TIMING PERCEPTION OF SEQUENTIAL COMPLEX TONES CONNECTED WITH GLIDES OF A FORMANT FREQUENCY

Satomi Tanaka[1], Minoru Tsuzaki[2], Eriko Aiba[1] and Hiroaki Kato[3]
[1]*Graduate School of Music, Kyoto City University of Arts, Kyoto, Japan*
[2]*Faculty of Music, Kyoto City University of Arts, Kyoto, Japan*
[3]*ATR, Cognitive Information Science Laboratories/NICT, Kyoto, Japan*
m06902@kcua.ac.jp, minoru.tsuzaki@kcua.ac.jp, m06901@kcua.ac.jp, kato@atr.jp

Abstract

Temporal relations between sound events can be perceived without critical difficulty although it is difficult to be defined acoustically. This study was designed to investigate what acoustical change functions as an effective cue to "mark" the changes in sounds, i.e., the arrival of a new event. A perceptual experiment was conducted to test whether any significant difference existed between the starting and ending points of frequency glides as the arrival of a new event. The glide was realized by changing the center frequency of a band-pass filter to simulate a formant change in speech, while a fundamental frequency was fixed to prevent harmonic components from changing frequencies. The results showed that the sensitivity was more accurate for the starting point than for the ending point, similar to the case of sinusoidal frequency glides.

The sensitivity for temporal deviations at the starting or ending points of change in a sound signal has been experimentally studied. The authors previously reported that in some conditions, sensitivity for temporal deviation was significantly higher at the starting point of frequency glide than at the ending point using sinusoidal tones (Tanaka et al., 2006a, b). This result suggests that for detecting temporal deviation the starting point of change in sound can be a more important cue than the ending point. However, no systematic difference was observed between them in conditions where the resonance frequency or formant frequency of a complex tone glided instead of the fundamental frequency (Tsuzaki et al., 2006).

These results suggest that the starting point of change became an important cue only when the pitch of the sound changes. However, since the performance level indicates that detection was much harder in the formant glide experiment than in the sinusoidal frequency glide experiment, perhaps some cues other than formant glide might lead to the opposite result. To investigate whether sensitivity for temporal deviation shows different result between sinusoidal and formant glides, an experiment replicated the same method as Tsuzaki et al. (2006) by modified stimuli parameters.

Procedure

Stimuli --- Standard stimuli

Our experiment detected temporal deviation from an isochronous structure realized by the intervals between the starting or ending points. Standard stimuli had isochronous structures, while the compared stimuli had temporal deviations.

In our experiments, the sound stimulus had three identical base parts and three identical upper parts that were alternated and connected with glides of formant frequency

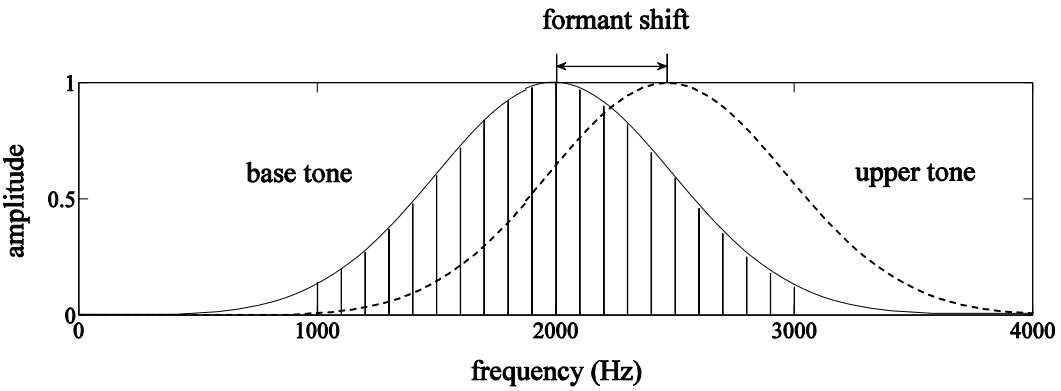

Fig. 1. Amplitude spectrum and envelopes of tones. Frequency separation between each component was 100 Hz corresponding to fundamental frequency of stimulus.

without change in spectral shape. Figure 1 shows an amplitude spectrum and envelopes of the base and upper parts. The fundamental frequency was fixed at 100 Hz throughout the whole stimulus sequence. To generate stimuli, a band-pass filter whose frequency response followed a Gaussian function was designed. The center frequency of the filter was 2000 Hz for the base part, and the half-power bandwidth was 1150 Hz. There were two conditions of upper parts whose center frequencies were 2241 or 2481 Hz. These frequencies were selected based on the biological constraint of the auditory system, i.e., the bandwidth of the frequency analysis processing widened as frequency increased. In this study, these band widths were calculated using equivalent rectangular bandwidth (ERB) described as

$$\mathrm{ERB} = 24.7\,(4.37F/1000+1) \qquad (1)$$

where F indicates the center frequency in Hz. The two frequency range conditions were 1 ERB (2241 Hz) and 2 ERB (2481 Hz). The separation between the center frequencies of the base and upper parts in the 1-ERB condition was almost identical as the bandwidth of the auditory filter at these center frequencies while that in the 2-ERB condition clearly exceeded it.

There were two temporal point conditions, starting and ending, as schematically shown in Fig. 2. The temporal intervals between either the starting or ending points of the formant glides were fixed at 400 ms, while those between each other point were randomized. By this mean, an isochronous structure was established exclusively between the starting points, or vice versa. Accordingly, the duration of the formant glide parts varied one by one ranging from 80 to 240 ms, since the real start point of the whole sequence as well as the ending point could add unnecessary cues for timing. To minimize such effects, a long leading part of 2000 ms and a trailing part of 1000 ms flanked the core part of the stimulus. Both the leading and trailing parts were tapered linearly 500 ms, which was assumed to be gradual enough to avoid unnecessary timing cues.

Stimuli --- Comparison stimuli

The spectral structures of the comparison stimuli were identical to the standard stimuli, while the temporal points fixed at the standard stimuli deviated. Figure 3 shows a schematic figure as an example. In a comparison stimulus, one interval fixed at 400 ms in a standard stimulus was lengthened by T, another was shortened by T, and the other two remained as they had been. Therefore, the entire duration was identical as the standard stimulus to avoid the differences of the entire duration from becoming a cue. All varieties of fluctuated intervals

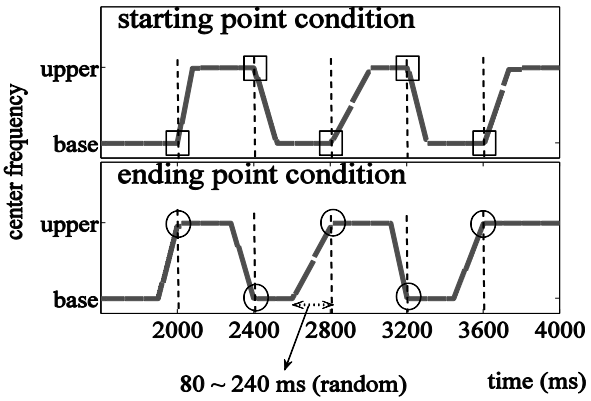

Fig. 2. Schematic figures of standard stimuli whose intervals between each starting or ending point were fixed.

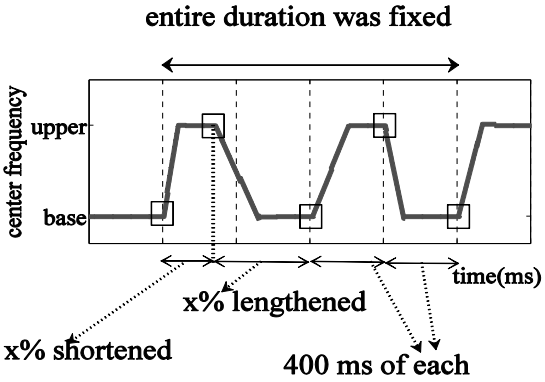

Fig. 3. Schematic figure of a comparison stimulus with no isochronous structures.

were contained in 12 patterns. Five deviational ranges were set: 5, 9, 16, 28, and 50% of 400 ms.

Method

All processes in this experiment were controlled on Mac OS X (Apple), using KYMA X as a sound processing application and Capybara-320 as a DSP black box (Symbolic Sounds Co.). Participants heard stimuli through SENNHEISER HD600 headphones. The experimental method had two intervals and two alternative forced choices, where the inter-stimulus interval was 800 ms. The order of the presentation of standard or comparison stimulus was randomized. Participants were instructed to choose the standard, isochronous stimulus from two intervals. Feedback was given visually after each response. There were 960 trials per participant, two combinations of deviational points (starting or ending points), two formant frequency ranges (1 or 2 ERB), five deviational rates (5, 9, 16, 28, or 50% of 400 ms), 12 deviational interval varieties, two presentation orders, and two repetitions. Participants were five undergraduate and graduate students (including one of the authors) ranging in age from 19 to 29. They had normal audition, and they were paid for participation.

Results

Figure 4 shows the results of the experiment. The height of each bar indicates the percentage of correct answers. The dark and light bars correspond to the ending and the starting point condition, respectively. The left panel shows the results of the 1-ERB condition and right panel the 2-ERB condition.

Under all but one condition, the correct answer rate of the starting point condition was higher than the ending point condition. Particularly in three conditions, (16% and 28% temporal deviation on 1-ERB formant frequency shift, and 28% of 2-ERB), significant differences in the correct rate were observed between the starting and ending point conditions ($p = 0.01$ at 16% on 1-ERB condition, $p =< 0.0001$ at 28% on 1-ERB, and $p < 0.0001$ at 28% on 2-ERB). When temporal deviation was small, i.e., 5% and 9%, in the 1-ERB condition, participants could not distinguish between standard and compared stimuli. However in the 2-ERB condition, the correct rates seemed slightly higher than the 1-ERB condition. In particular, the correct rate of the ending point condition at 5% deviation showed prominent growth compared with the 1-ERB condition, so that only this condition showed the

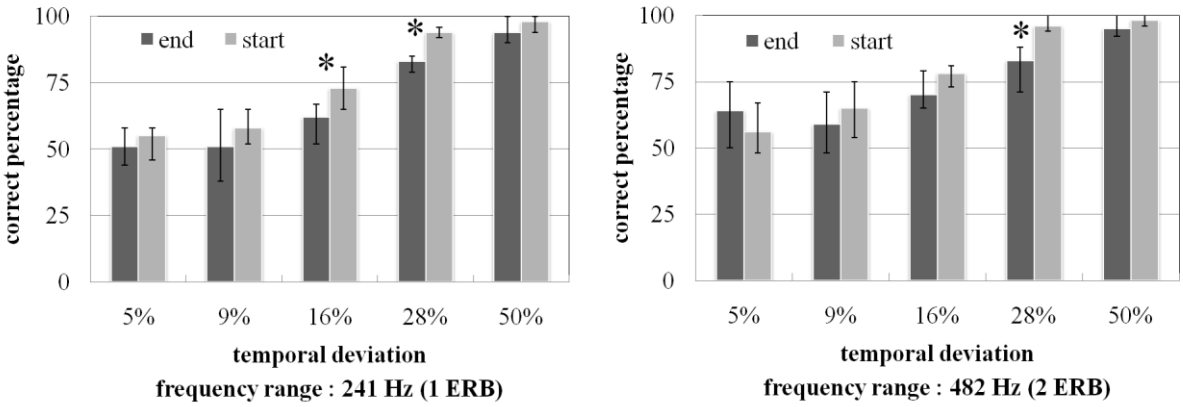

Fig. 4. Percentage of correct answers for each condition. Conditions marked by asterisk show significant differences between starting and ending point conditions.

opposite results from the other conditions. Although the reversal tendency in 2-ERB, 5% condition was not statistically significant, it could reach the significant criterion if the test was repeated by some additional participants. At the current stage, it is difficult to speculate what caused this reversal if any. At least, we should observe a certain reserve in concluding that the starting point provide more effective cue than the ending point for the temporal judgment.

Discussion

Experiment results show that higher sensitivity for temporal deviations on the starting point of change than the ending point was also observed when the change was realized by formant frequency glide. This indicates that the advantage of the starting point over the ending point is not peculiar to the frequency change of sinusoidal tones, validating our hypothesis that the starting point of a change in a sound signal can provide a more important cue to detect temporal deviation, such as a change in timing. However, the result also suggests that the possibility still remain for the ending point of the change to be more detectable than the starting point in certain cases. Further studies must clarify causes of the opposite result.

Acknowledgements

This work was supported by the Grant-in-Aid for Scientific Research A, No. 16200016, Japan Society for the Promotion of Science.

References

Tanaka, S., Tsuzaki, M., and Kato, H. (2006a). Timing perception of portamento: Difference in detectability of temporal deviation between the starting point and the ending point. *Proceedings of the Autumn Meeting of the Japanese Society for Music Perception and Cognition.* pp. 101-104.

Tanaka, S., Tsuzaki, M., and Kato, H. (2006b). Timing perception of tonal sequences connected with glides, *The Journal of the Acoustical Society of America, 120, 3084.*

Tsuzaki, M., Tanaka, S., and Kato, H. (2006). Timing perception of sequences with no obvious onset: Isochrony detection in sequences of complex tones connected by formant glides, *The Journal of the Acoustical Society of America, 120, 3084.*