Perception, 2015, volume 44, pages 198–214

doi:10.1068/p7753

Effects of frequency separation and diotic/dichotic presentations on the alternation frequency limits in audition derived from a temporal phase discrimination task

Shoko Kanaya[1,2,3], Waka Fujisaki[2], Shin'ya Nishida[4], Shigeto Furukawa[4], Kazuhiko Yokosawa[1]
[1]Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan; [2]National Institute of Advanced Industrial Science and Technology (AIST), Japan; [3]Japan Society for the Promotion of Science; [4]NTT Communication Science Laboratories, NTT Corporation, Japan; e-mail: skanaya@l.u-tokyo.ac.jp
Received 19 March 2014, in revised form 2 December 2014

Abstract. Temporal phase discrimination is a useful psychophysical task to evaluate how sensory signals, synchronously detected in parallel, are perceptually bound by human observers. In this task two stimulus sequences synchronously alternate between two states (say, A-B-A-B and X-Y-X-Y) in either of two temporal phases (ie A and B are respectively paired with X and Y, or vice versa). The critical alternation frequency beyond which participants cannot discriminate the temporal phase is measured as an index characterizing the temporal property of the underlying binding process. This task has been used to reveal the mechanisms underlying visual and cross-modal bindings. To directly compare these binding mechanisms with those in another modality, this study used the temporal phase discrimination task to reveal the processes underlying auditory bindings. The two sequences were alternations between two pitches. We manipulated the distance between the two sequences by changing intersequence frequency separation, or presentation ears (diotic vs dichotic). Results showed that the alternation frequency limit ranged from 7 to 30 Hz, becoming higher as the intersequence distance decreased, as is the case with vision. However, unlike vision, auditory phase discrimination limits were higher and more variable across participants.

Keywords: temporal phase discrimination, binding, audition, vision, cross-modal, time

1 Introduction

The human brain processes sensory information in parallel. It is necessary, then, for the brain to bind these separately processed signals into correct representations of the original sources and objects (Treisman, 1996). Sensory signals that originate from a single source or object usually arise simultaneously. Thus, temporal synchrony between signals is considered a reliable cue to bind these signals.

The temporal phase discrimination task (Aghdaee & Cavanagh, 2007; Amano, Johnston, & Nishida, 2007; Arnold, 2005; Bartels & Zeki, 2006; Clifford, Holcombe, & Pearson, 2004; Fujisaki & Nishida, 2010; Holcombe & Cavanagh, 2001; Holcombe & Judson, 2007; Maruya, Holcombe, & Nishida, 2013; Moutoussis & Zeki, 1997; Nishida & Johnston, 2002; Victor & Conte, 2002) is a useful psychophysical method to gain insight into how humans integrate information from different sensory channels. In this task each stimulus sequence is composed of two alternating states of a specific sensory attribute at a specific temporal frequency over several seconds (eg A-B-A-B… and X-Y-X-Y…). Participants report the relative phase relationship between them (ie whether A is in synchrony with X or Y). In general, the task becomes more difficult as the alternation frequency increases. The critical alternation frequency beyond which participants cannot perform this task is termed the temporal frequency threshold or the alternation frequency limit. It changes depending on a variety of stimulus factors, supposedly reflecting the temporal characteristics of the underlying binding process.

The phase discrimination task has been used to investigate visual bindings (eg within or between colors, luminance levels, orientations, and motion directions) and cross-modal bindings (Fujisaki & Nishida, 2010; Holcombe & Cavanagh, 2001; Holcombe & Judson, 2007; Maruya et al., 2013). Binding can occur at various levels ranging from early (peripheral) to late (central) in the perceptual system (Treisman, 1996). The early bindings are thought to be relatively fast, while the late bindings are thought to be relatively time-consuming. In general, the results indicate that the alternation frequency limit is relatively high when two stimulus sequences are close in space or similar in features, and relatively low when stimuli are distant from each other (see also Aghdaee & Cavanagh, 2007; Forte, Hogben, & Ross, 1999; Holcombe & Cavanagh, 2001; Victor & Conte, 2002). This is consistent with the idea that fast peripheral mechanisms are responsible for the binding of close stimulus pairs, while slow central mechanisms are responsible for binding distant pairs (Fujisaki & Nishida, 2010; Holcombe, 2009). In addition, the similarity in the alternation frequency limit for binding across different attributes suggests the possible existence of a general central binding mechanism (Fujisaki & Nishida, 2010).

The purpose of this study is to extend this line of research to the purely auditory domain. We examined whether the alternation frequency limit could be reliably measured for auditory temporal phase discrimination tasks, and if so, whether the limit showed properties similar to those found for vision, in particular the effect of distance. Several auditory studies have investigated auditory binding (or grouping and segregation) performance in humans using tasks that might tap the same mechanisms as those tapped by the temporal phase discrimination task (Carlyon, 1991, 1994; Furukawa & Moore, 1996, 1997; Richards, 1987; Strickland, Viemeister, Fantini, & Garrison, 1989; Yost & Sheft, 1989). Some prior studies have suggested effects of intersequence distance.[1] However, since the tasks were not exactly identical to the temporal phase discrimination task, we cannot tell what the upper limit of alternation frequency is, nor whether the limit is faster or slower than that of, for example, visual color–color binding. Once we can characterize auditory bindings using the same tasks and the same performance measures as those used in other modalities, we will be able to directly compare binding mechanisms across different modalities (for details, see section 6 and figure 7), and obtain a more global and coherent view on the mechanisms of feature binding in humans based on temporal synchrony.

In two experiments we measured the alternation frequency limits while varying the distance between two oscillating auditory stimulus sequences. Experiment 1 varied frequency separation [equivalent rectangular bandwidths (ERBs) of 0.4 and 3.3] between two dichotically presented stimulus sequences (each sequence was presented monaurally and the two different sequences were presented to different ears). In experiment 2 two oscillating stimulus sequences were presented diotically (to the same ears), or dichotically to different ears.

2 General method

2.1 *Participants*

The participants were university students who had normal hearing and normal or corrected-to-normal vision. Informed consent was obtained from the participants before the experiment began. The experiment was approved by the Institutional Review Board of University of Tokyo and was performed in accordance with the Declaration of Helsinki.

[1] See section 6 for details.

2.2 *Apparatus*

Experiments were conducted in a quiet, dark room. A small, dim light on the table illuminated the response keys. Experiments were controlled with MATLAB (MathWorks) on a PC (Epson MT 7500). Auditory stimuli were presented via headphones (Sennheiser HDA 200) through a USB Audio System (Creative SB1240). Feedback, by color change of the fixation point, was presented on a CRT monitor (Sony CPD G520, frame rate 160 Hz).

2.3 *Stimuli*

Two oscillating stimulus sequences, which consisted of two tones appearing alternately (ie A-B-A-B… and X-Y-X-Y…), were used (figure 1a). For convenience, we refer to the sequences as AB and XY. Each sequence lasted 6 s. Several different alternation frequencies, typically six ranging from 1.4 to 8 Hz in half-octave steps, were used in the experiment (figure 1b). To prevent participants from judging the phase relationship from the onset and offset of the sequences, each had 2 s cosine ramps at onset and offset (figure 1b). In addition, the relative phase of the two sequences started with a random phase and gradually shifted to the intended phase relationship (0° or 180°) over the initial 2 s.

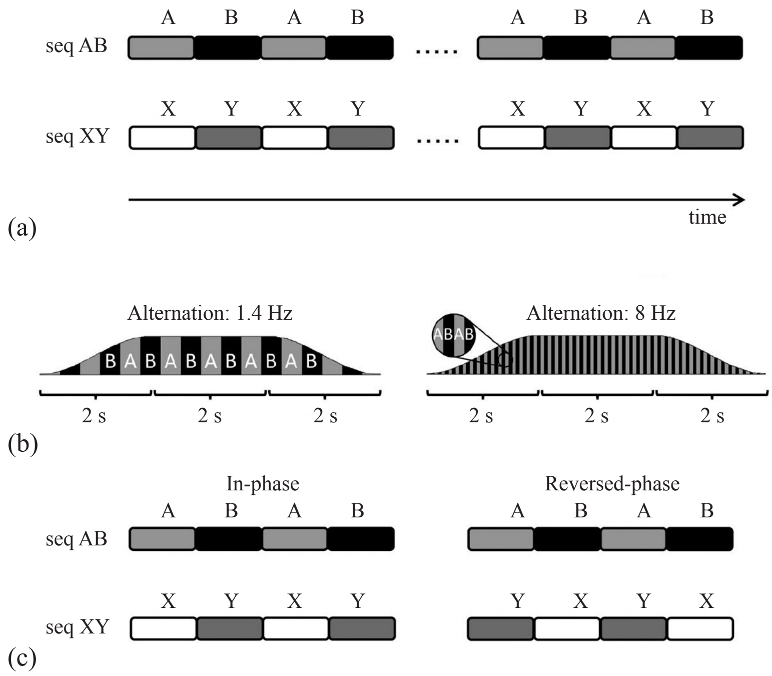

Figure 1. A schematic illustration of the temporal phase discrimination task. (a) The structure of the oscillating stimulus sequences (AB and XY sequences). (b) To prevent participants from judging the phase relationship from the onset and the offset of the sequences, a sequence had 2 s cosine ramps both at onset and offset. (c) For convenience, we named the phase relationship in which A is synchronized with X the ‘in-phase’ and that in which A is synchronized with Y the ‘reversed-phase’ sequence. Note: seq = sequence.

2.4 *Procedure*

In each trial the stimulus sequences (ie AB and XY) were presented simultaneously. Participants made a two-alternative forced-choice response about the phase relationship (ie whether A is in synchrony with X or Y) by pressing one of two response keys. For convenience, we named the phase relationship in which A is synchronized with X as the ‘in-phase’ and that in which A is synchronized with Y as the ‘reversed-phase’ (figure 1c). The participants were instructed to ignore the first 2 s and judge the phase relationship of the last 4 s. To prevent the participants from incorrectly assigning them to the response keys, feedback (in-phase or out-of-phase) was given by changing the color of the fixation point.

One block consisted of 12 trials. The first two were practice trials that consistently started with the in-phase and concluded with the reversed-phase. Data were collected from the subsequent ten trials, which included a randomized presentation of five in-phase and five reversed-phase trials. The alternation frequency was fixed within a block. One session consisted of several blocks with different alternation frequencies, typically six alternation frequencies ranging from 1.4 to 8 Hz in half-octave steps; higher frequencies were added when necessary. There were two types of sessions: ascending (from the lowest to the highest frequency) and descending (from the highest to the lowest frequency).

3 Experiment 1

Experiment 1 measured participants' performance on temporal phase discrimination tasks in which the frequency separation between two pure-tone sequences at the closest point was varied from near to far.

Previous research in the visual modality has indicated that the alternation frequency limits drop when two stimulus sequences are spatially separated (Aghdaee & Cavanagh, 2007; Forte et al., 1999; Holcombe & Cavanagh, 2001; Victor & Conte, 2002). In the visual system, spatial information is topographically represented at the retinal level. In the auditory domain, spatial information is represented at a relatively high cortical level, where the information from each ear is integrated. Thus, manipulation of the spatial location of sounds is not analogous to the manipulation of the spatial location of visual stimuli. A more plausible auditory equivalence of spatial topography in vision is the topographical representation of stimulus frequency, known as tonotopic organization, which occurs already at the level of the basilar membrane. The retina and basilar membrane are also highly equivalent in the sense that they are each the initial receptor surface where the light or sound energy is transduced. Therefore, in experiment 1 we measured the limits of the auditory temporal phase discrimination tasks, manipulating the frequency separation between two auditory stimulus sequences as an analogy of spatial separation in vision. The two stimulus sequences were separately presented to opposite ears (dichotic presentation).

3.1 *Method*

3.1.1 *Participants.* The participants were twelve university students aged 20–24 years (mean = 21.5 years, SD = 1.17 years).

3.1.2 *Stimuli.* There were two conditions: near and far (figure 2a). In the near condition AB and XY sequences alternated between 294 and 494 Hz (close to D4 and B4 in musical note names) and 523 and 1109 Hz (close to C5 and C#6). In the far condition AB and XY sequences alternated between 262 and 367 Hz (close to C4 and F#4) and 622 and 880 Hz (close to D#5 and A5). These carrier frequencies were chosen so that none of the frequencies shared the same musical note tone and so that none of the combinations had a simple harmonic relationship (eg 2 : 1, 3 : 2, 4 : 3). Each individual stimulus sequence was composed of two stationary pure tones, which were individually amplitude-modulated and superimposed. The waveform modulating each component was a square wave with a modulation depth of 100%. The phase of a waveform modulating one component was lagged 180° behind the other. To reduce spectrum splatter caused by transient changes of frequency components, the lead and the tail edges of the square wave were modulated by 10 ms cosine ramps from the onset and offset of the squares. Each sequence was presented to one of the two ears with a sound pressure level of 60 dB.

3.1.3 *Procedure.* In both the near and far conditions each sequence was presented dichotically (a different sequence in each ear). The frequency separation between two sequences at their closest was small (around 0.4 ERB[(2)]) in the near condition and relatively large (around

[(2)] One ERB represents the equivalent rectangular bandwidth of a single auditory filter (Glasberg & Moore, 1990). See section 6.

3.3 ERB; figure 2a, bottom) in the far condition (see Glasberg & Moore, 1990). The whole frequency bands covering two sequences in the near and far conditions were not largely different (figure 2a).

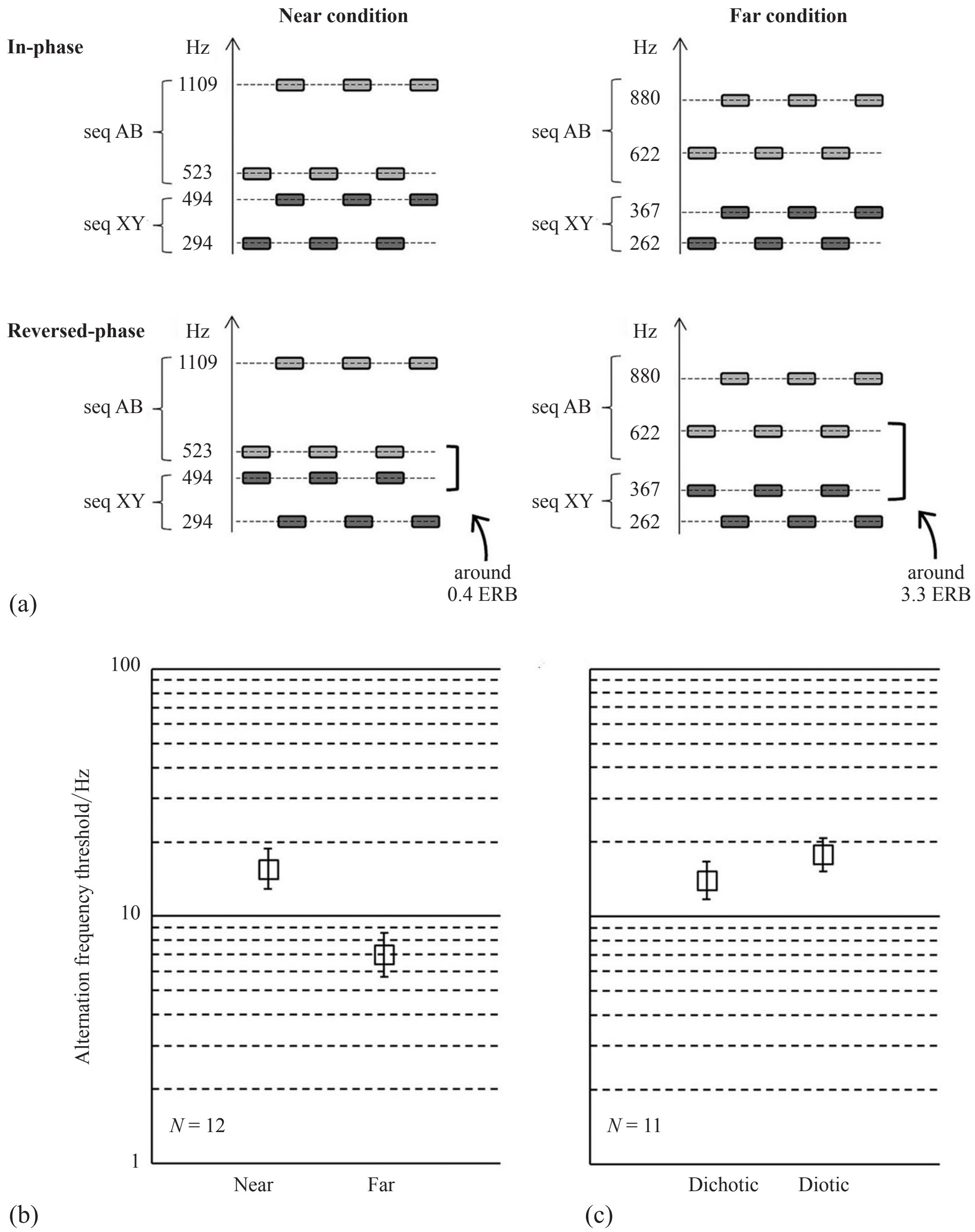

Figure 2. The temporal phase discrimination performance in experiments 1 and 2. (a) The auditory stimuli and experimental conditions of experiments 1 and 2. Sequence AB (seq AB) and sequence XY (seq XY) are two oscillating stimulus sequences used in each condition. For convenience, we named the phase relationship in which A is synchronized with X as 'in-phase.' The two frequencies presented simultaneously are separated irrespective of the condition. We named the phase relationship in which A is synchronized with Y as the 'reversed-phase'. Two frequencies presented simultaneously become very close intermittently in the near condition, whereas they are consistently well separated in the far condition. (b) The alternation frequency limits in the near and far conditions (experiment 1). (c) The alternation frequency limits in the dichotic and diotic conditions (experiment 2). Notes: ERB, equivalent rectangular bandwidth; SE, standard error; error bars = ±1 SE.

The critical alternation frequency beyond which participants could not make temporal phase discriminations was estimated by calculating the proportion of correct responses for each of six alternation frequencies, ranging from 1.40 to 8.00 Hz in half-octave steps. For participants whose proportion of correct answers at 8.00 Hz was above 75%, additional higher frequencies were used.(3) Participants performed the task for at least one ascending and one descending session for both conditions. For some participants, more sessions (three at most)(4) were added. The ear to which each sequence was presented was counterbalanced between sessions.

3.2 *Results and discussion*

The obtained psychometric function was fit by a logistic function using the maximum likelihood method to estimate the 75% correct point.(5) The fitting function used in the present study is defined by the following equation.

$$P = 0.5 \frac{\gamma}{1 + \exp[\alpha(\ln \chi - \ln \beta)]} + 0.5 \,,$$

where α denotes the slope, β the horizontal position, and γ the maximum proportion correct.

Ten out of twelve participants had a higher threshold (alternation frequency limit) in the near condition than in the far condition (figure 3).

Figure 2b shows the alternation frequency limits of the near and far conditions. The data are geometric averages over all participants.(6) The alternation frequency limit was around 16 Hz in the near condition and around 7 Hz in the far condition.

Relatively large individual differences were observed for the alternation frequency limits in both conditions, and there was a concern that the distribution was not normal. Therefore, we conducted a nonparametric statistical test. Wilcoxon's signed rank test indicated that the alternation frequency limit between the two conditions differed significantly ($Z = -2.32$, $p < 0.05$). The alternation frequency limit was significantly higher in the near condition than in the far condition.

However, for many participants the alternation frequency limits were replaced by the highest alternation frequency used in the experiment (26.67 Hz), because the proportion correct at this alternation frequency was above 75%. This caused a ceiling effect, and it became difficult to find a significant difference between conditions. Therefore, we added an analysis using the proportion of correct answers (with inverse sine transformation) at every alternation frequency. Specifically, we conducted a two-way repeated-measures ANOVA with the experimental conditions (far versus near) and alternation frequencies (1.4–8 Hz) as factors. The highest alternation frequency used in this analysis (ie 8 Hz) was the highest one which all participants experienced in both of the two conditions. There were main effects of experimental conditions ($F_{1,11} = 7.15$, $p < 0.05$) and alternation frequency ($F_{5,55} = 16.50$, $p < 0.01$). The results of multiple comparisons, using Bonferroni corrections, on the main effect of the alternation frequencies are shown in table 1. The interaction between experimental condition and alternation frequency was not significant. This result indicated

(3) Additional frequencies were selected for each participant in the following order of increasing frequency: 11.43, 16.00, 20.00, 26.67.

(4) For participants YO and ST1 data for additional alternation frequencies in the near condition (over 8 Hz) were collected in only one ascending or descending session.

(5) For participants MM, YO, AH, HM, KY, and TS the maximum alternation frequency used replaced the threshold frequency, since the estimated proportion of correct answers on the fitted psychometric function was not below 75 % within the range of the alternation frequencies used in the near condition.

(6) We used geometric means instead of arithmetic means because the alternation frequencies used in the present study were manipulated along a logarithmic axis.

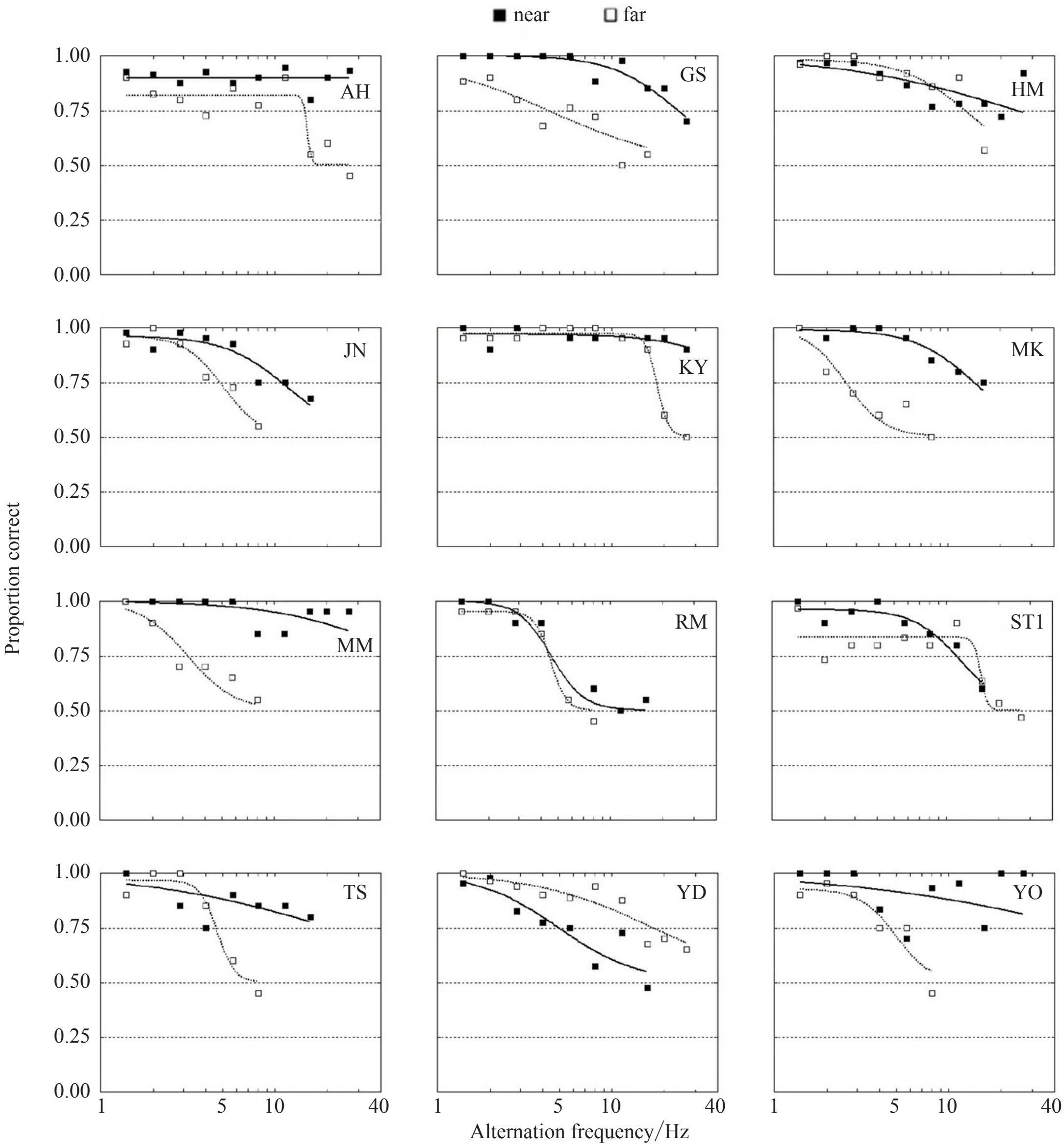

Figure 3. The proportion correct in experiment 1 for each participant as a function of alternation frequency. The symbols represent the proportion of correct answers in each condition, and the lines are fitted psychometric functions. Black symbols and solid lines indicate the near condition, while white symbols and dotted lines indicate the far condition.

Table 1. The results of multiple comparisons for the main effect of alternation frequencies in experiment 1. Statistically significant comparisons are shown in bold.

Alternation frequency/Hz						
	1.4	2	2.86	4	5.71	8
1.4	–	1.000	0.213	**0.005**	**0.004**	**0.001**
2	–	–	1.000	0.869	0.172	**0.018**
2.86	–	–	–	0.868	0.107	**0.004**
4	–	–	–	–	0.588	**0.003**
5.71	–	–	–	–	–	**0.033**
8	–	–	–	–	–	–

that the alternation frequency limit of temporal phase discrimination in audition increases as the frequency separation becomes small and decreases when it becomes large, as is the case for spatial separation in vision.

4 Experiment 2

The results of experiment 1 indicated that the alternation frequency limit was higher when the frequency separation between two auditory stimulus sequences was small. This suggests that the intersequence separation is a factor controlling (lowering) the alternation frequency limit of the temporal phase discrimination task not only in vision, but also in audition. To explore this issue further, we sought to examine whether channel separation would also affect the alternation frequency limit of feature binding.

The manipulation of frequency separation in experiment 1 was analogous to the manipulation of spatial separation in visual within-attribute binding. However, the two stimulus sequences in experiment 1 were presented to different ears. Therefore, in experiment 1 the most fundamental channels were separated. To make the situation more similar to a visual condition where two visual stimuli are respectively projected on two nearby positions on the retina, we have to present two auditory stimuli that are very close in frequency to a single ear (a single basilar membrane).

Experiment 2 compared conditions where two stimulus sequences were presented to either the same ear (diotically) or opposite ears (dichotically). If the channel separation lowers the alternation frequency limit, we could expect a higher limit for the diotic condition than for the dichotic one.

4.1 *Method*

The method was identical to that in experiment 1 except for the following aspects.

4.1.1 *Participants.* Participants were twelve university students aged 20–25 years (mean = 21.83 years, SD = 1.59 years) who had not participated in experiment 1.

4.1.2 *Stimuli.* There were two conditions: dichotic and diotic. Two auditory stimulus sequences identical to those in the near condition in experiment 1 were presented to each ear in the dichotic condition and to both ears simultaneously in the diotic condition. We chose the near condition, in which the closest frequency separation was 0.4 ERB. When presented diotically, the two signals interacted at the level of basilar membrane in a single auditory filter. The sound pressure level at each ear was 60 and 63 dB in the dichotic and the diotic conditions, respectively.

4.1.3 *Procedure.* The critical alternation frequency beyond which participants could not make temporal phase discriminations was estimated by the proportion of correct responses for eight alternation frequencies, ranging from 1.40 to 16.00 Hz in half-octave steps. For participants whose proportion of correct answers at 16.00 Hz was above 75%, additional higher alternation frequencies (up to 26.67 Hz) were used. Participants performed the task for a minimum of one ascending and one descending session for each condition. For some participants, more sessions (three at most) were added.[7]

4.2 *Results and discussion*

The data of one participant (NH) were eliminated from the analysis, as his performance in the dichotic condition was not stable. However, as this was an interesting case in which the task was possible in the diotic, but not the dichotic condition, the data are shown in figure 4.

[7] For participant TI one more ascending session was conducted with additional alternation frequencies (over 16 Hz) in the dichotic condition.

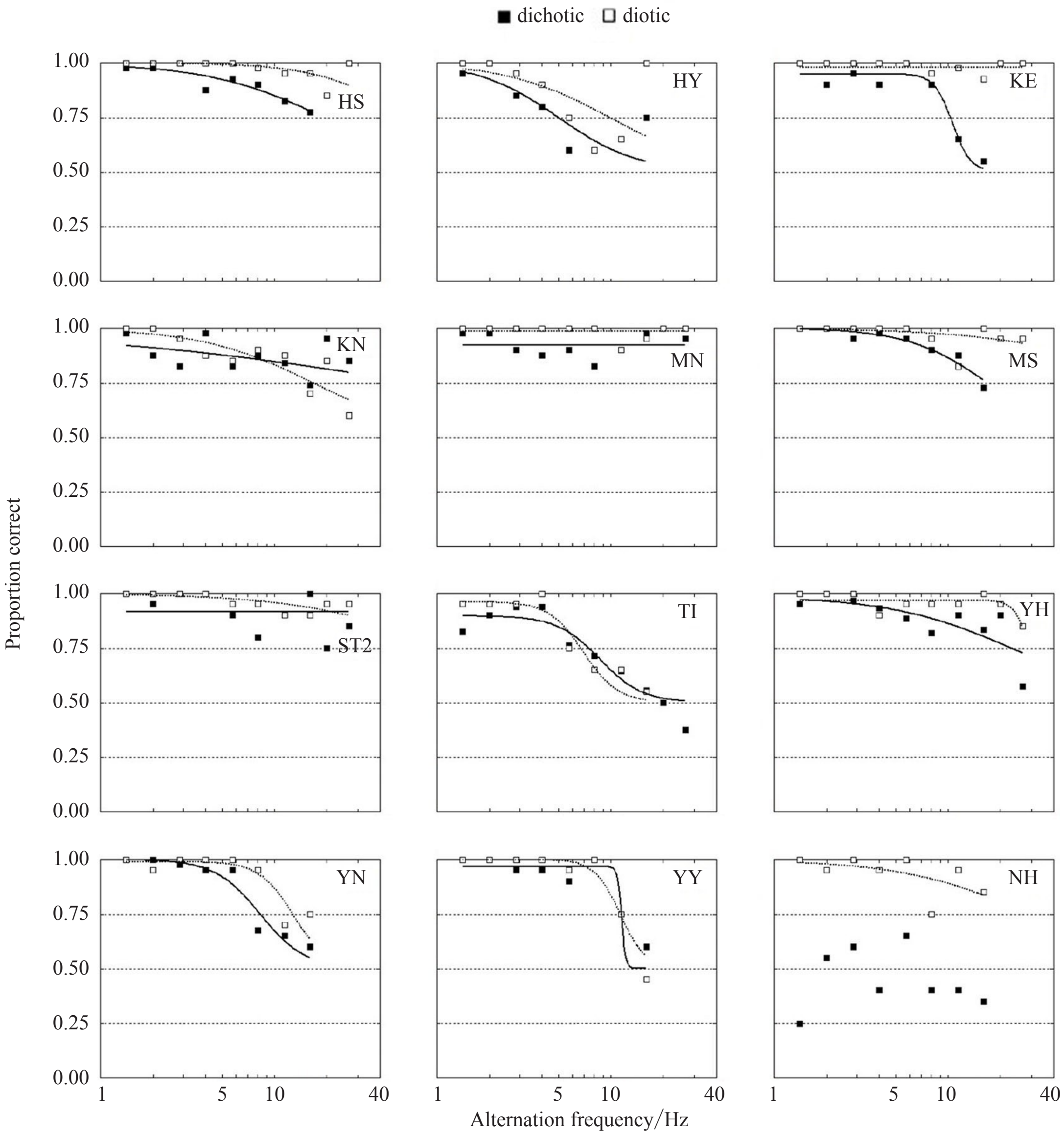

Figure 4. The proportion correct in experiment 2 for each participant as a function of alternation frequency. The symbols represent the proportion of correct answers in each condition, and the lines are fitted psychometric functions. Black symbols and solid lines indicate the dichotic condition, while white symbols and dotted lines indicate the diotic condition. The data of participant NH were excluded from the analysis.

Because the dichotic condition in the present experiment was the same as the near condition of experiment 1, we compared the results of these two conditions. As in experiment 1, we conducted a Wilcoxon's signed rank test. There were no significant differences between the two experiments ($Z = -0.80, p = 0.42$), suggesting that similar limits were observed, regardless of the different participants.

The results of the dichotic and diotic conditions are shown in figure 4.(8) For six of eleven participants, excluding participant NH, the thresholds were higher in the diotic than in the

(8) For participants KN, MN, KE, YH, HS, MS, and ST2 the maximum alternation frequency used replaced the threshold frequency because the estimated proportion of correct answers on the fitted psychometric function was not below 75% of the range in the alternation frequencies used in one of the conditions.

dichotic condition. Figure 2c shows the alternation frequency limits of these conditions (geometric averages over eleven participants). The alternation frequency limit was around 18 Hz in the diotic and around 14 Hz in the dichotic condition. A Wilcoxon's signed rank test showed that the alternation frequency limit between the two conditions differed marginally significantly ($Z = -1.72$, $p = 0.09$). The alternation frequency limit was significantly higher in the diotic than in dichotic condition.

As in experiment 1, we conducted a two-way repeated-measures ANOVA with the experimental conditions (dichotic versus diotic) and alternation frequency (1.4–16 Hz) as factors. The highest alternation frequency used in this analysis was the highest one at which participants underwent in both conditions (16 Hz). As a result, there were main effects of the experimental condition ($F_{1,10} = 39.05$, $p < 0.01$) and alternation frequency ($F_{7,70} = 15.09$, $p < 0.01$). The results of multiple comparisons on the main effect of the alternation frequencies, with Bonferroni correction, are shown in table 2. There were no interactions between experimental condition and alternation frequency.

Table 2. The results of multiple comparisons for the main effect of alternation frequencies in experiment 2. Statistically significant comparisons are shown in bold.

Alternation frequency/Hz								
	1.4	2	2.86	4	5.71	8	11.43	16
1.4	–	1.000	1.000	0.610	**0.035**	**0.004**	**0.000**	**0.010**
2	–	–	1.000	1.000	0.432	**0.026**	**0.000**	**0.006**
2.86	–	–	–	1.000	0.346	0.060	**0.000**	0.051
4	–	–	–	–	1.000	0.259	**0.009**	0.292
5.71	–	–	–	–	–	1.000	0.177	1.000
8	–	–	–	–	–	–	1.000	1.000
11.43	–	–	–	–	–	–	–	1.000
16	–	–	–	–	–	–	–	–

Participant NH, whose data are shown in figure 4, could perform the task well in the diotic but not the dichotic condition. This may suggest that some peripheral cues were available in the diotic condition that were not available in the dichotic condition, and participant NH could not perform the task without these cues. A previous study (Strickland et al., 1989) reported similar individual differences; some participants could successfully detect the phase disparity between envelopes of two amplitude-modulated signals presented diotically, but not dichotically, in some conditions.

5 Subsidiary experiment

In experiment 1 when the two stimulus sequences were close in frequency (the near condition), the frequency separation between A and B (or X and Y) was decreased as compared with the far condition. Therefore, it is not clear whether the effects are due to the frequencies to be bound being close or the frequencies to be segregated being far. To address this issue, we conducted a subsidiary experiment in which two of the authors (SK and WF) participated. The conditions of the subsidiary experiment are shown in figure 5. In the subsidiary experiment AB and XY sequences were both alternations of two pure tones separated by 4 semitones (within-sequence distance). The tones were combined to make three intersequence distance conditions (1 semitone condition, 7 semitone condition, and 13 semitone condition). Within-sequence distance was fixed to 4 semitones and did not vary depending on the condition. The frequency separation between 1 semitone and 7 semitone (6 semitones) was equal to the frequency separation between

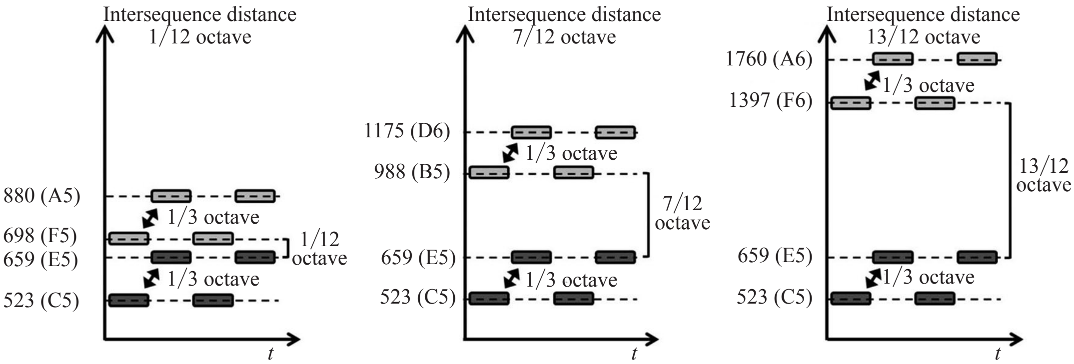

Figure 5. Stimulus configuration of a subsidiary experiment. There were three intersequence distance conditions (1 semitone condition, 7 semitone condition, and 13 semitone condition). Within-sequence distance was fixed at 4 semitones and did not vary depending on the condition.

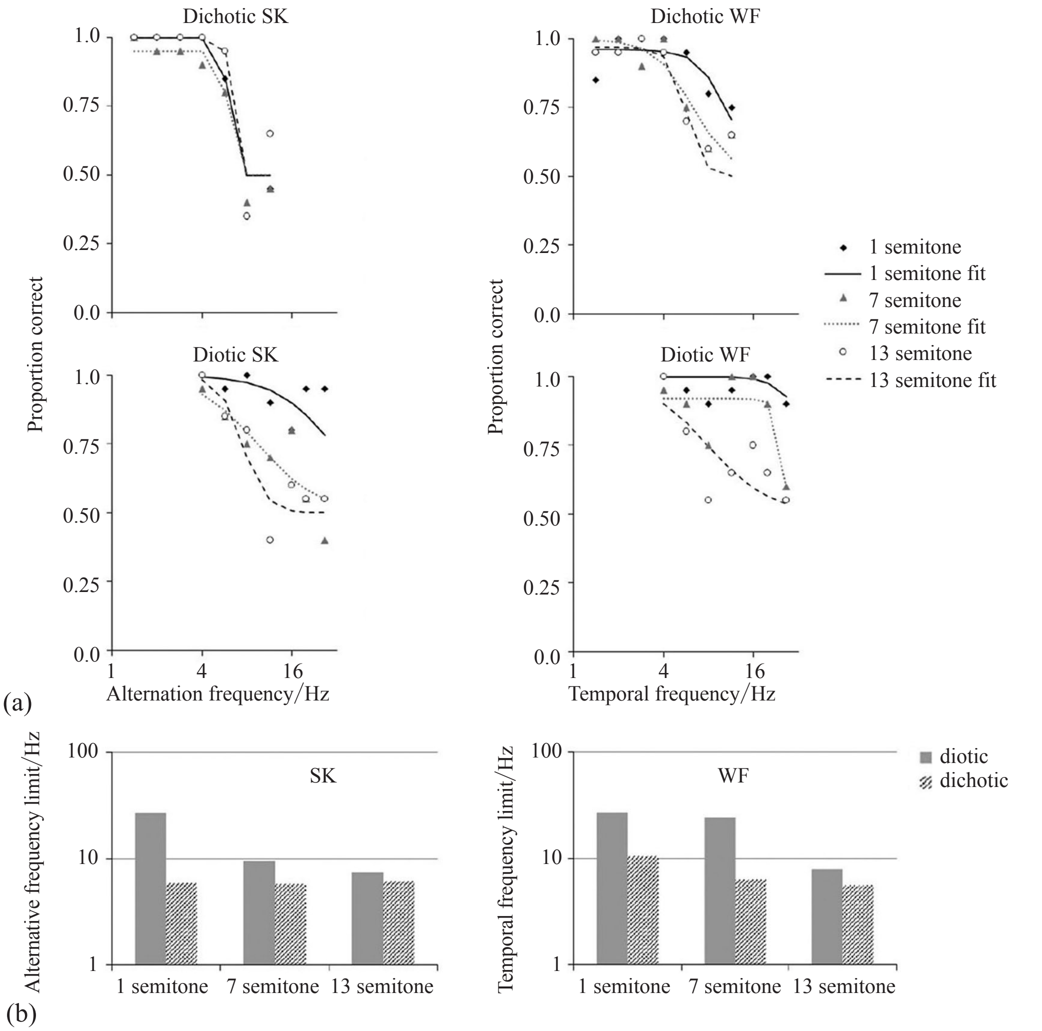

Figure 6. Results of a subsidiary experiment performed by two of the authors. Two stimulus sequences were presented either dichotially or diotically. (a) The proportion correct for each participant as a function of alternation frequency. The symbols represent the proportion of correct answers in each condition, and the lines are fitted psychometric functions. (b) The alternation frequency limits in each condition. For both participants the maximum alternation frequency (26.67 Hz) replaced the threshold frequency because the estimated proportion of correct answers on the fitted psychometric function was above 75% of the range in the alternation frequencies used in the 1 semitone condition with the diotic presentation.

7 semitones and 13 semitones (6 semitones). These stimulus parameters were optimized so that none of the frequencies shared the same musical note names. The intersequence distances were systematically manipulated while keeping the within-sequence distance fixed. In the subsidiary experiment the whole frequency band was shifted from the frequency bands used in experiments 1 and 2 to the upper region so that the single tones would consist of relatively large numbers of cycles. Specifically, the lowest pure-tone frequency was 523.25 Hz (C5). We set the duration of cosine ramps which modulated the lead and the tail edges of the square wave to 5 ms, which was the half length of that in experiments 1 and 2.

The results are shown in figure 6. When the stimuli were presented diotically, the effect of intersequence frequency separation was evident. In the 1/12 octave condition where the intersequence distance was smallest, the alternation frequency limits were replaced by the highest alternation frequency (26.67 Hz), because the proportion of correct answers at this alternation frequency was above 75%. When the stimuli were presented dichotically, the alternation frequency limit in the 1/12 octave condition was higher than in the other conditions at least in one participant (WF), while the effect of intersequence distance was not as clear as in diotic presentation. These results suggest that the alternation frequency limits are dependent on the intersequence distance even when the within-sequence distance is fixed.

6 General discussion

We measured the limits of temporal phase discrimination between two alternating auditory stimulus sequences while changing the distance between two stimulus sequences in two experiments. The alternation frequency limits were high when the frequency separation was small (experiment 1) and even higher for the diotic presentation (experiment 2). Similar to vision, temporal binding became more difficult in audition as the intersequence distance increased.

We also conducted an analysis using the proportion of correct answers at every alternation frequency. Specifically, we conducted a two-way repeated-measures ANOVA with the experimental condition and alternation frequency as factors. The results suggested that the proportions of correct answers decreased as the intersequence distance increased. Here, the lack of significant interactions in the two-way ANOVA in both experiments is important, and merits discussion, because it apparently contradicts the general trend that participants can easily perform the temporal phase discrimination task at the lowest alternation frequency, regardless of condition. Individual data (figures 3 and 4) show that the percent correct was nearly 100% at the lowest alternation frequency, regardless of condition, but varied at higher alternation frequencies depending on the condition (near and far in experiment 1; dichotic and diotic in experiment 2). However, some participants did not show this trend. For example, in figure 3 AH, GS, and ST1 showed a higher percent correct for the near condition, while many other participants showed equally high percent correct for both conditions at the lowest alternation frequency. This may be the reason why the statistical tests on the data of all participants did not show any significant interactions. Currently, it is not clear why the percent correct differed for some participants according to condition, even at the lowest alternation frequency. One possible explanation is that various cues were available, and task performance was dependent on the cue used by each participant. For example, when a participant relies on a low-level cue (such as one that is available only before the binaural processing stage), the task can be performed more easily for the diotic condition than for the dichotic condition, even at the lowest alternation frequency.

The present study demonstrated that distance (whether it be frequency separation or channel separation) is a critical factor in the alternation frequency limits of the temporal phase discrimination task. Here, we discuss the results of this study in terms of whether two signals interact in a single auditory filter or not. Auditory filters are part of a widely accepted model describing how frequency analysis functions in the peripheral auditory system. This frequency analysis is thought to be mediated by a filter bank composed of many bandpass filters—auditory filters—with different center frequencies (Moore, 2003). One ERB represents the frequency width of this filter, and ranges from 11% to 17% of the center frequency (Patterson, 1976). When two signals are input to independent auditory filters, they are resolved. In contrast, when they are input to a single auditory filter, they are not resolved. In experiment 2 of the present paper the intersequence distance was around 0.4 ERB. Thus, in the diotic condition two stimulus sequences are input to a single auditory filter and not resolved. This may cause low-level cues such as patterns of amplitude modulation at a single auditory filter to contribute to the better performance in the diotic condition than in the dichotic condition. Despite using slightly different tasks, previous studies have reported an effect of frequency separation consistent with the present findings. For example, Richards (1987) investigated the ability to discriminate between simultaneously presented 100 Hz wideband noise with envelopes that were either similar or dissimilar, and found that discrimination was quite difficult when the center frequencies of the noise were widely separated. Yost and Sheft (1989) showed that the detection of phase disparity between two amplitude modulation waves imposed on two carriers becomes more difficult as the frequency separation increases. Strickland (1989) also reported similar trends with a higher modulation frequency.

The present study is also broadly consistent with the so-called 'auditory stream segregation' phenomenon, which has been widely reported in the audition literature (Bregman & Campbell, 1971; Bregman & Pinker, 1978; Dannenbring & Bregman, 1978; Moore & Gockel, 2002; van Noorden, 1975). In a typical auditory stream segregation experiment two tones that differ in frequency (A and B) are presented alternately in a sequence (eg ABA_ABA_ABA …; where '_' represents a silent period). Listeners tend to report either the 'one-stream' percept, in which the A and B tones are grouped to form a single coherent stream, or the 'two-stream' percept, in which the A and B tones are independently grouped to form two parallel streams. When the listener perceives two streams, judging of the temporal relationship between the A and B tones is known to be difficult (eg Bregman & Campbell, 1971; Micheyl, Hunter, & Oxenham, 2010; Warren, Obusek, & Farmer, 1969), implying the failure of perceptual binding (or comparisons) across sensory channels (ie frequency channels in this case) by the auditory system. The probability of the two-stream percept tends to increase with increasing frequency separation between the A and B tones and with increasing speed of alternation. This tendency is consistent with the present results, and thus implies that the paradigm adopted in the present study can tap the mechanisms underlying auditory stream segregation.

The present study demonstrated that the temporal phase discrimination task, which has been utilized in vision and cross-modal studies, is also applicable to audition. Because the same tasks were utilized, the results of the present study of audition can be directly compared with those of previous studies that studied the visual and cross-sensory modalities. The alternation frequency limits of temporal phase discrimination tasks are thought to reflect the temporal characteristics of the mechanisms involved in binding at various stages in the perceptual system. Therefore, with the use of this task, we can investigate the temporal characteristics of various mechanisms mediating visual, auditory, and cross-modal bindings at the same stage. We believe this is the most important contribution of the present paper.

Figure 7 summarizes the alternation frequency limits of the present and previous studies. Previous reports on the alternation frequency limits of visual temporal phase discrimination tasks have revealed that the alternation frequency limits are high when two stimulus sequences have the same attribute (eg luminance–luminance) and are spatially close (20 Hz or more), while these limits are relatively low when they are spatially separated (under 10 Hz) (Aghdaee & Cavanagh, 2007; Forte et al., 1999; Fujisaki & Nishida, 2010). Treisman (1996)

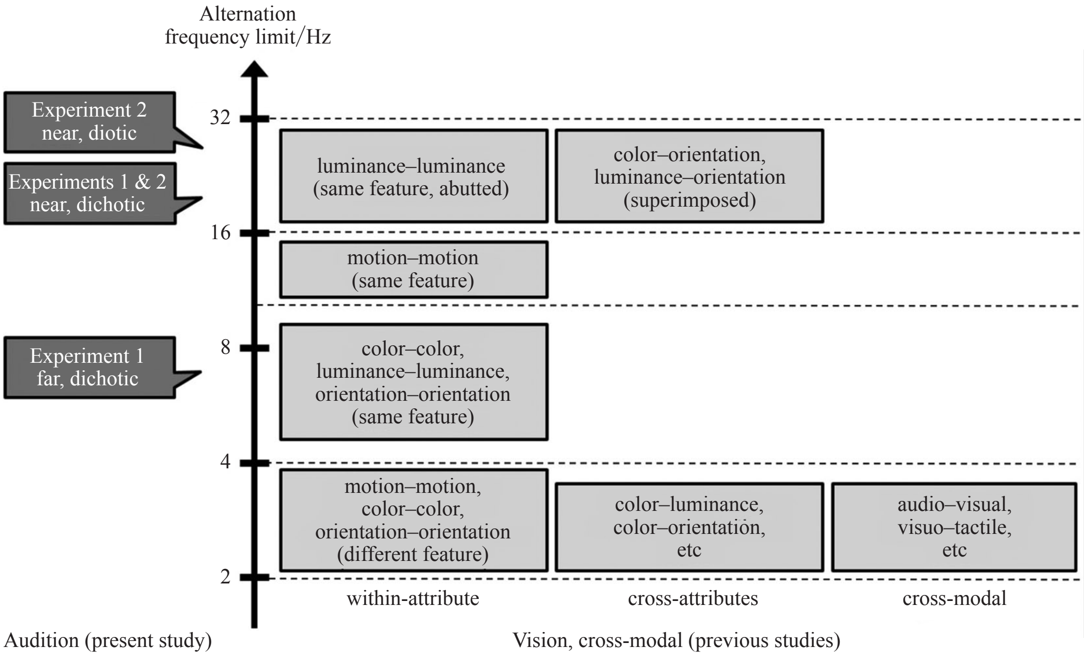

Figure 7. The alternation frequency limits of the present study compared with those of previous studies (Aghdaee & Cavanagh, 2007; Amano et al., 2007; Arnold, 2005; Bartels & Zeki, 2006; Clifford et al., 2004; Forte et al., 1999; Fujisaki & Nishida, 2010; Holcombe & Cavanagh, 2001; Maruya et al., 2013; Moutoussis & Zeki, 1997; Nishida & Johnston, 2004; Victor & Conte, 2002).

has pointed out that binding can be mediated by various mechanisms, including early or automatic mechanisms and late mechanisms that require attention. In previous studies the fast temporal phase discrimination between two visual stimulus sequences within a short distance was thought to be mediated by a low-level mechanism that detects various cues derived from interactions between stimuli. For example, when two luminance-modulated visual stimuli are very close in space, one can discriminate whether the phases of luminance modulation are matched or not even at quite high modulation frequencies (20–30 Hz) (Aghdaee & Cavanagh, 2007; Forte et al., 1999; Holcombe & Cavanagh, 2001; Victor & Conte, 2002). In this case, cues such as subjective contours at the counterphasing edges have been reported to be available. These cues are thought to be processed at relatively early stages in the perceptual system. However, the slow temporal phase discrimination between visual stimulus sequences over a long distance may be mediated by higher level mechanisms (Aghdaee & Cavanagh, 2007; Forte et al., 1999; Holcombe & Cavanagh, 2001; Victor & Conte, 2002). For example, for cross-modal binding (eg vision–audition, vision–touch, and audition–touch) where the intersequence distance would be largest, very low alternation frequency limits are observed (Fujisaki & Nishida, 2010). Indeed, binding across different sensory modalities is thought to be mediated by attention-demanding late processes (Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006; Fujisaki & Nishida, 2007, 2008). Similar to the findings for vision, the present study demonstrated that temporal binding in audition became difficult as the intersequence distance increased. This suggests that an underlying binding principle may be common to vision and audition.[9]

The data illustrated in figure 7 show that the alternation frequency limit decreased to around 2–3 Hz when two stimuli have different attributes (eg color–luminance, color–orientation) or modalities (eg vision–audition, vision–touch, audition–touch; Amano et al., 2007; Arnold, 2005; Bartels & Zeki, 2006; Fujisaki & Nishida, 2010). It is important to note that if features are different within the same attribute (such as a red–green or a blue–yellow alternation of a color attribute), or if alternations of the same attribute are both orientation changes (such as different tilts), the alternation frequency limits drop to around 2–3 Hz, similar to the alternation frequency limits for different attributes and modalities (Bartels & Zeki, 2006; Fujisaki & Nishida, 2010; Maruya et al., 2013). All stimuli used in the present study were tone bursts, and the perceptual and physical distances were defined by manipulation of the fundamental frequencies and whether the presentation was dichotic or diotic. That is, we asked participants to bind two tone bursts. In vision, this is similar to binding luminance with luminance or color with color. Therefore, the comparison was between the same attribute at various distances. The alternation frequency limits obtained with the present study were around 7 Hz and 14–16 Hz for the far and near conditions, respectively, in experiment 1, and around 18 Hz for the diotic condition in experiment 2. In general, this pattern of results is similar to that found in vision studies using the same attributes. However, the alternation frequency limits were slightly higher for audition than for vision, and individual differences were large, especially in the diotic condition of experiment 2. For example, the alternation frequency limits reached more than 26.67 Hz for some participants, and one participant could not perform the task in the dichotic condition, but performed the task well in the diotic condition (see figure 4). The large individual differences may suggest that multiple cues from peripheral to mid- or high-level audition were available and different participants utilized different cues. The different alternation frequency limits obtained here may reflect the several different mechanisms at work in auditory temporal binding.

The present study utilized pure tones with different frequencies (and different musical note names). This is analogous to the same attribute with different feature conditions in vision. In the case of vision, the alternation frequency limits drops to around 2–3 Hz when two features are different, but this was around 7 Hz in the case of audition (the far condition of experiment 1 in the present study), which seems higher than vision.

The reason why different alternation limits were observed between vision and audition is not clear at present, although one possible interpretation is that vision and audition are not exactly parallel, as follows. In vision, the information from each eye is represented independently until it reaches V1. In contrast, in audition, the information from each ear reaches the auditory cortex through the cochlea, the cochlear nucleus (first relay point) in the brainstem, the superior olivary complex (second relay point) in the brainstem, and several other structures. In this pathway the information from each ear is integrated in the superior olivary complex. Moreover, there is a feedback pathway from the superior olivary complex to the cochlear nucleus. Thus, the information from each ear is integrated at a relatively early stage in the auditory pathway. However, the sound frequency analysis is first performed by the basilar membrane or the auditory nerve fibers, and the information from each ear is independent at this stage.

[9] The auditory cortex and the peripheral structures have selectivity for different temporal frequencies, and the temporal resolution of peripheral neurons is higher than that of cortical ones (Chi, Ru, & Shamma, 2005; Joris, Schreiner, & Rees, 2004). This seems to suggest a similarity to vision—namely, that cortical neurons are more sluggish than subcortical. However, in the auditory system time-varying signals can be represented in multiple ways (Lu, Liang, & Wang, 2001). Therefore, to directly compare audition and vision is difficult.

The present study did not utilize different auditory attributes besides pure tone frequencies. Although it is difficult to define auditory attributes analogous to visual attributes (eg color, orientation, motion), investigating the alternation frequency limits of auditory binding between different attributes would be of interest.

Acknowledgments. SK is supported by Grant-in-Aid for JSPS Fellows 11J08278, DC1. WF and KY are supported in part by JSPS Grants-in-Aid for Scientific Research 25380997 and 23119707, respectively.

References

Aghdaee, S. M., & Cavanagh, P. (2007). Temporal limits of long-range phase discrimination across the visual field. *Vision Research*, **47**, 2156–2163.

Amano, K., Johnston, A., & Nishida, S. (2007). Two mechanisms underlying the effect of angle of motion direction change on colour–motion asynchrony. *Vision Research*, **47**, 687–705.

Arnold, D. H. (2005). Perceptual pairing of colour and motion. *Vision Research*, **45**, 3015–3026.

Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research*, **46**, 2280–2286.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244–249.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, **32**, 19–31.

Carlyon, R. P. (1991). Discriminating between coherent and incoherent frequency-modulation of complex tones. *The Journal of the Acoustical Society of America*, **89**, 329–340.

Carlyon, R. P. (1994). Further evidence against an across-frequency mechanism specific to the detection of frequency-modulation (FM) incoherence between resolved frequency components. *The Journal of the Acoustical Society of America*, **95**, 949–961.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, **118**, 887–906.

Clifford, C. W. G., Holcombe, A. O., & Pearson, J. (2004). Rapid global form binding with loss of associated colors. *Journal of Vision*, **4**(12):8, 1090–1101.

Dannenbring, G. L., & Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, **24**, 369–376.

Forte, J., Hogben, J. H., & Ross, J. (1999). Spatial limitations of temporal segmentation. *Vision Research*, **39**, 4052–4061.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **273**(1588), 865–874.

Fujisaki, W., & Nishida, S. (2007). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Research*, **47**, 1075–1093.

Fujisaki, W., & Nishida, S. (2008). Top-down feature-based selection of matching features for audio-visual synchrony discrimination. *Neuroscience Letters*, **433**, 225–230.

Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **277**(1692), 2281–2290.

Furukawa, S., & Moore, B. C. J. (1996). Across-channel processes in frequency modulation detection. *The Journal of the Acoustical Society of America*, **100**, 2299–2311.

Furukawa, S., & Moore, B. C. J. (1997). Dependence of frequency modulation detection on frequency modulation coherence across carriers: Effects of modulation rate, harmonicity, and roving of the carrier frequencies. *The Journal of the Acoustical Society of America*, **101**, 1632–1643.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, **47**, 103–138.

Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, **13**, 216–221.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, **4**, 127–128.

Holcombe, A. O., & Judson, J. (2007). Visual binding of English and Chinese word parts is limited to low temporal frequencies. *Perception*, **36**, 49–74.
Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, **84**, 541–578.
Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience*, **4**, 1131–1138.
Maruya, K., Holcombe, A. O., & Nishida, S. (2013). Rapid encoding of relationships between spatially remote motion signals. *Journal of Vision*, **13**(2):4, 1–20.
Micheyl, C., Hunter, C., & Oxenham, A. J. (2010). Auditory stream segregation and the perception of across-frequency synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, **36**, 1029–1039.
Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.). San Diego, CA: Academic Press.
Moore, B. C., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica united with Acustica*, **88**, 320–333.
Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **264**, 393–399.
Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology*, **12**, 359–368.
Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, **59**, 640–654.
Richards, V. M. (1987). Monaural envelope correlation perception. *The Journal of the Acoustical Society of America*, **82**, 1621–1630.
Strickland, E. A., Viemeister, N. F., Fantini, D. A., & Garrison, M. A. (1989). Within- versus cross-channel mechanisms in detection of envelope phase disparity. *The Journal of the Acoustical Society of America*, **86**, 2160–2166.
Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, **6**, 171–178.
van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.
Victor, J. D., & Conte, M. M. (2002). Temporal phase discrimination depends critically on separation. *Vision Research*, **42**, 2063–2071.
Warren, R. M., Obusek, C. J., & Farmer, R. M. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, **164**(3879), 586–587.
Yost, W. A., & Sheft, S. (1989). Across-critical-band processing of amplitude-modulated tones. *The Journal of the Acoustical Society of America*, **85**, 848–857.