function of tones is the same as those of speech segments: tones form minimal tone pairs to differentiate meanings, similar to those minimal segmental pairs. However, tones differ from speech segments in acoustic properties. Tone is a suprasegmental feature with pitch variation superimposed on the whole syllable. Pitch changes of tones are slow, around several hundred milliseconds, whereas formant transitions of stop-consonants are fast, around several dozen milliseconds.

Only a few studies have worked on the mechanisms of tone processing in comparison to speech segments. A tone monitoring experiment [2] shows that tonemes are processed at a separate level than speech segments. In some ERP [3] and fMRI [4] experiments, tone processing is claimed to have different neural mechanisms from speech segments. However, whether such difference is due to the bottom-up or top-down processing remains unclear.

Hemispheric lateralization patterns of different speech components provide a good indicator of the bottom-up and top-down processing. The recent ERP study [7] reveals a right hemispheric dominance in tone perception, in opposite to stop-consonant perception. Some PET [5] and fMRI [6] experiments that compare the acoustic processing of fast or slow changing signals also show that the left hemisphere has a greater activity to sounds with fast changing acoustic properties than the right hemisphere, whereas the right hemisphere has a greater activity to sounds with slow changing acoustic properties than the left hemisphere. These results reflect a typical bottom-up processing of acoustic signals.

Based upon these results, the authors of [7] suggest that right hemisphere takes a major role in perceiving tones when doing auditory processing in an early time window, whereas left hemisphere shows greater activation when doing language related tasks in a late time window. They cite some PET [8] and fMRI [9] experiments to support the latter claim. By comparing hemodynamic changes during tone perception in tone and non-tone language speakers, those experiments show that when perceiving the same tonal syllables, there are more left hemisphere activations in tone language speakers than non-tone language speakers. Another IMRI study [10] that compares the native tone language speakers with non-native tone language speakers further shows that there are more left inferior frontal gyrus activations in native speakers than non-native speakers. All these results reflect not only the influence of language experience but also a strong top-down mechanism in perceiving tones.

All evidence points out a separation of the bottom-up and top-down processing, indicated by the opposite lateralization patterns of tones: Right hemisphere has an advantage in the bottom-up processing of pitch variation in tones, whereas left hemisphere has an advantage in the top-down processing of lexical semantics in tones. In other words, tone perception is a unified process. Following this perspective, the authors of [7] propose a serial model separating tone processing into two stages: an early auditory processing stage before and around 200 ms, in which there is a right hemisphere dominance of tone processing; and a late processing stage after 200 ms, in which the left hemisphere advantage starts to emerge. However, their work only manipulates acoustic properties of the stimuli and adopts the MMN paradigm to examine the ERP component around 200 ms (early stage), without considering the semantics of tones. In addition, the neuroimaging results cited by the authors to support the top-down tone processing [8, 9] do not reveal the time course of the top-down processing.

In the next section, we will further discuss the serial and parallel perspectives on the bottom-up and top-down processing during language perception, and then, propose our parallel model of tone perception.

2. Serial and Parallel Models
There has been a heated debate on whether language processing is a serial or parallel process. This question is intertwined with the issue of the bottom-up and top-down
processing. On the one hand, if language processing (or speech perception) is a pure bottom-up process, there should be a series of processing stages. In other words, language processing follows a serial model. On the other hand, if the top-down processing and bottom-up processing interact simultaneously, multiple stages may coexist within the same time. In other words, language processing follows a parallel model. The temporal relationship between the top-down and bottom-up processing could provide direct evidence to support either the serial or the parallel model.

2.1. Serial Model

In [11], the author proposes an auditory sentence processing model that favors the serial processing. According to this general model, the acoustic analysis is the first stage of language processing, which is before 100 ms. Then, the identification of phonemes occurs, which is around 100 ms. Since the syntactic processing is an automatic process happening before the semantic processing, the identification of word form and word category comes one by one immediately after the phoneme identification between 100 and 200 ms. The identification of lemma and morphological information is the next stage, which is between 200 and 300 ms. After that, the integration of semantic and morphosyntactic information happens around 300 to 500 ms. In the end, the reanalysis and repair process becomes the last stage, which is around 600 ms. These processes form a four stage cognitive processing model: phonological segmentation and sequencing, syntactic structure building, semantic relation forming or thematic role assignment, and syntactic integration. The author also lists some supportive ERP evidence for this model.

If such a serial model is correct, the manipulation of semantics could not affect the ERP components that occur early after the stimuli onset. However, recent ERP experiments [12-16] have provided the evidence of semantic modulation of the early ERP components. In addition, apart from assuming that syntactic processing is an automatic process in favor of the bottom-up processing to explain the early effect of syntactic manipulation, an alternative view that the top-down syntactic processing also takes place in the early time window makes sense as well. Whether the semantic or syntactic top-down processing takes place in the early time window is dependent on the nature of the task that subjects are doing. If the task is language irrelevant, it would be difficult to observe the online top-down processing, and the automatic bottom-up processing will dominate all processing stages. In such a case, the automatic processing could still reflect the top-down processing formed by long-term language experience. However, if the task is language relevant, the top-down processing, based on long-term memory or language experience, could take place, even in the early time window.

2.2. Parallel Model

Several ERP studies, based on language tasks, have revealed an early semantic processing that reflects the top-down processing around 100 to 200 ms after the stimulus onset [12-15]. In [16], also based on the MMN paradigm, the real words and pseudo words are shown to elicit significantly different ERP amplitudes at the preattentive stage. All these results indicate that the top-down semantic processing also occur in the early time window, and it can form an automatic processing observable by task-irrelevant paradigms such as MMN.

Together with other evidence, it is suggested that both the bottom-up and top-down processing exist in the early stage of language processing. However, it is still unclear whether both types of processing take place in the late stage on or after 300 ms and whether the top-down processing suppresses the bottom-up processing especially in the late stage.

A recent ERP study examines both the top-down and bottom-up processing by comparing the real word and pseudo-word comprehension with and without context [17]. The results support an early top-down processing starting around 200 ms after the stimulus onset, and a bottom-up processing existing in both the early (around 100 ms) and late (around 400 ms) time windows. This evidence supports a parallel way in which both types of processing are unified. In addition to this single evidence, more ERP or MEG experiments are called for to discern both types of processing, by examining the modulation effect at all levels including acoustics and phonetics, phonology, semantics, and syntax and clarifying whether both types of processing can be observed in both the early (100 to 200 ms) and late (300 to 500 ms) time windows.

2.3. Parallel Model of Tone Perception

Tone perception contains both the top-down and bottom-up processing, because this process is to comprehend lexical meanings by interpreting acoustic signals. As an acoustic signal, the pitch variation of tones elicits greater right hemisphere activation, mainly due to the auditory processing of tones. This reflects the bottom-up processing [7]. Meanwhile, in conveying lexical semantics, tones elicit greater left hemisphere activation, shown by comparing tone language speakers with non-tone language speakers, or native or non-native tone language speakers [8, 9]. This reflects the top-down processing.

According to the serial model, the bottom-up auditory processing of acoustic properties of tones should exist in the early time window, while the top-down interpretation of semantic meaning of tonal syllables should exist only in the late time window. Therefore, the lateralization pattern of tone perception should be a right hemisphere advantage in the early time window and a left hemisphere advantage in the late time window. However, if the parallel model is correct, the lateralization pattern should be bilateral in both the early and late time windows, but the degree of lateralization depends on the strength of the top-down or bottom-up processing, which can be modulated by the semantic or acoustic properties of the syllable.

Considering all these, we propose a parallel model of tone perception. Figure 1 shows the conceptual diagrams of the serial model (according to [7]) and our parallel model of the top-down and bottom-up processing of tone perception. In the following section, we will provide the results of three ERP experiments to support this parallel model.
3. Tone Perception and ERP Components

There are two ways widely adopted to explore the top-down and bottom-up processing during tone perception. One way is to observe the lateralization patterns of the ERP components and compare these patterns with previous results. The other way is to manipulate the experimental conditions and compare the relative lateralization patterns of the ERP components. In this paper, in order to explore the temporal relationship of the bottom-up and top-down processing, we adopt both ways and examine ERP components in the early and late time windows, based on three ERP experiments.

3.1. Auditory N1 Component

Auditory N1 component is the first positive going ERP component in response to the auditory input, and it has a fronto-central topographic distribution. The peak of the N1 component usually appears around 100 ms, and the amplitude of it changes according to the acoustic properties of sounds, and the states of participants such as alertness [18]. Lateralization of the N1 component has been reported to correlate with tasks as well as cues [19, 20]. MEG has the same source as ERP components. Lateralization of MEG is in accordance with lateralization of ERP [21]. The MEG N1 component is called N1m or M100. In language tasks, the N1m has a left lateralization pattern [22], whereas in music or pitch perception tasks, N1m has a right lateralization pattern [23].

Our first ERP experiment explores the lateralization of the N1 component based on a tone identification task. According to the general lateralization of language functions [20], the ERP pattern should be left lateralized. However, according to the cue dependent hypothesis [5], the lateralization in processing of sounds with relatively fast changing acoustic properties should be more left than that in processing of sounds with relatively slow changing acoustic properties.

The materials of this experiment include six Cantonese tones: three level tones and three contour tones. 20 native Cantonese speakers participate in this experiment. They are university students with the average age of 22. Six characters were shown on the screen at 1100 ms after the auditory stimulus onset (this is to avoid the interference with the ERP recording). Participants were asked to judge which character was the one they heard.

EEG data was collected by a 128-channel EEG system with Geodesic Sensor Net (EGI Inc., Eugene, OR, USA). The impedances of all electrodes were kept below 50 kΩ at the beginning of the recording. Subjects were instructed to blink naturally, and sit still and comfortably during the experiment. Eye blinks and movements were monitored through electrodes located above and below each eye and at the outer canthi. Signals were sampled at 1000 Hz with a 0.01-100 Hz band-pass filter during the recording. The original reference electrode was at the vertex and the ERPs were re-referenced to the averaged-reference during the data processing.

During the offline processing, the recorded continuous EEG data were filtered by a 40 Hz low-pass filter and segmented from -200 ms to 1000 ms by referring to the stimulus onset. All segments with an amplitude change exceeding 100 μV in the vertical eye channels and all electrodes, or with the voltage fluctuation exceeding 45 μV in the horizontal eye channels were excluded from further analyses. The baseline correction was done from -200 ms to 0 ms.

The overall correctness was 94.8%, with tone 5 having the lowest correctness at 89.9%. The auditory N1 component showed a peak around 118 ms, and it had a frontal central distribution. The C3 and C4 electrodes were recruited for analysis. A 2-way repeated measures ANOVA was conducted with lateralization (C3/C4) and rate of pitch changes (level/contour tones) as two factors. The N1 component showed a general left lateralization pattern (F (1, 19) = 9.261, p < 0.01), and the level/contour tone factor interacted significantly with the lateralization factor (F (1, 19) = 4.457, p < 0.05). The post-hoc analysis on the average N1 amplitude showed that the contour tones elicited a greater left lateralized N1 component than the level tones (p < 0.05), and both types of tone elicited a significant left lateralized N1 component (level tones: p < 0.025; contour tones: p < 0.01). Figure 2 shows the average N1 amplitudes elicited by both level and contour tones.

![Figure 2: The average N1 amplitudes in the left and right homologue electrodes in our first experiment.](image)

3.2. Auditory P2 Component

Auditory P2 component is the second positive going ERP component. It has a central distribution and its peak amplitude is usually around 200 ms [24]. Lateralization of the P2 component is reported to be related to both acoustic properties and tasks [25]. The correspondent MEG component is the P2m or M200. Previous research reports a general lateralization of the P2m in doing language related tasks [26].
and the acoustic property of the sound also affects the lateralization of the P2m [27].

Our second ERP experiment explores the lateralization of the P2 component based on two tasks: the dichotic listening task and the lexical decision task with phonological priming in Mandarin.

In both tasks, we design two factors, semantics (meaningful vs. non-meaningful words) and acoustic property (stop-consonant vs. tone conditions) factors. The semantic factor is expected to elicit greater left lateralization of the P2 components, which is in line with the general lateralization pattern of the language task. However, the acoustic property factor is expected to show a greater left lateralization pattern in the stop-consonant condition with fast changing acoustic property, and a greater right lateralization (or less left lateralization) pattern in the tone condition with slow changing acoustic property.

3.2.1. Dichotic Listening Task

In this task, participants hear two different syllables simultaneously and respectively in their left and right ears. We adopt a two-by-two design. The first factor, semantics, has two levels, meaningful and non-meaningful words. In the condition of meaningful words, the two different syllables simultaneously presented are meaningful Mandarin monosyllabic words; in the condition of non-meaningful words, both syllables are built up by non-existing combinations of valid Mandarin phonemes. The second factor, acoustics, also has two levels, stop-consonant and tone. In the condition of stop-consonants, the two syllables in the two ears differ only in initial stop-consonant; in the condition of tone conditions, they differ only in lexical tones.

The EEG recording and offline ERP processing procedures were the same as the above mentioned tone identification experiment, expect for the sampling rate (250 Hz), segmentation parameters (-100 ms to 900 ms), and baseline correction range (-100 ms to 0 ms). For the rest of the ERP experiments, all parameters were kept the same as the current one.

The task of participants is to remember the two syllables played simultaneously in their two ears. After hearing the two syllables, participants need to respond, according to the Chinese character “左/右” (left/right) shown on the screen, both the consonant and vowel of the corresponding side of the auditory input by pressing the responding pad.

Thirty-two healthy university graduate students participate in the experiment. They are all native Mandarin speakers with the average age of 27 (±4.2).

A 3-way repeated-measures ANOVA was conducted to analyze the behavioral results with semantics, acoustic property and ear advantages as three factors. A significant 3-way interaction was found ($F(1, 31) = 13.549, p < .001$), with significant interactions between the semantic and lateralization factors ($F(1, 31) = 5.636, p < .05$), and between the acoustic property and lateralization factors ($F(1, 31) = 7.199, p < .012$).

The post-hoc analysis on the average amplitudes between 152 and 248 ms of the P2 component in the C3 and C4 electrodes showed a greater left lateralization in the conditions of meaningful words, compared with the conditions of non-meaningful words ($p < .05$). As for the acoustic property factor, the P2 component of the conditions of stop-consonants showed a greater left lateralization pattern than that of the conditions of tones ($p < .025$). Figure 3 shows the average P2 amplitudes in different conditions.

3.2.2. Lexical Decision and Phonological Priming Task

In this task, the prime syllable either has the same consonant as the first syllable of the target disyllabic word to be judged, or shares the tone with the first syllable of the target word. In the control condition, there are no common phonemes or onomemes between the prime and target syllables. The task is to judge whether the target syllabic word is a real word or not. Participants are asked to ignore the prime syllable and just focus on the target syllable. Thirty-one healthy university graduates participate in this task. They are native Mandarin speakers whose average age is 25 (±2.3).

A significant priming effect of the real stop-consonants was observed in the behavioral data. The ERP data showed similar results as the dichotic listening task, with respect to the P2 component. The P2 averaged amplitude of the priming effect revealed a significant interaction between the semantics and lateralization factors ($F(1, 30) = 5.636, p < .024$), and between the consonant/tone priming factor and lateralization factors ($F(1, 30) = 4.810, p < .036$). This result was in accordance with that in the dichotic listening task: both semantics and acoustic properties affect lateralization.

3.3. N400 Component

Auditory N400 component appears in the late time window around 300 to 500 ms with a negative going potential, when the integration of the target to the context is unexpected. The semantic violation experiment is usually used to elicit the N400 component [28]. The phonological priming experiment also shows the N400 component, when comparing the control condition with the priming condition [29, 30]. The auditory N400 component has a more frontal distribution, compared with the visual N400 component [31]. The source of the N400 is believed in the temporal and frontal brain areas [32, 33].
The lateralization of the N400 component should also be left lateralized in language tasks [34].

Our third ERP experiment adopted both the priming and semantic violation in sentences tasks. The priming task was already illustrated in the previous section. A further analysis of the N400 component showed that the semantics factor interacted significantly with the lateralization factor (F(1, 30) = 10.440, p < .003). Meaningful words had greater left lateralized N400, compared with non-meaningful words. Such lateralization pattern was also observed in the previous experiment.

In order to examine whether the acoustic property also affects the lateralization of the N400 component, we also conduct the semantic violation in sentences task, in which participants need to judge whether the last syllable in a sentence is consistent with the previous context or not. There are two semantic violation conditions. In one condition, the consonant of the last syllable of the sentence is changed to form an irrelevant syllable; in the other condition, the tone of the last syllable is changed to form an irrelevant syllable. In the control condition, the last syllable is consistent with the previous context.

Twenty-nine healthy university graduate students participate in the experiment. They are native Mandarin speakers, whose average age is 26 (±2.8). The correctness is 94.3%.

A 2-way repeated measures ANOVA was conducted with violation types (consonant/tone) and lateralization (left/right) as two factors. The N400 component showed a general left lateralization pattern. There was also a significant interaction between the lateralization and consonant/tone violation induced N400 component (F(1, 28) = 5.990, p < .021). The post-hoc analysis revealed that the consonant violation induced N400 had a significant left lateralization (p < .036), whereas the tone violation induced N400 did not, though the component had a greater amplitude in the left electrodes. Figure 4 shows the average N400 amplitudes in these conditions.

4. Discussions and Conclusions

The above explorations of the auditory ERP components in both the early and late time windows reveal that a general left lateralization, due to the top-down processing, exists across different time windows of processing. Meanwhile, the bottom-up processing also exists in both the early and late time windows, as shown by the modulation on the lateralization by acoustic properties of input signals. The general left lateralization is in accordance with the modulation on the relative lateralization pattern by controlling whether the input stimuli are meaningful or not.

Although the lateralization and its interactions with the top-down (e.g. semantics) and bottom-up (e.g. acoustic properties) factors were explored in different experiments and using different tasks, we obtained consistent lateralization patterns throughout the early N1/P2 and late N400 ERP components. There are three general lateralization patterns: 1) these ERP components are left lateralized; 2) the linguistic manipulation of meaningful or meaningless words modulates the lateralization pattern in a way that the meaningful words generate a greater left lateralization pattern; and 3) the manipulation on the physical properties of the auditory input modulates the lateralization pattern in a way that the faster changing cues generate a greater left lateralization pattern. The second and third effects are regarded to reflect respectively the top-down effect from linguistic experience and the bottom-up effect from the acoustic/phonetic processing. The first effect was also ascribed to the top-down effect, because the left lateralization is generally consistent with the relative lateralization changes modulated by the top-down effect. Moreover, the left lateralization of speech processing resulting from the linguistic experience was confirmed by the previous experiments [8, 9], and such linguistic experience is also a top-down effect in speech processing.

All these experimental results support the parallel model incorporating both the top-down and bottom-up processing. This model emphasizes that the top-down processing is a pre-determined process, making the relevant hemisphere/brain regions get prepared for the task. When the stimuli come in, they will evoke the responses from the correspondent hemisphere/brain regions, and these responses may adjust or even alter the relative degrees of lateralization.

Meanwhile, in [7], the authors use the MMN paradigm to support the serial model. However, they only manipulate the acoustic properties of the stimuli, without considering the semantic properties of the stimuli. Therefore, they can only identify the bottom-up processing in the MMN component, related only to acoustic properties of the stimuli. The experimental paradigm such as MMN can detect automatic preattentive processing. Such processing is due to either the acoustic properties, or the long-term language experience, the latter of which reliably is shown in [12-16]. Therefore, the MMN component can not only reflect the bottom-up processing but also the top-down processing relevant for semantics and syntax.

In conclusion, our three ERP experiments of tone perception collectively support that both the bottom-up and top-down processing are coexistent, taking place simultaneously in both the early and late time windows. In other words, tone perception is a parallel process incorporating both the bottom-up and top-down processing.

5. Acknowledgements

This work was supported in part by the SHIAE funding in the Chinese University of Hong Kong and the Society of Scholars in the Humanities in the University of Hong Kong. We thank colleagues from the Language Engineering Laboratory in the Chinese University of Hong Kong for their support.

6. References


