

AUDITORY EVENTS THAT SEEM TO BE FORMED ILLUSORILY FROM OFFSET CUES

Yoshitaka Nakajima^{1),2)}, Tsuyoshi Kuroda³⁾, and Tomomi Terada¹⁾

1) Department of Human Science, Kyushu University,

2) Center for Applied Perceptual Research, Kyushu University
815-8540 Fukuoka, Japan
nakajima@design.kyushu-u.ac.jp

3) École de Psychologie, Université Laval, Québec, Canada

Abstract

It was revealed that a very short tone can be perceived only from an offset cue without any preceding change that can indicate an onset. We presented a harmonic glide tone complex of 2 s that ascended or descended between 200 and 400 Hz. The tone had five glide components, among which one was turned off in the temporal middle. Some listeners could hear a very short illusory tone around the temporal middle embedded in the complex. Subjective properties, i.e., pitch, duration, and loudness, of the tone were estimated employing a method of adjustment, but only with listeners who could hear the illusory tone constantly. Effects of individual listeners were observed systematically, suggesting that the illusory tone was created subjectively depending on each individual's perceptual framework. The subjective duration of the illusory tone, measured as the point of subjective equality, was always within a range of 40 to 450 ms in four among five participants—far longer than 20 ms, the lower limit of the adjustment determined as the sum of the rise and the fall time. We suggest that the tone was constructed from a real offset and an illusorily restored onset.

When several harmonic glide components ascend or descend together, and when one of these components is turned off around the temporal middle, some listeners can perceive a short tone embedded in the pattern. We call this subjective tone a termination tone. It is known that onsets alone can be perceived as an auditory stream; Steiger's (1980) informal observations as discussed in Bregman (1990) were probably the first systematic examples. However, it has not yet been reported that offsets alone can be perceptual cues of auditory events. One of the closest cases may be the illusory phenomenon reported by Sasaki et al. (2010), in which auditory events were constructed without onset cues. On the other hand, their stimulus patterns still included a discontinuous temporal change that could indicate the presence of an onset. We thus attempted to construct a stimulus pattern inducing an illusory short tone purely from an offset cue.

According to our pilot study, a termination tone could be heard only when the listener tried to hear one, and this happened only to some listeners. Furthermore, a termination tone, even when perceived, did not have a physical counterpart. All these things made it impossible for us to apply a well-established method to the investigation of this phenomenon. However, the phenomenon seemed very important to study how human listeners construct auditory events from limited perceptual cues. It should not be too unreasonable to assume that any auditory event must begin with an onset (Nakajima et al., 2000; Remijn et al., 2007; Kuroda et al., 2009; Sasaki et al., 2010), and, if so, it may be necessary for the auditory system to restore an onset when it detects a cue of an offset without detecting a cue of an onset to be connected to it.

We generated various sound patterns trying to find a simple pattern in which a short tone could be perceived from an offset cue without an onset cue or a cue of any sudden change preceding it. This attempt had not been as successful as we had hoped, but at least we found stimulus patterns in which some listeners could perceive short tones without corresponding onset cues. There was a problem, however: such an illusory tone did not appear spontaneously, but was heard only when the listener tried to hear a short tone near the middle of a pattern. In order to report this phenomenon in public, the fact that some listeners reported the perception of short tones did not seem sufficient. We aimed at clarifying what kind of tone the listener could hear without the presence of a corresponding physical tone. Even if our theoretical framework may be modified or changed in the future, the phenomenon itself is curious enough to be examined in this way.

Preliminary listening tests revealed that the illusory short tone, once heard, had pitch, loudness, and duration, as should be the case in most tones we hear. We decided to measure these subjective properties of the tone employing a method of adjustment. The listeners adjusted the frequency, duration, and level of a comparison pure tone to make the subjective properties as similar as possible to those of the illusory tone. Because it was difficult to specify physical cues for the perception of the short tone, we could not design an experiment in which real short tones would serve as control standard stimuli, as in conventional experiments on perceptual illusions. We somehow depended on a naïve assumption that the physical properties of the comparison tone corresponded to specific subjective properties, and this was a limitation of the present study. On the other hand, there should have been no better first step to grasp the nature of the illusory phenomenon.

Experiment

Stimulus patterns of five harmonic glide components, among which one was aborted, served as standard patterns. It was questioned whether and how a short tone could be heard out around the temporal middle. We also utilized control patterns in which no component was aborted. In a preliminary phase of the experiment, we selected listeners who could hear out illusory short tones stably, and we used twelve standard patterns that were the same as used in the main phase of the experiment except that the rise and fall times were 20 ms. The rise and fall times were changed to 10 ms in the main phase of the experiment in order to meet a technical requirement. Each stimulus pattern was presented three times. To each presentation, the listener responded whether he/she could hear a short tone near the middle of the pattern. We selected listeners who could hear a short tone in 90% or more of the non-control presentations. Among 12 listeners, 6 were selected, and 5 of them could hear a short tone in all the non-control presentations. This preliminary phase showed that the illusory phenomenon was not very robust, but that it could not be neglected.

In the main phase of the experiment, each standard pattern was followed by a short pure tone burst that served as comparison. The frequency, duration, and level of the comparison were adjusted together in the same trial, and we obtained a pure tone burst of similar pitch, loudness, and duration in the subjective domain in each trial.

Method

Participants:

Twelve students, five males and seven females, of Kyushu University took part in the first phase of this experiment. They were between 19 and 25 years old. Among the six listeners who participated in the main phase, three were male, and three female. Authors TK (male) and TT (female) were included.

Stimulus patterns:

All the stimulus patterns consisted of five harmonic glide components of 2 s (Figure 1). The lowest component corresponded to the continuous change of fundamental frequency, from 200 to 400 Hz in the ascending condition, or from 400 to 200 Hz in the descending condition. The frequency change was at a constant rate in logarithmic frequency. In control patterns, this fundamental component and four additional glide components, corresponding to the second to the fifth harmonics, were played together. In the other condition, one of the five glide components was aborted in the temporal middle, thus halving its duration to 1 s. In total, there were 2 control patterns (ascending/descending) and 10 non-control patterns [2 (ascending/descending) x 5 (aborted components)]. Each component had a rise time and a fall time of 10 ms of cosine-shaped ramps. The rise and the fall time had been 20 ms in the preliminary phase to reduce the audible spectral splatter (Miyasaka & Sakai, 1980). They were, however, shortened to 10 ms in the main phase to match with the rise and the fall time of the comparison tone, which were fixed at 10 ms to make it possible to measure very short subjective duration.

Each component of standards was 65 dB SPL when measured with an artificial ear (Brüel & Kjær 4153) and a sound level meter (Node 2071). Each comparison, which was a pure tone burst with 10-ms rise and fall times, was presented 4.0-4.5 s after the end of the standard. All the stimulus patterns were generated and presented with a computer (Frontier KZFM71/N), a D/A converter (Fostex VC-8), a low-pass filter (NF DV8FL), an equalizer (Roland RDQ-2031), a headphone amplifier (Stax SRM-313), and headphones (Stax SR-303). The sampling frequency and resolution were 44.1 kHz and 16 bits, and the stimulus patterns were presented diotically to the participant in a sound-proof room.

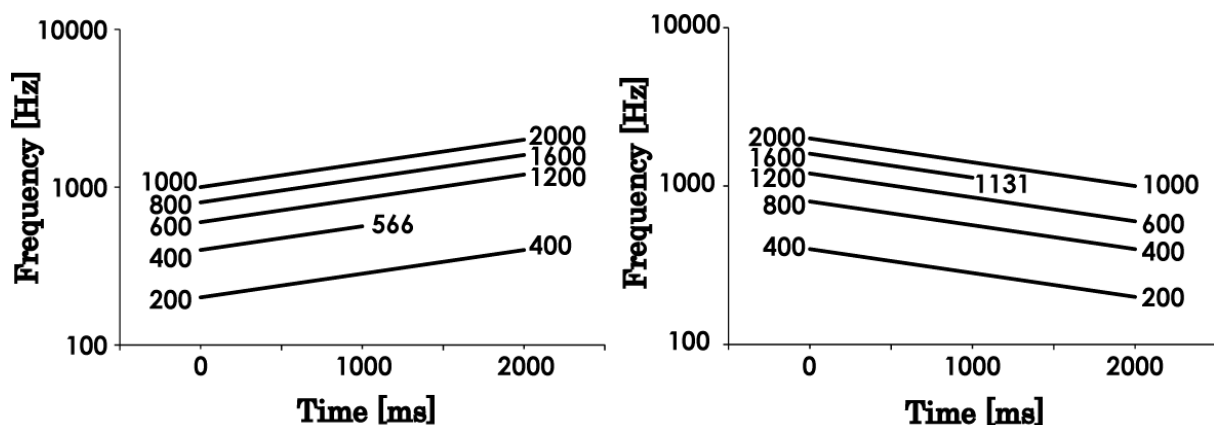


Figure 1 Examples of stimulus patterns.

Procedure:

In order to draw a rough sketch of the illusory tone that might be heard when one of the components was aborted, the participant was asked to generate a real pure tone burst in each trial that was equivalent to the illusory tone in terms of pitch, duration, and loudness.

All the standard stimulus patterns were presented once to the participant before the measurement trials. The participant initiated each presentation by clicking a button on the computer screen, and the stimulus pattern started after a silent period chosen randomly between 1.1 and 1.5 s. In each of the measurement trials that followed, a standard stimulus pattern and a comparison pure tone burst were presented in this order. A method of adjustment was applied. The participant adjusted the comparison by manipulating three slide controllers on the screen to change the frequency, the duration, and the intensity (converted to

a sound pressure level), but these physical properties were not mentioned in the instructions. The participant continued manipulating these properties until the pitch, the duration, and the loudness of the comparison were the same as those of the illusory short tone. The initial comparison values of frequency, duration, and intensity were chosen randomly within certain ranges. If the participant did not change any of these values, the initial values were considered as inappropriate, and the same trial was performed again without the participant knowing that the same standard was repeated. The slide controllers were labeled as pitch, duration, and loudness.

The initiation of each presentation was the same as above, and the silent period between the standard and the comparison in each presentation was chosen randomly between 4.0 and 4.5 s. One experimental block consisted of measurement trials for the twelve standard patterns in random order preceded by two warm-up trials utilizing the last two standard patterns to appear in the block. Each participant performed nine measurement blocks, among which the first was for training not to be analyzed. The physical values for the comparison were limited within the ranges 100-3200 Hz, 20-800 ms, and 20-78 dB SPL. These ranges were determined from pilot experiments, but it was necessary to change the duration range for one participant, and another participant showed that the frequency range was not wide enough in one trial. These incidents took place partially because the individual differences were far bigger than we expected, and because it was difficult to expect PSEs of the illusory tones that did not exist physically. How the individual cases were treated will be described in the next section.

There were also cases in which the participant could not hear a short tone. If this happened, the participant pressed a button for “could not hear a short tone,” and this trial was finished without measured values. When the participant was satisfied with the adjustment, the final values of frequency, duration, and level were recorded as the points of subjective equality (PSEs). These PSEs were analyzed separately to get a rough sketch of the illusory termination tone, but it is to be noted that these three values basically work as a single (multidimensional) measure of the illusory tone.

Results and Discussion

Among the six participants (L1..L6), one (L2) was not able to hear a short tone in 26 out of the 80 trials, and this participant was excluded from further statistical analyses. Another participant (L6) was not able to hear a short tone in 11 trials, but such cases were limited to the patterns in which the fourth harmonic was aborted; this participant was included in the analyses except for these patterns. In one participant (L5), the duration PSE reached the upper limit of 800 ms in 14 trials, and the whole experiment was carried out again for this participant increasing the upper limit for duration adjustment to 2500 ms; the new data were treated as the experimental data. In another participant (L1), the frequency PSE reached the upper limit 3200 Hz in one trial. Because this was the only anomaly of this kind in this participant, and because we adopted non-parametric statistical procedures which were basically immune to a few extreme values, we treated the data as they were.

The median PSEs of frequency, duration, and intensity (converted to sound pressure levels) for each standard stimulus were calculated for each participant (Table 1). No clear tendencies may be observed in these results, but an important point is that the PSEs were within certain ranges. The PSEs of duration are especially to be noticed (Figure 2). There were considerable individual differences, and it seemed reasonable to separate one participant (L5) as exceptional; the duration PSEs never reached nor exceeded 500 ms in the other four participants, but the corresponding PSEs in this participant were above 500 ms in most cases, i.e., in 70 out of the 79 trials in which a termination tone was audible. In the other four

participants (L1, L3, L4, and L6), the duration PSEs were limited within a range between 42 and 435 ms. The values were far below the length of the aborted component, 1 s, and it is unlikely that the participants judged its duration. The values, however, were considerably

Table 1. Median points of subjective equality.

Participant	Ascending patterns					Descending patterns				
	Frequency (Hz)									
	Aborted component					Aborted component				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
L1	1670.3	1054.1	1063.3	1513.4	827.9	1586.2	1080.7	1552.9	1714.3	1150.3
L3	271.5	279.9	392.3	570.1	325.0	312.9	316.5	315.0	314.6	324.5
L4	279.8	581.6	1063.6	2204.9	2297.4	296.9	582.2	1029.0	1438.8	1596.6
L5	263.1	273.1	724.7	543.6	830.8	355.8	333.8	919.9	989.0	911.1
L6	267.2	293.4	209.0		588.4	295.8	311.2	395.9		433.6
Median	271.5	293.4	724.7	1041.7	827.9	312.9	333.8	919.9	1213.9	911.1
Offset frequency	282.8	565.7	848.5	1131.4	1414.2	282.8	565.7	848.5	1131.4	1414.2

Participant	Duration (ms)									
	Aborted component					Aborted component				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
L1	91.5	136.5	122.0	113.0	98.5	98.5	121.5	128.5	112.5	126.5
L3	250.0	198.0	176.5	87.5	159.5	216.0	181.0	176.0	208.5	187.5
L4	116.0	94.5	98.0	105.0	93.0	124.0	103.0	118.0	124.5	108.0
L5	743.5	718.5	647.0	596.0	644.0	716.0	676.0	680.0	631.0	645.0
L6	65.5	55.5	61.0		53.5	65.5	49.0	59.0		63.0
Median	116.0	136.5	122.0	109.0	98.5	124.0	121.5	128.5	166.5	126.5

Participant	Sound Pressure Level (dB)									
	Aborted component					Aborted component				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
L1	56.9	56.4	50.3	49.9	48.8	58.0	54.7	44.8	45.5	52.7
L3	58.1	51.7	53.2	57.2	51.8	58.8	58.0	61.1	51.3	55.7
L4	50.8	52.3	52.8	42.8	45.2	49.9	48.3	45.8	47.2	43.35
L5	60.3	58.8	45.5	47.7	51.9	55.2	51.7	48.5	48.5	55.15
L6	53.9	48.8	46.4		51.7	53.8	53.2	46.3		50.75
Median	56.9	52.3	50.3	48.8	51.7	55.2	53.2	46.3	47.9	52.7

larger than the lower limit of the duration adjustment; the participants could make the values smaller, but did not. In fact, there were many cases in which duration PSEs were above 100 ms (as is clear in Figure 2). It is, therefore, unlikely that only a short period around the fall time was involved in the perceptual construction of the termination tone.

For the frequency PSEs, effects of the aborted component may be observed for each participant. There seems to be a similarity between the results for the ascending patterns and for the descending patterns in each participant. The present data, however, do not seem enough to discuss this issue further. The intensity PSEs were mostly in a range of 40-60 dB SPL, and it should not be a strange idea, considering the effects of masking, that the auditory system utilized the last portion of the aborted component of 65 dB SPL to construct the illusory tone.

Our tentative hypothesis is that the offset of the aborted component was not sufficient alone to construct an auditory event, and that an onset was restored subjectively when the participant was somehow forced to judge the duration of a short event in the present experimental procedure (Nakajima et al., 2000; Sasaki et al., 2010). This hypothesis works also for the exceptional participant (L5) whose duration PSEs were longer than those obtained from the other participants. The problem of the present paradigm was that we were not able

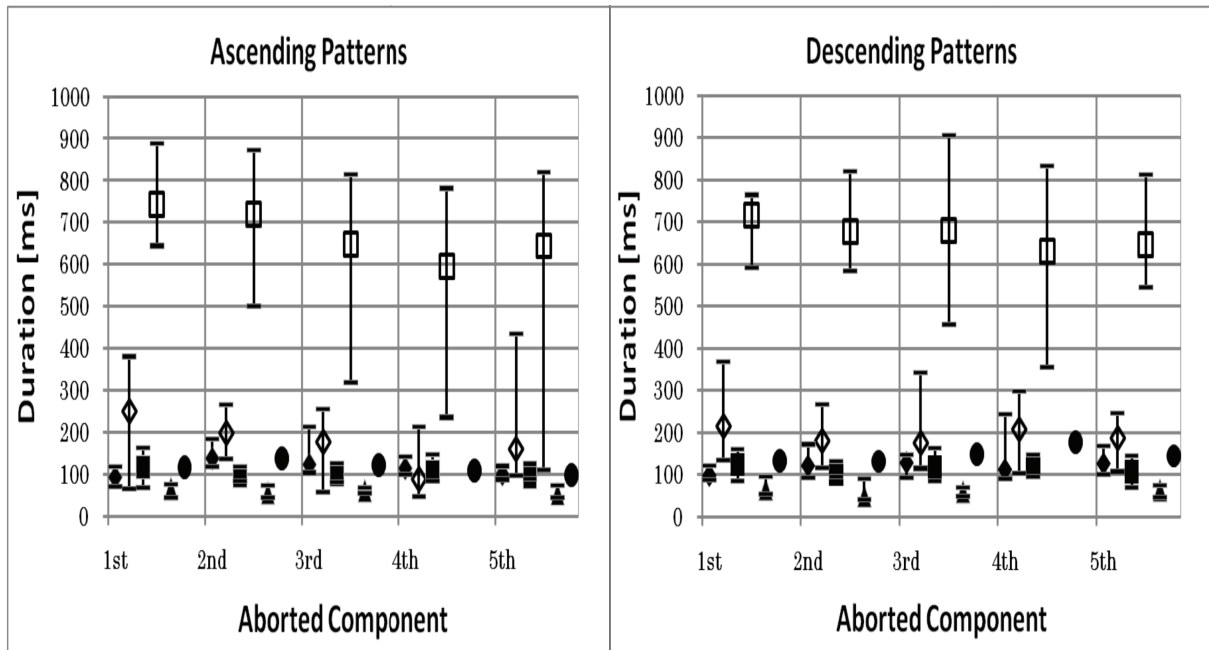


Figure 2 Median points of subjective equality of duration. The median and the range of the measured values for each participant are indicated for each stimulus pattern. The five participants (L1, L3, L4, L5, L6) selected for the present analyses are shown by different point shapes from left to right. For one participant (L6), data for the fourth components could not be obtained. The median of the individual median values for each stimulus pattern is indicated at the right end with a solid circle.

to compare the obtained PSEs with some kind of control values in order to clarify the underlying perceptual mechanism quantitatively, but at least we were able to describe the subjective nature of the illusory tone in a rough manner. The individual differences were very big, but the illusory tone could be interpreted as a subjective counterpart of a portion, or a chimera of portions, of the stimulus pattern.

Acknowledgments

Ger Remijn gave us valuable comments. Yuki Kamimura gave us technical assistance. This work was supported by the Japan Society for the Promotion of Science (19103003 in fiscal 2007-2010; 20653054 in fiscal 2008-2010).

References

- Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Kuroda, T., Nakajima, Y., Tsunashima, S., & Yasutake, T. (2009). Effects of spectra and sound pressure levels on the occurrence of the gap transfer illusion. *Perception*, 38, 411-428.
- Miyasaka, E., & Sakai, H. (1980). Detectability of switching transients. *Journal of the Acoustical Society of Japan*, 36, 244-252. (in Japanese)
- Nakajima, Y., Sasaki, T., Kanafuka, K., Miyamoto, A., Remijn, G., & ten Hoopen, G. (2000). Illusory recouplings of onsets and terminations of glide tone components. *Perception & Psychophysics*, 62, 1413-1425.
- Remijn, G., Nakajima, Y., & Tanaka, S. (2007). Perceptual completion of a sound with a short silent gap. *Perception*, 36, 898-917.
- Sasaki, T., Nakajima, Y., ten Hoopen, G., van Buuringen, E., Massier, B., Kojo, T., Kuroda, T., and Ueda, K. (2010). Time stretching: Illusory lengthening of filled auditory durations. *Attention, Perception, & Psychophysics*, 72, 1404-1421
- Steiger, H. (1980). Some informal observations concerning the perceptual organization of patterns containing frequency glides. Technical Report, McGill University, Montreal.