ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 104 (2018) 817 – 820
DOI 10.3813/AAA.919231

Temporal Order Judgment Reveals Local-Global Auditory Processes

Xiangbin Teng[1)], Yue Sun[1)], David Poeppel[1,2)]
[1)] Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt, Germany. xiangbin.teng@gmail.com
[2)] Department of Psychology, New York University, New York, NY, USA 10003

Summary
Speech signals can be considered as acoustic sequences composed of local units (e.g. phonemes) which form global acoustic patterns (e.g. syllables). Extraction of speech information at both local and global scales is essential for comprehension. To decipher this process, we employed the temporal order judgement (TOJ) paradigm and investigated how the auditory system processes acoustic sequences. We selected four vowel segments of 30 ms and generated short acoustic sequences. We then examined listeners' performance on TOJ of the vowel sequences using a same-different paradigm. The data showed that acoustic changes on a local scale caused by reversing vowel segments modify TOJ performance. Furthermore, the effect of local changes was attenuated when inter-onset interval between vowel segments increases, where segments can be recognised individually. A follow-up experiment showed that recognition of each segment was modulated by segment position and indicated that positions of acoustic segments contribute differently to TOJ. The results suggest that listeners perform TOJ by perceiving global patterns of acoustic sequences, which are further modulated by acoustic details on a local scale. Acoustic information on the local and global scales determines concurrently identification of short acoustic sequences.

PACS no. 43.66.-x, 43.66.Mk, 43.70.Fq, 43.71.-k, 43.71.Rt

1. Introduction

Speech comprehension requires listeners not only to process elementary acoustic segments, such as phonemes and syllables, but also to encode their temporal order [5]. For example, 'cat' and 'act' are two different words composed of the same phonemes and only differ by the temporal order of phonemes. Therefore, extracting temporal order of acoustic elements is a fundamental process in speech perception [19].

Previous research used sequences composed of artificial acoustic stimuli, such as pairs of hisses, buzz, tones, and clicks, and asked listeners to judge temporal order between different components. The minimum onset interval between sounds necessary for temporal order judgement (TOJ) was found to be less than 30 ms [1, 5, 6, 7, 8]. However, it was argued that listeners could perceive global changes caused by different temporal orders of the two components to perform TOJ [18]. Accordingly, other studies used sequences with more than two acoustic components and repetitively presented those sequences to prevent listeners from perceiving global patterns [2, 19]. An onset interval of longer than 100 ms was often found.

Received 7 March 2018,
accepted 8 September 2018.

The two time constants found potentially reflect two distinct processes involved in TOJ – one process relies on perceiving changes of acoustic sequences on a global scale, which are induced by the changes of the temporal order of local elements; the other process relies on recognition of each local element and the temporal order is later determined cognitively. However, it is still unknown how these two processes, processing acoustic details at a local scale and perceiving global patterns, interact with each other and jointly determine TOJ.

Here, we investigate this question using a novel paradigm. Specifically, we used sequences composed of four short vowel segments, which had distinct acoustic structures and can be recognized individually. We then introduced changes of temporal order by reversing the order of vowel segments. Two experimental manipulations were made. First, we varied the positions of the segments that are reversed to examine whether local changes influences TOJ. Secondly, we manipulated the length of intervals between segments, which influence recognition of individual segments [4, 9, 15]. We hypothesize that if recognition of acoustic sequences involves processes that operates on both local and global scales, then both the position of the changes on a local scale and the size of intervals should modulate TOJ.

2. Experiment 1

We tested whether listeners can differentiate between the four short vowel segments presented in isolation, as stimuli types affect listeners' performance on TOJ [19].

2.1. Methods

2.1.1. Participants

Ten English native speakers (age 18 to 23 years; 7 female; one left-handed) gave written consent and participated in the experiment. None of the participants had hearing loss or neurological abnormalities according to participants' self-report. We conducted all experiments in accordance with procedures approved by the NYU committee on Activities Involving Human Subjects.

2.1.2. Stimuli and procedures

Four English vowels spoken by a female speaker (close front, close-mid back, near-open near-front, and open-mid near-front) were used as stimuli (The stimuli can be found at https://edmond.mpdl.mpg.de/imeji/collection/rZWJgrvQrz2AP8DL?q=). Amplitudes of all tokens were normalized individually to 60 dB SPL. A segment of 30 ms was chopped with a rectangular window from the middle of each vowel and was used in all experiments.

A match-to-sample paradigm was used to examine the discriminability between different vowel segments (Figure 1, top). On each trial, the participants were first presented with one of four vowel segments and 700 ms later with two vowel segments sequentially as match alternatives, one of which was the same as the sample. The two samples had inter-onset intervals uniformly distributed between 400 ms and 600 ms. The participants had to choose which one of the two vowel segments matched the sample by pressing two buttons. No feedbacks were provided as we would like to test how well listeners can differentiate the vowel segments without previous experiences. Forty trials were presented for each comparison between two vowel segments.

All stimuli were presented using MATLAB (The MathWorks, Natick, MA) at 16 bit, with a sampling rate of 44.1 kHz using headphones (Sennheiser HD 380 Professional, Sennheiser Electronic Corporation, Wedemark, Germany). The d-prime value corresponding to the 100 percent accuracy is 4.5 as a half incorrect trial was added.

2.2. Results and Discussion

The discriminability between vowel segments are shown in Figure 1(bottom). The data showed that d-prime values for all pairs of vowel segments are above 3.5 and close to 4.5. Listeners have no difficulty discriminating vowel segments from each other. This result confirms that the acoustic information within 30 ms for each vowel segment suffices for the auditory system to differentiate different segments [9]. Therefore, difficulties of listeners to identify temporal order of an acoustic sequence composed of such vowel segments cannot be due to the inability to recognize individual components.

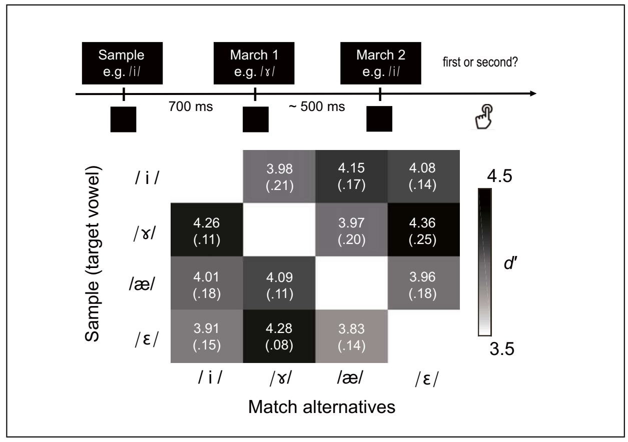

Figure 1. Top: the match-to-sample paradigm using in Experiment 1. Bottom: the results of d-prime value of the match-to-sample. The vertical axis labels the sample; the horizontal axis labels the vowel segment which is different from the sample in the match pair. The scale bar shows d-prime value. In each box of the confusion matrix, the numbers show the group-averaged d-prime value and one standard error over participants (in parentheses).

3. Experiment 2

In Experiment 2A, we examined how TOJ is modulated by the position of order reversal on the local scale and the size of interval between components. In Experiment 2B, we further tested how the position of vowel segments influence recognition of individual vowel segments.

3.1. Methods

3.1.1. Participants

Ten English native speakers (age 18–27 years; 7 females) in Experiment 2A and ten English native speakers (age 18–27 years; 7 females) in Experiment 2B gave written informed consent to participate in the study. All participants were right-handed with no known hearing deficits or neurological abnormalities.

3.1.2. Stimuli and procedures

In Experiment 2A, we created vowel sequences by concatenating randomly the four vowel segments with different temporal orders. The experimental procedure was a same-different task, where two vowel sequences were presented sequentially with an inter-sequence interval equally distributed from 400 to 600 ms. Participants were asked to determine whether the two vowel sequences were the same or different. Six types of temporal order reversing were defined as a function of the position of the two vowels that were involved (Figure 2,bottom): 1) reversing between the 1st and the 2nd segments; 2) between 2nd and 3rd; 3) between 3rd and 4th; 4) between 1st and 3rd; 5) between 2nd and 4th; 6) between 1st and 4th. The inter-onset intervals (IOI) between vowel segments were varied across six levels: 30, 50, 70, 110, 170 and 250 ms. Experiment 2A is of a 6 x 6 design with two within-participant factors of vowel reversing position and IOI. Each condition contained 20 trials, and a total of 720 trials were included in the final analysis. The trials were randomly divided into four blocks and presented in pseudorandom order.

In Experiment 2B, the experimental procedure was a one-interval two-alternative-force choice paradigm. A target vowel segment was first presented to the participants ten times before each block. On each trial, the participants were presented with one sequence of four vowel segments and had to determine whether the target vowel segment was presented in the first half (the first and second positions of the vowel sequence) or in the second half (the third and fourth positions). Target positions were binned into two categories in the analyses: Boundary (the first and fourth positions) and Middle (the second and third positions). IOI was also manipulated as in Experiment 1A. Experiment 2B was of a 2 x 6 design with two factors of target vowel position and IOI. The participants were tested in four blocks, with a specific vowel segment used as target for each block. Each block contained 360 trials (30 trials for each condition).

3.2. Results and Discussion

Results of Experiment 2A and 2B are shown in Figure 2 and 3, respectively. To measure the effects of reversing position and IOI in Experiment 2A, we conducted a two-way Reversing position x IOI repeated measures ANOVA (rmANOVA) on d-prime values (Figure 2). A significant main effect was found for Reversing position ($F(5,45) = 64.27$, $p < 0.001$) and for IOI ($F(5,45) = 88.84$, $p < 0.001$). The interaction effect between Reversing position and IOI was also significant ($F(9,225) = 8.83$, $p < 0.001$). We then measured at what IOI the effect of reversing position is significant by conducting a one-way rmANOVA at each IOI. We found a significant main effect of the reversing position at IOIs from 30 ms to 170 ms ($p < 0.01$, Bonferroni correction applied)). To measure the effects of segment position in Experiment 2B, we conducted a two-way Segment position x IOI rmANOVA on d-prime values. A significant main effect was found for Segment position ($F(1,9) = 39.25$, $p < 0.001$) and for IOI ($F(5,45) = 35.65$, $p < 0.001$). The interaction effect between Reversing position and IOI was also significant ($F(5,45) = 6.18$, $p < 0.001$). We then measured at what IOI the effect of Segment positi on is significant by conducting a one-way rmANOVA at each IOI with the segment position as the main factor. We found a significant main effect of the reversing position at IOIs from 30 ms to 110 ms ($p < 0.05$, Bonferroni correction applied)(Figure 3).

Our results from Experiment 2A confirm the previous finding that listeners perceive global patterns of acoustic sequences to identify temporal order [2, 19], as the participants' performance should not be modified by the reversing position if they adopt a strategy to first recognize each component and then identify their temporal order. As IOI increases and vowel segments are separated further apart, the effects of reversing position are attenuated. Experiment 2B showed that there is a position effect of recognition of each component (Figure 3, bottom). The vowel segments on the boundary positions can be better recognized, which explains the effect of reversing position in Experiment 2A.

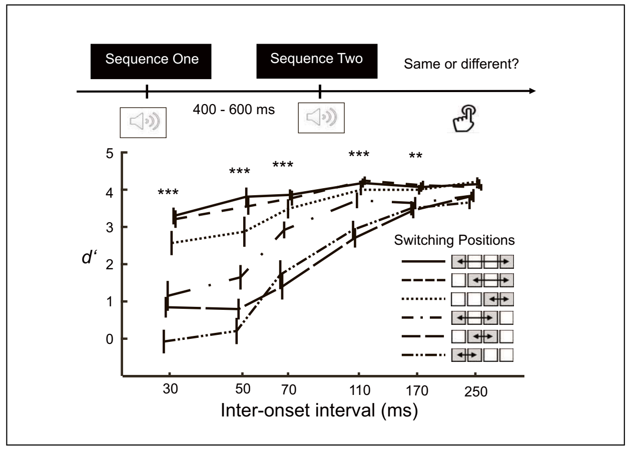

Figure 2. Top: the same-different paradigm in Experiment 2A. Bottom: results of Experiment 2A. The vertical axis represents d-prime value and the horizontal axis inter-onset interval. The line style codes for switching positions of vowel segments. The shaded boxes and the double arrows in the legend indicate that the positions of vowel segments in the first sequence is reversed in the second sequence.

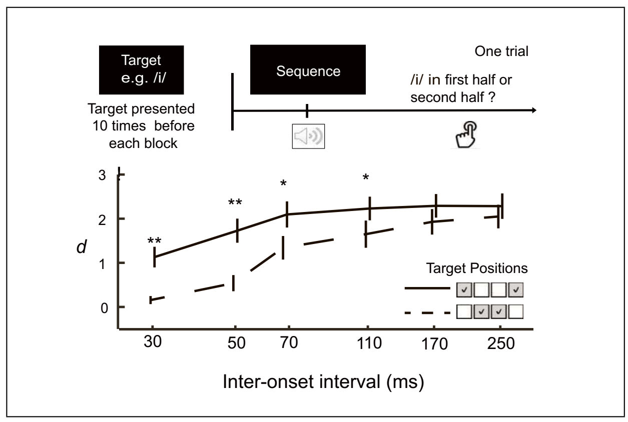

Figure 3. Top: the experimental paradigm in Experiment 2B. Bottom: d-prime value of Experiment 2B. The line style codes for segment position, boundary positions (solid) and central positions (dashed). The data show that vowel segments are easier to be recognized in the boundary positions than in the central positions.

4. General Discussion

We showed that TOJ involves auditory processes of both extracting local details and identification of global patterns. When IOI is short (e.g. <170 ms), TOJ relies on perception of global pattern of acoustic sequences, which are modulated by details of temporal order reversal on the local scale. When IOI increases over 170 ms and each acoustic component can be recognized, the effect of local acoustic changes disappears. Our study here, though seemingly simple, reveals complicated auditory processes in speech perception – acoustic information, local and global, needs to be extracted concurrently to form a holistic percept [14, 16].

The results of reversing position in Experiment 2A and segment position in Experiment 2B echo findings from studies on forward and backward masking [10, 11]. As the vowel segments in the middle of acoustic sequences are masked by both the preceding and following segments, the masking effects probably lead to worse recognition of these segments comparing to of those at the boundary positions. This finding suggests that acoustic information in different temporal positions within acoustic sequences contributes differently to forming the globally perceived pattern of acoustic sequences. The effects of local acoustic details are modulated by IOI. As the time constants found in studies on forward and backward masking are often of tens of milliseconds [12, 13], the effect of local acoustic changes should not occur for IOIs longer than 100 ms. However, in our study, this effect persists with an IOI as long as 170 ms. This result is in line with previous findings that more than hundreds of milliseconds are needed to recognize individual components in acoustic sequences.

The findings of the present study echo a recurrent theme on resolution and integration of the auditory system [3, 15]. The auditory system needs to integrate acoustic information over a long timescale to perceive global patterns while extracting acoustic information on a short timescale to decipher fast acoustic changes. Our results here lend a support to a proposal that concurrent local-global processes exist in the auditory system [14, 16].

Acknowledgement

This research was supported by NIH R01 2DC05660 to DP and the Max-Planck-Society.

References

[1] D. E. Broadbent, P. Ladefoged: Auditory perception of temporal order. J Acoust Soc Am **31** (2005) 1539–1539.

[2] W. L. Cullinan, E. Erdos, R. Schaefer, M. E. Tekieli: Perception of temporal order of vowels and consonant-vowel syllables. J Speech Lang Hear R **20** (1977) 742.

[3] E. De Boer: Auditory time constants: A paradox? Time resolution in auditory systems. Springer, 1985.

[4] O. Ghitza, S. Greenberg: On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica **66** (2009) 113–126.

[5] I. J. Hirsch: Auditory perception of temporal order. J Acoust Soc Am **31** (1959) 759–767.

[6] I. J. E. C. Hirsh, J. Sherrick: Perceived order in different sense modalities. J Exp Psychol **62** (1961) 423–432.

[7] M. Kanabus, E. Szelag, E. Rojek, E. Poppel: Temporal order judgement for auditory and visual stimuli **62** (2002) 263–270.

[8] J. A. S. Kinney: Discrimination in auditory and visual patterns. The American Journal of Psychology **74** (1961) 529.

[9] D. W. Massaro: Preperceptual images, processing time, and perceptual units in auditory perception. Psychol Rev **79** (1972) 124.

[10] B. C. J. Moore: Masking patterns for synthetic vowels in simultaneous and forward masking. J Acoust Soc Am **73** (1983) 906.

[11] B. C. J. Moore: Temporal integration and context effects in hearing. J Phonetics **31** (2003) 563–574.

[12] A. J. Oxenham: Mechanisms and mechanics of auditory masking. The Journal of Physiology **591** (2013) 2375–2375.

[13] A. J. Oxenham, B. C. J. Moore: Modeling the additivity of nonsimultaneous masking. Hear Res **80** (1994) 105–118.

[14] D. Poeppel: The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time. Speech Commun **41** (2003) 245–255.

[15] X. Teng, X. Tian, D. Poeppel: Testing multi-scale processing in the auditory system. Scientific Reports **6** (2016) 34390.

[16] X. Teng, X. Tian, J. Rowland, D. Poeppel: Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. PLoS Biol **15** (2017) 1–29.

[17] R. L. J. van Eijk, A. Kohlrausch, J. F. Juola, S. van de Par: Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. Percept Psychophys **70** (2008) 955–968.

[18] R. M. Warren, C. J. Obusek: Identification of temporal order within auditory sequences. Percept Psychophys **12** (1972) 86–90.

[19] R. M. Warren, C. J. Obusek, R. M. Farmer, R. P. Warren: Auditory sequence: Confusion of patterns other than speech or music. Science **164** (1969) 586–587.