Dichotic continuity illusion for steady-state versus frequency-gliding sounds

Tsuyoshi KURODA1*; Simon GRONDIN2; Makoto MIYAZAKI1
1 Shizuoka University, Japan
2 Université Laval, Canada

ABSTRACT
A sound with a silent gap is perceived as continuous when this gap is filled with a louder noise. This phenomenon, called the continuity illusion, has been posited to relate with peripheral simultaneous masking. However, we found that the illusion could occur in dichotic listening, i.e., when the discontinuous sound and the inserted noise were presented at different ears. This indicates that the occurrence of the continuity illusion could not be predicted with the peripheral excitation pattern in a single ear. Moreover, the dichotic mode of the illusion more clearly occurred when the frequency of the discontinuous sound ascended or descended in time than when it remained stable. Different neural mechanisms appeared to underlie the perceived continuity of steady state sounds and that of frequency moving sounds. In this proceeding paper, we report the results of a preliminary experiment that was conducted before a series of main experiments for the present project.

Keywords: Continuity illusion, Dichotic listening, Frequency modulation
I-INCE Classification of Subjects Number: 63

1. INTRODUCTION
A sinusoidal sound with a silent gap is perceived as continuous when the gap is filled with a louder noise [1] (Figure 1). This continuity illusion has been explained in terms of the relation with peripheral masking [2]. The auditory system interprets the missing part of the discontinuous sound as being present but masked by the inserted sound, resulting in the perceptual restoration of the missing part (the masking-criteria hypothesis). However, some limitations of this explanation have been voiced recently [3–6]. For example, Riecke et al. [5] indicated that the illusion depends on the overall loudness rather than on the peripheral masking properties of the inserted sound.

Figure 1 – The continuity illusion. A sine tone with a silent gap (a-a’) is perceived as continuous when the gap is filled with a louder noise (b-b’).

* tkuroda@inf.shizuoka.ac.jp
However, an evidence challenging the masking-criteria hypothesis had already been found before the masking-criteria hypothesis was proposed in 1972 [2]. A research group directed by L. Elfner [7,8] reported that the continuity illusion occurred in dichotic listening, i.e., when the discontinuous and inserted sounds were presented at different ears. The occurrence of the continuity illusion thus cannot be attributed to the peripheral excitation pattern in a single ear, and the illusion is involved with neural mechanisms after the cross of neural signals from the two ears. However, those previous researches used a noise as a discontinuous sound and a tone burst as an inserted sound. This structure (noise-tone-noise) is contrary to the one used recently; a tone and a noise are used as a discontinuous and an inserted sound, respectively (tone-noise-tone).

Bennett et al. [9] provided the data suggesting the occurrence of the continuity illusion in dichotic listening with the tone-noise-tone structure, but, in their experiment (Experiment 3), the frequency of the discontinuous sound moved around the gap; in other words, the experiment had no condition where the frequency remained stable. Previous studies have provided no conclusive arguments concerning whether frequency glides and steady sounds differed in the occurrence of the continuity illusion in monaural listening [10,11]. We thus conducted the present preliminary experiment in order to examine, firstly, whether the dichotic continuity illusion could be replicated with the tone-noise-tone structure as well as the noise-tone-noise structure, and, secondary, whether the illusion would be modulated by the frequency dynamics of sounds in dichotic and monaural listening. The experiment had four types of target sounds whose continuity was to be judged by participants: an ascending glide, a sinusoidal sound of steady frequency, a descending glide, and a band noise (Figure 2). For the first three cases, the inserted sound (distractor) was a band noise (tone-noise-tone), while, for the last case, it was a sinusoidal burst (noise-tone-noise).

2. METHOD

2.1 Participants

Sixteen volunteers (9 females), aged 20-39 years, were recruited from Yamaguchi University. They self-reported having normal hearing, and consented to participate by signing a form approved by the institutional ethics review board. One additional participant was recruited but the data from this participant was not kept for the analysis because of a low performance of discriminating the physical continuity of the noise target in the learning block where no distractor was presented (the correct probability did not exceed the criterion even in the third round of the learning block).

2.2 Stimuli and Apparatus

Digital signals of sound stimuli were sampled at 44100 Hz and quantized to 16 bits. These signals were converted into analogue ones in an USB audio interface (Onkyo SE-U33GXVII). The stimuli were presented from headphones (Sennheiser HD 650) via an amplifier (Teac A-H01). There were four types of target sounds (Figure 2). The ascending target was a single frequency
glide moving from 500 to 2000 Hz linearly in logarithmic frequency. The descending target was a glide moving in the opposite direction. The steady target was a sinusoidal sound fixed at 1000 Hz. The noise target was a band noise between 500 and 2000 Hz. Each target lasted 1000 ms and was presented from the left or the right channel of the headphones.

The band noise was a mixture of 500 sinusoidal components. The initial phase of each component was decided randomly. Its frequency was also decided randomly but with a constraint that the resulting spectrum kept almost white. The noise was generated on each trial and thus non-frozen.

The target was physically continuous or discontinuous. In the latter case, a gap of 12 ms was inserted in the temporal middle of the target. To avoid spectral splatter, amplitude rose and decayed during 8 ms. These ramps formed a shape of a half part of a cosine curve in time-power domain (not time-amplitude domain). The boundary of the gap duration was defined as a 3-dB down points (Figure 3). Thus, the 12-ms gap included the last half (4 ms) of the decay ramp of the target, zero points of 4 ms, and the first half (4 ms) of the rise ramp of the target.

The target was accompanied with no distractor (the control condition), with a distractor at the opposite ear (the dichotic condition), with a distractor at the same ear (the monaural condition). The distractor, lasting 12 ms, was a 500-2000 Hz band noise for the ascending, steady and descending targets whereas it was a 1000 Hz sinusoidal burst for the noise target. The distractor was centered at the temporal middle of the target. When the target was physically discontinuous, the ramps edging the gap were overlapped with those of the distractor (Figure 3). The target and the distractor were presented at 30 dB and 50 dB above the detection threshold, respectively. The details of the threshold measurement are explained below.

Consequently, the experiment was based on a 2 (target ears) × 2 (physical continuities of targets) × 3 (distractor ears plus a control) × 4 (target types).

![Power envelope around the gap of physically discontinuous targets.](image)

### 2.3 Procedure

#### 2.3.1 Continuity Judgment (Main Task)

The four targets were presented in separate sessions. Each of the four sessions was divided into five experimental blocks, between which a few-second break was inserted. In each block, the twelve stimuli (= 2 target ears × 2 target continuities × 3 distractor conditions) were presented twice each in a random order, resulting in 24 trials. Participants were instructed to judge whether the target was continuous or discontinuous after the presentation of the stimulus. They pressed the left and the right button of a mouse for responding “continuous” and “discontinuous,” respectively. The next stimulus began 1-2 s after the participant’s response.

Each session began with a learning block, where only the no-distractor condition was presented. The four stimuli (= 2 target ears × 2 target continuities) were presented three times each in a random order, resulting in 12 trials. The same task as in the experimental block was used, but the feedback message was presented on a computer display after the participant’s response. This learning block was repeated until participants correctly discriminated the physically continuity of the target in more than 10 trials (i.e., the correct proportion exceeded 90%). In fact, all participants, except the one removed from the analysis as mentioned above, fulfilled this criterion by the end of the third round of the learning block in each session. Then, they moved on to the first experimental block.
2.3.2 Threshold Measurement (for Intensity Calibration)

Elfner and Homick [8] set the presentation level of the inserted sound and of the discontinuous sound at 50 dB and 30 dB above the threshold, respectively. These values were based on the findings of the masking experiments [12,13] where the threshold level for detecting a signal was not influenced by a masker presented at the opposite ear until the masker was boosted up to 50 dB above threshold. Thus, Elfner and Homick attempted to avoid a sound leakage across ears in their dichotic-listening experiment.

Following this previous study, we measured the detection threshold at the beginning of the experiment. There were a rough measurement and a fine measurement. The former was conducted first. In the rough measurement, the experimenter adjusted the volume of the amplifier so that the participant could just detect a signal noise that was normalized at a certain level in the digital (.wav) file. This level is called the reference level. The signal noise was a 300 ms, 500-2000 Hz band noise with the same spectral property as in the main experiment.

The fine measurement was then conducted based on the method of limits (see [14]). There were four blocks, each consisting of four trials. The ear at which the signal noise was presented was constant throughout each block. The order of the left-ear (L) block and the right-ear (R) block was LRLR for half participants and RLRL for the others. Each block consisted of two ascending-series (A) and two descending-series (D) trials. The order of these trials was ADAD for half participants and DADA for the others. The ascending series began with the presentation of the signal at a (undetectable) level that was 8, 10, or 12 dB lower than the reference level. After the signal presentation, participants judged whether the signal could be detected or not by clicking on a pane on the computer display; thus, they were expected to respond “no” for the first presentation in the ascending series. The signal was then increased by 2 dB. The presentation was repeated until participants responded “yes” for two or three consecutive presentations. The descending series began with the presentation of the signal at a (detectable) level that was 8, 10, or 12 dB higher than the reference level, and the signal was decreased by 2 dB per presentation. The threshold was a median of two consecutive levels where the participant’s response changed from “no” to “yes” for the ascending series and from “yes” to “no” for the descending series. When participants exhibited an unexpected response for the first presentation in each series (i.e., “yes” for the ascending and “no” for the descending series), the experimenter re-adjusted the volume of the amplifier and restarted the measurement. The first two blocks were regarded as a practice, and thus, the results of the four trials for each of the last two blocks were averaged. The resulting value was used for calibrating the intensity of each sound used in the main experiment.

3. RESULTS

The first of the five experimental blocks for each session was regarded as a practice and removed from the analysis. Furthermore, the target-ear conditions were collapsed. The dependent variable for each experimental condition was thus based on 16 responses (4 blocks × 2 responses × 2 target ears). The probability of responding “continuous” was estimated for each experimental condition and then subject to the arcsine transformation in order to conform the linearity of the scale. The mean arcsine transform score for each experimental condition is shown in Figure 4. Several tendencies could be found from the figure when the target was physically discontinuous: The probability was higher in the dichotic and monaural conditions than in the no-distractor condition. Moreover, in the dichotic condition, the probability was higher for the ascending and descending targets than for the steady and noise targets.

An analysis of variance (ANOVA) with repeated measures according to a 3 (distractor locations) × 4 (target frequency states) design was conducted separately on the two target continuities (physically continuous and discontinuous). Degrees of freedom were adjusted with Greenhouse-Geisser epsilon against potential violation of sphericity. For the physically continuous target, the distractor-location effect, \( F(1.43, 21.46) = 12.195, p < .001, \eta^2_p = .448 \), as well as the interaction, \( F(3.99, 59.9) = 3.482, p = .013, \eta^2_p = .188 \), was significant, but the frequency-state effect was not significant, \( F(2.77, 41.62) = .764, p = .512, \eta^2_p = .048 \). For the physically discontinuous target, the frequency-state effect, \( F(1.95, 29.23) = 14.877, p < .001, \eta^2_p = .498 \), the distractor-location effect, \( F(1.99, 29.85) = 58.874, p < .001, \eta^2_p = .797 \), and the interaction, \( F(3.63, 54.38) = 4.089, p = .007, \eta^2_p = .214 \), were all significant. Because the interaction was significant in each case, the simple main effects of the frequency state and of the distractor location were tested with a one-way repeated-measures ANOVA with the Greenhouse-Geisser correction. When the simple main effect
was significant, pairwise comparisons were conducted according to the Holm method. The results of the post hoc analysis involved with the frequency state are reported in Figure 4. Note that the results of the post hoc analysis involved with the distractor location are not reported in the figure for avoiding complicated data presentations but those results can be summarized as follows: For the physically continuous target, the “continuous” probability was decreased in the monaural condition compared with the no-distractor condition, especially when the steady target was used. For the physically continuous target, the probability was increased in the dichotic and monaural conditions compared with the no-distractor condition, regardless of the frequency state.

Figure 4 – Mean arcsine transform score of the “continuous” probability in each experimental condition for the physically continuous target (upper) and for the physically discontinuous target (lower). Bars are 95% confidence intervals. The results of the post hoc contrasts involved with the frequency state are also shown (*p < .05, **p < .01, ***p < .001).
4. DISCUSSION

The continuity illusion has previously been demonstrated with the presentation of the distractor and the target at the same ear [1–5,10,11]. However, it could take place in the present study even when these sounds were presented at different ears with the tone-noise-tone structures as well as the noise-tone-noise ones [7,8]. Indeed, the probability of responding that the target was continuous was increased in the dichotic and monaural conditions compared with the no-distractor condition. Moreover, the probability was higher when the target’s frequency ascended or descended than when it remained stable, suggesting a stronger illusion for frequency dynamic sounds than for static sounds. Given that in the present study the monaural condition did not lead to much differences between the target types (see also [11]), dichotic listening might activate a restoration mechanism specialized for frequency dynamic sounds.

Unexpectedly, the physically continuous target was perceived as discontinuous in the monaural condition, especially when the frequency was stable. The same effect was suggested in a very limited number of studies [9,15]; most studies used only physically discontinuous sounds. It should be a future avenue to examine what factor causes the perceived discontinuity of physically continuous sounds. We should keep in mind that this issue may influence the interpretation of the differences between the frequency states in the monaural condition in the present study.

It may be a technical limitation of the present study that open-type headphones were used. This type might be different from what was used in the previous studies [12,13] that the present presentation levels were based on. We were thus unsure if a sound leakage across ears was completely washed out in the present experiment. It should be more appropriate to use closed-type headphones for the purpose of diminishing a sound leakage across ears.

5. CONCLUSIONS

The continuity illusion can take place in dichotic listening regardless of the structure (tone-noise-tone or noise-tone-noise) of sound patterns, and its occurrence may be more robust when the frequency of the discontinuous sound moves temporally than when it remains stable. Experiments now in progress are more systematically examining the requirements for the occurrence of the dichotic continuity illusion. We will report the results of these experiments in Inter-Noise 2016.

ACKNOWLEDGEMENTS

This work was made possible by research grants awarded to TK by JSPS KAKENHI Grant Number 15K21195, 16H01502; to MM by JSPS KAKENHI Grant Number 26119521, 16H01866; and to SG by the Natural Sciences and Engineering Council of Canada, without any commercial or financial relationships that could be regarded as a potential conflict of interest.

REFERENCES