

**Auditory Kappa Effects Can Be Explained by Perceptual Grouping Based on Feature  
Similarity**

Carolyn Kroger, Deborah R. Fu, Renee M. Banakis Hartl, Anahita H. Mehta

Department of Otolaryngology – Head and Neck Surgery, University of Michigan

Address correspondence:

Carolyn Kroger, Ph.D.

Kresge Hearing Research Institute

Department of Otolaryngology – Head and Neck Surgery

University of Michigan

Ann Arbor, MI 48109

Email: [ckroger@med.umich.edu](mailto:ckroger@med.umich.edu)

ORCID: 0000-0002-2331-0963

Keywords: kappa effect, time perception, perceptual interactions, auditory grouping, auditory motion

### Abstract

Pitch, timing, and space in auditory perception interact, causing changes in one dimension to distort judgments of others (Boltz, 1998; Henry et al., 2009; Jones et al., 2006). The auditory kappa effect exemplifies this, as larger pitch changes between sequential tones create an illusion of longer time intervals (Cohen et al., 1954). Traditionally studied with three-tone sequences ('A-X-B'), this effect shows that varying the middle tone's pitch relative to the first and third tones skews time interval perception. The prevailing *auditory motion hypothesis* (Henry & McAuley, 2009) proposes that a "pitch velocity" is inferred from a consistent pitch trajectory. We tested this against the *auditory grouping hypothesis*, which proposes that feature similarity affects perceived timing. Experiment 1 tested the kappa effect under varying pitch separations, revealing slower pitch velocities produced larger effects, contradicting the motion hypothesis. Experiment 2 examined inconsistent pitch trajectories to challenge the stability of pitch velocity referents, and still found kappa effects, further supporting the auditory grouping hypothesis. Experiment 3 investigated the kappa effect in auditory space, confirming that sounds presented closer in space are judged as occurring closer in time. These findings suggest that the kappa effect results from the grouping of auditory events by feature similarity, opposing the auditory motion hypothesis.

## Introduction

Making sense of complex auditory environments requires the integration of various perceptual features such as pitch, spatial location, and onset timing into coherent objects and events (Bizley & Cohen, 2013). Such theoretically dissociable acoustic cues produce reliable interactions in perception, as in the case of auditory *kappa* and *tau* effects, where greater pitch separation between successive tones leads to longer reported time intervals between them and longer time intervals between tones leads to greater reported pitch separation, respectively (Cohen et al., 1954; Henry & McAuley, 2009; Henry et al., 2009). In classic kappa tasks, two intervals are defined by three discrete events (sometimes termed ‘A-X-B’) and participants compare the duration of the first and second interval, reporting which interval is relatively longer or shorter (between ‘A-X’ or between ‘X-B’). Kappa effects have been shown for visual, auditory, and tactile stimuli, and represent a general temporal bias where events separated by greater distance (e.g., in physical space or other feature space) are judged to be separated by longer time intervals (Cohen et al., 1953; Cohen et al., 1954; Suto, 1952).

The prevailing explanation for the kappa effect is an imputed velocity model, based on the theory of ‘phenomenological motion’ that links successive discrete events across space and time, creating the percept of unified motion (Price-Williams, 1954). The foundational kappa study in vision presented sequences of three flashes of light with varying distances between them, and participants reported longer durations between more spatially-distant flashes (Cohen et al., 1953; Huang & Jones, 1982). Spatial manipulations were performed by keeping the distance between the first and third flash (‘A’ and ‘B’) constant and changing the location of the second (‘X’) flash to be closer to either to A or B. According to Jones and Huang (1982), participants impute uniform motion across sequences with an assumption of constant velocity, and thus judge

the farther X distance (from A or B) as taking longer to traverse than the shorter one, leading to longer duration judgments for the greater distance.

The imputed velocity model was later translated from the visuospatial to auditory pitch domain based on an *auditory motion hypothesis*, which proposes that sequences of pitches have motion-like properties. This model was applied to the auditory kappa effect by computing a “pitch velocity” (change in pitch over time) for tone sequences (Henry & McAuley, 2009). In these experiments, “pitch space” is an auditory analog to the visuospatial manipulations studied previously. Henry and McAuley (2009) found that auditory kappa effects for tone sequences could be predicted by an imputed pitch velocity model that related the objective ( $t$ ) and expected ( $E(t)$ ) time intervals based on average pitch velocity between A and B,  $V = \Delta p / \Delta t$ , with a free weighting parameter,  $w$ . If this weighting parameter,  $w$  is equal to 1, perceived timing is based solely on the objective timing  $t$ , and no kappa effect is observed. If  $w = 0$ , perceived timing is based solely on the expected timing, where  $E(t) = \Delta p / V$ . Thus, the value of  $w$  represents the extent of response bias in temporal judgments based on imputed pitch velocity,  $V$ , and negatively correlates with the magnitude of the kappa effect. Their model equation is thus:

$$R = f[w t + (1 - w)E(t)] \quad (1)$$

Where  $R$  is a response probability (i.e., “shorter” vs “longer” time interval) as a function of the relative weighting of objective and expected event timing. In their study, Henry and McAuley (2009) found support for the *auditory motion hypothesis* by demonstrating smaller  $w$  values in their model for higher-velocity sequences, indicating stronger kappa effects for faster-moving pitch which is proposed to induce stronger motion percepts. Other support comes from studies showing that properties of perceptual motion apply to pitch sequences, such as the use of pitch-time trajectory to predict when and where a future melodic note will occur, as demonstrated by

pitch-based “representational momentum” effects (Johnston & Jones, 2006), and the presence of “inertia” in musical composition (for a review, see: Hubbard, 2017). There is also some linguistic evidence, such as the use in Western European languages of spatialized conceptual and linguistic metaphors including “rising” or “falling” to ascribe motion qualities such as direction of change in pitch over time (for a review, see: Eitan, 2013).

A notable shortcoming of this hypothesis, however, is the failure to adequately account for auditory kappa effects in non-pitch features such as spectral centroid and speech vowel quality that are not known to induce auditory motion percepts. Marty et al. (2019) manipulated both the pitch and brightness (spectral centroid) of tones in an A-X-B kappa paradigm. They found kappa effects for sequences of tones changing only in brightness with fixed pitch (fundamental frequency). Further, they found that the magnitude of the brightness-only kappa effect was not significantly different from sequences with both features changing together (i.e., increasing in both pitch and brightness) and the pitch-only kappa effect was significantly smaller than the other two conditions. This study indicated that relative brightness was a stronger cue for kappa effects than pitch, contrary to predictions of the *auditory motion hypothesis*.

In another study, Shigeno (1986) investigated kappa effects for A-X-B sequences of vowel sounds, where A and B were brief synthesized /i/ and /e/ sounds, and X took on a range of intermediate sounds that did not belong to either phonetic category. Importantly, each possible target X sound was manipulated to be more similar to either the /i/ or /e/ sound. Results showed an auditory kappa effect where participants reported the presentation of X to be closer in time to the vowel sound that was more similar based on the first three formant frequencies. These experiments call into question the *auditory motion hypothesis* as an explanation for auditory

kappa effects, as spectral changes associated with timbral brightness or vowel quality are examples of cues for category boundaries rather than auditory motion.

An alternative explanation for the kappa effect which can account for the presence of this effect in experiments involving pitch features or category boundary cues is an *auditory grouping hypothesis*. This hypothesis proposes that two auditory stimuli more similar along a given feature dimension are perceptually grouped as part of an auditory object and thus judged as occurring closer in time than a third, more distinct stimulus. A well-known effect of changing the relative frequency and timing of sequential tones is “auditory streaming” whereby successive tones can be heard as belonging to one coherent stream or object, or to separate streams (Bregman & Campbell, 1971; Moore & Gockel, 2002; van Noorden, 1975). Sequential auditory stimuli are perceptually grouped as part of a stream based on acoustic feature similarity such as pitch, spectral shape or timbre, spatial location, or temporal proximity, where sounds that are more similar on a given dimension are more likely to belong to the same stream (for comprehensive reviews see: Bizley & Cohen, 2013; Moore & Gockel, 2012; Snyder & Alain, 2007). Additionally, the grouping of auditory stimuli into streams influences the perception of temporal relations among stream elements, with judgements of temporal gaps in stimuli across groups or streams being overestimated in duration relative to those within a group (Fitzgibbons et al., 1974; Micheyl et al., 2010; Thorpe & Trehub, 1989).

Considering the stimuli traditionally used to examine the auditory kappa effect, where participants are required to make temporal judgments about sequential tones with varying features in both temporal and non-temporal domains, certain time intervals may be difficult to discriminate while other features (e.g. pitch) are more easily discriminable. When temporal grouping cues are ambiguous, as in the case of nearly-equivalent inter-onset intervals between

events, the listener may perceive auditory stimuli that share other feature similarities as belonging to one stream and more dissimilar events as belonging to a separate stream. In previous studies on auditory kappa effects, stimuli that share more feature similarity were judged to be closer in time than those that were more distinct. Beyond frequency and spectral content, kappa effects have been demonstrated for variations in sound intensity (loudness) and sounds presented sequentially to the same or different ear (Alards-Tomalín et al., 2013; Bausenhardt & Quinn, 2018). These studies manipulated acoustic features that could serve as cues for auditory grouping/streaming, suggesting that a general perceptual mechanism, rather than a motion-based process underlies auditory kappa effects.

The present study proposes a series of experiments to test the *auditory motion hypothesis* against the *auditory grouping hypothesis* as a parsimonious explanation for auditory kappa effects, as well as extending the auditory kappa effect to sequences moving through auditory space. The first experiment revisits the pitch-velocity effects found by Henry and McAuley (2009) by testing for differences in magnitude of auditory kappa effects across velocity conditions created by changing the overall pitch separation between tones while keeping the overall temporal separation constant. In their study, Henry and McAuley (2009) manipulated pitch velocity by increasing or decreasing the inter-onset interval (IOI) between bounding tones (A and B), such that the total pitch separation between A and B was constant but sequence duration ( $IOI_T$ ) took on one of three values (728 ms, 1000 ms, or 1600 ms). For sequences with longer IOIs between bounding tones ( $IOI_T$ ), the pitch velocity was slower based on  $V = \Delta p / IOI_T$ . They interpreted larger kappa effects in faster pitch velocity conditions as evidence for stronger motion percepts according to the *auditory motion hypothesis*. However, this velocity effect cannot be disentangled from an overall temporal separation effect, as velocity and  $IOI_T$  were

perfectly correlated. We followed up on this experiment by comparing the magnitude of kappa effects across pitch velocity conditions created in the opposite manner, by holding  $\text{IOI}_T$  constant across four different pitch separation conditions ( $\Delta p$ ). The *auditory motion hypothesis* predicts that kappa effects will be stronger for sequences with greater pitch separation between A and B (faster pitch velocity) than sequences with smaller overall pitch separation. However, the *auditory grouping hypothesis* predicts stronger kappa effects for sequences with smaller pitch separations due to stronger grouping cues among tones that are closer in pitch.

In a second experiment, we directly tested the *auditory motion hypothesis* against the *auditory grouping hypothesis* by introducing a change in auditory motion direction in each sequence to prevent the formation of a stable velocity referent across sequences. Each A-X-B tone sequence took on one of the four possible pitch separation (“pitch velocity”) conditions presented in Experiment 1, however, the pitch of target tone X was never between A and B. Thus, each sequence either had a downward-then-upward pitch motion trajectory, or an upward-then-downward trajectory. Because sequences within a given pitch velocity were presented in randomized order, the pitch trajectory per-trial was unpredictable, and uniform motion could not be imputed across the three tones. Thus, the *auditory motion hypothesis* predicts no kappa effect for these inconsistent pitch trajectory sequences. However, because the target X pitch was always closer to either A or B pitch, the *auditory grouping hypothesis* predicts a kappa effect based on pitch similarity, where the two tones that are closer in pitch are judged as closer in time due to grouping.

Finally, in a third experiment, we followed up on previous studies to test whether there is an auditory spatial kappa (ASK) effect. Although the original kappa effect in vision was based on spatial manipulations and the corresponding auditory effect in pitch-space has been described



with a velocity-based model initially developed for the visual paradigm, to our knowledge this effect has not been reliably shown for 3-sound sequences moving through auditory space.

Previous studies that have tested temporal perception for auditory sequences with virtual changes in spatial location have found inconclusive or mixed results. One study by Sarrazin et al. (2007) had participants reproduce the durations of seven inter-onset intervals separating eight sounds presented at equal- or variable-spatial distances. They simulated spatial separation between sounds by manipulating the presentation delay between two speakers placed two meters apart, where half of trials had constant virtual spatial distance between all tones and the other half had variable virtual distances. They found that participants produced more variable IOIs for sequences that simulated variable spacing compared to sequences that simulated equal spacing. However, there was no distance-duration correlation, such that participants did not produce longer intervals for larger apparent distances, in other words, no kappa effect. Roy et al. (2011) tested for ASK effects with single-interval stimuli, where two tones played from loudspeakers that were separated by 1.1 to 3.3 meters horizontally. They found that participants reported shorter inter-onset interval durations for sounds that were farther apart in space, the inverse of the kappa effect. These findings are surprising in light of the *auditory motion hypothesis*, which would predict strong kappa effects for sounds moving through auditory space.

However, a stronger test of ASK effects using real instead of simulated auditory space and a classic two-interval task is needed. Testing for an ASK effect, similar to previously observed visual and auditory pitch kappa effects, is important for understanding which auditory features give rise to kappa effects and whether they depend on spatial/motion percepts or feature-based grouping principles.

### **Experiment 1: Auditory Kappa Effect with Pitch-Based Velocity Manipulation**

This experiment was carried out as a partial replication and extension of a previous study examining pitch-based auditory kappa effects across sequences with different pitch velocities (Henry & McAuley, 2009). The present experiment compared the magnitude of observed kappa effects across tone sequences which varied in total pitch separation (with fixed temporal separation,  $IOI_T$ ) to create different pitch velocities. In contrast to the Henry and McAuley study, the total inter-onset interval ( $IOI_T$ ) between bounding tones ('A and B') was fixed across all conditions and pitch velocity was manipulated by varying the total pitch distance between tones A and B. Our velocity conditions included the fastest pitch velocity from Henry and McAuley (2009) as well as one slower and two faster conditions. All timing conditions of the target ('X') tone and three out of five target pitch conditions were exactly matched to Experiment 2 of their study. If total pitch velocity, and not total duration, of a tone sequence drives differences in the magnitude of the kappa effect, then we expect to replicate previous pitch velocity effects where kappa effects are larger for faster (i.e., larger pitch separation) than slower velocity sequences, as predicted by the *auditory motion hypothesis*.

#### **Method**

*Participants.* Twenty-four naïve participants were recruited from the University of Michigan (UM) and a community-wide health research recruitment database (<https://umhealthresearch.org>). Seven participants were excluded from analysis due to accuracy below 80% in the easiest timing conditions, indicating a failure to complete the experiment task as instructed. The final sample included seventeen participants ( $n = 14$  female,  $n = 3$  male) between the ages of 19 and 37 years ( $M = 24.1$ ,  $SD = 5.1$  years). Thirteen participants reported having some formal music training, ranging from 2 to 16 years ( $M = 8.1$ ,  $SD = 4.2$  yrs).

All participants underwent an audiogram and had pure-tone thresholds below 25 dB at octave frequencies from 250 – 8000 Hz. They were compensated at a rate of \$15 per hour for their participation. This study was approved by the Institutional Review Board - Michigan Medicine at the University of Michigan and all participants provided informed consent prior to participation.

*Design and Stimuli.* Experiment 1 implemented a 4 X 5 X 8 within-subjects design. Four pitch velocities (6 ST/728 ms, 8 ST/728 ms, 14 ST/728 ms, 18 ST/728 ms) were crossed with five target pitch levels (-2, -1, 0, +1, +2 ST) and eight target timing levels (+/- 4%, 8%, 12%, and 16%). Target tone pitch was set relative to the pitch midpoint (0 ST) between bounding tones and is represented as a signed ST deviation from this midpoint. Target tone timing was set relative to the theoretical bisection point (364 ms), which is half of the total inter-onset interval ( $IOI_T/2$ ) separating bounding tones and is represented as a signed percentage deviation from this bisection interval.

Stimuli were sequences of three 100 ms pure tones presented in increasing-frequency (pitch) order. Tones were presented in an ‘A-X-B’ paradigm, with a fixed temporal separation of 728 ms inter-onset interval (IOI) between bounding (A and B) tones. Within each block of trials, bounding tones had a fixed pitch separation of 6, 8, 14, or 18 semitones (ST). This produced the four pitch velocity conditions of 6 ST/728 ms, 8 ST/728 ms, 14 ST/728 ms, and 18 ST/728 ms (Figure 1 –Left panel). Tone A pitch was always fixed at a C4 (261.63 Hz) and tone B pitches corresponding to each velocity condition were: F#4, (369.99 Hz), G#4 (415.30 Hz), D5 (587.33 Hz), and F#5 (739.99). This range of pitch velocities was chosen to match the velocity showing the strongest kappa effect (8 ST/728 ms) in the previous study by Henry & McAuley (2009) and extended slightly slower (6 ST/728 ms) and faster (14 ST/728 ms & 18 ST/728 ms). Trials were

blocked within a pitch velocity, and the order of pitch velocity condition was counterbalanced across participants using a Latin-square design.

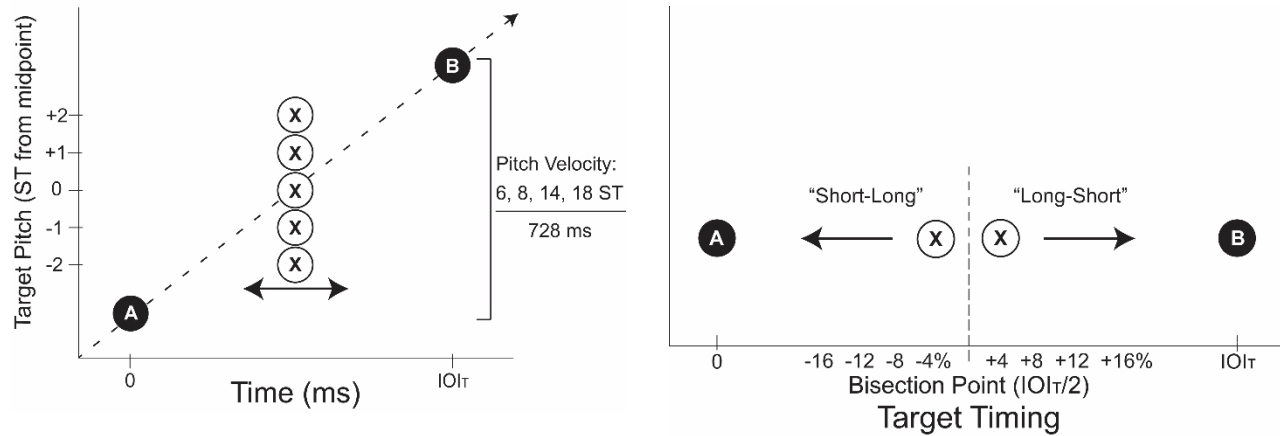


Figure 1. Experiment 1 stimulus schematics. *Left*: Five possible target ('X') pitch conditions. Bounding sounds 'A' and 'B' were fixed in pitch and temporal separation per block. The pitch of X (open circles) varied between 0 (pitch midpoint) and 2 semitones (ST) higher and lower from trial-to-trial. Four pitch velocity conditions (calculated as pitch distance in semitones over time) were created by changing the pitch separation between A and B while keeping the time interval between them ( $IOI_T$ ) constant. Dashed line represents the expected pitch trajectory according to the imputed pitch velocity model (Henry & McAuley, 2009). *Right*: Eight possible target timing conditions for all three experiments. The target tone was presented early (-) or late (+) with respect to the temporal bisection point ( $IOI_T/2$ ) by 4, 8, 12 or 16 %. Early targets created "short-long" IOI patterns, and late targets created "long-short" patterns.

In all velocity conditions, the target tone ('X') varied in pitch and timing from trial-to-trial in the same manner. Target tone pitch varied relative to the 0 ST (baseline) condition at the pitch midpoint between tones A and B (Figure 1 – left). This baseline condition corresponded to (a target pitch of) 3, 4, 7, or 9 ST between tones A and B for each of the 4 velocity conditions, respectively. All other target pitches were 1 or 2 semitones (ST) closer to the pitch of A or B, creating 5 target tone pitch conditions: -2, -1, 0, +1, +2 ST (from pitch midpoint). All bounding pitch and target pitch manipulations were selected to exclude same-note or octave-equivalent (i.e., 0 or 12 ST) relations among pitches in a sequence.

Target tone onset timing was set relative to the sequence temporal bisection point ( $\text{IOI}_T/2 = 364 \text{ ms}$ ) and took on one of eight values ( $\pm 4, 8, 12, 16\%$ ), where negative values were early and positive values were late with respect to the bisection point (Figure 1 – Right panel). For example, the onset of a target tone presented at  $-12\%$  timing was  $43.68 \text{ ms}$  early with respect to the bisection point (i.e.,  $320.32 \text{ ms}$  after A). Early target timings created a pattern of “short-long” intervals and late timings created a “long-short” pattern. Target timings were selected based on previous studies on auditory kappa effects to include a range of IOIs that span sub- to supra-threshold temporal discrimination for psychometric analyses (described below). All target timing and target pitch conditions were fully crossed and presented in randomized order once per experimental run, with 8 runs of 40 trials presented for each pitch velocity condition. Thus, a total of 320 trials were presented for each pitch velocity block.

*Equipment and Procedure.* Participants listened to ‘A-X-B’ tone sequences over Sennheiser HD 650 headphones while sitting at a computer terminal in a sound-attenuating booth. Stimuli were delivered and responses recorded via a MATLAB program written using the AFC package (Ewert, 2013). Their task was to listen to each three-tone sequence, then respond whether the pattern of time intervals between tones was “long-short” or “short-long” by clicking a labeled button on the computer monitor. Participants were explicitly told to ignore the pitch of the tones and only respond based on the timing.

Before starting the experiment, participants were shown a short slide deck with task instructions and given the opportunity to ask clarification questions to the experimenter. They then completed a practice block with correct/incorrect feedback consisting of five trials (one per target pitch condition) at each of the four easiest ( $\pm 12$  and  $16\%$ ) target timing conditions for a total of 20 practice trials. Participants repeated the practice as necessary until they scored  $75\%$

correct or higher. Participants who failed to meet this criterion after three or more attempts were not allowed to participate in the experiment.

Following practice trials, participants completed four blocks of 320 experimental trials, one for each pitch velocity condition, for a total of 1280 trials. Each of the four blocks were separated by a mandatory break. After all experimental trials, participants completed a short survey about demographics, music training, and any strategies used to perform the experiment task. The entire experimental session lasted around 2 hours in total.

*Data Analysis.* Proportions of “long-short” responses were computed for each experimental condition for each participant, and psychometric functions were fit to individual data. Point of Subjective Equality (PSE) was calculated for each pitch level at each velocity by computing the value of  $x$  when  $y = .5$  (corresponding to 50% long-short responses), then PSEs were averaged across participants. The PSE represents the theoretical time interval where the target sound should be presented (relative to the temporal midpoint between bounding sounds) to produce two perceptually-equivalent inter-onset intervals (IOIs). Negative PSE values correspond to a leftward shift in the psychometric function and represent a bias to respond “long-short,” and positive PSE values correspond to a rightward shift and bias to respond “short-long.” A PSE of 0 would indicate no systematic timing bias with equal 50% rate for “long-short” and “short-long” responses at the temporal bisection point. Following Henry and McAuley (2009), a relative Constant Error (CE) measure was calculated by subtracting PSE from the bisection IOI (half of the total sequence IOI). This measure is represented as a percentage deviation which can be compared across perceptual experiments with differing base IOIs (Gescheider, 1985). Positive relative CE corresponds to negative PSE (representing a bias to respond “long-short” more) and vice-versa. As a measure of discrimination threshold, we calculated just-noticeable difference

(JND) as the semi-interquartile range, and percent correct responses were also calculated for each experimental condition.

Effects of target pitch and pitch velocity on CE and JND were analyzed using linear mixed effects models with the *lmer4* package (Bates et al., 2015) and p-values were obtained with the *lmerTest* package (Kuznetsova et al., 2017) in the *R* environment (R Core Team, 2024).

## Results

To test for kappa effects across pitch conditions in this experiment, a linear mixed-effects model was performed on the CE with target pitch (ST deviation from midpoint), pitch velocity, and their interaction as fixed effects and participant as a random effect. The effect of target pitch on CE was significant, where CE increased by 2.33% on average for every semitone (ST) shift in target pitch (95% CI [1.80, 2.86],  $p < .001$ ). Positive slopes indicate the presence of a kappa effect, where participants showed greater bias to respond “long-short” when the target pitch was closer to the pitch of tone B than tone A (Figure 2). There was no effect of pitch velocity on CE, however there was a significant interaction between target pitch and pitch velocity ( $\beta = -0.10$ , 95% CI [-0.14, -0.05],  $p < .001$ ), with the slope of target pitch and CE being largest for the slowest (6 ST/728 ms) pitch-velocity. Steeper slopes indicated a greater magnitude of kappa effect for slower compared to faster pitch velocities. Average CE across all velocity and target pitch conditions was 2.41%, and a one-sample t-test against 0 was significant ( $t(339) = 10.97$ ,  $p < .001$ , 95% CI [1.98, 2.84]), indicating a general response bias.

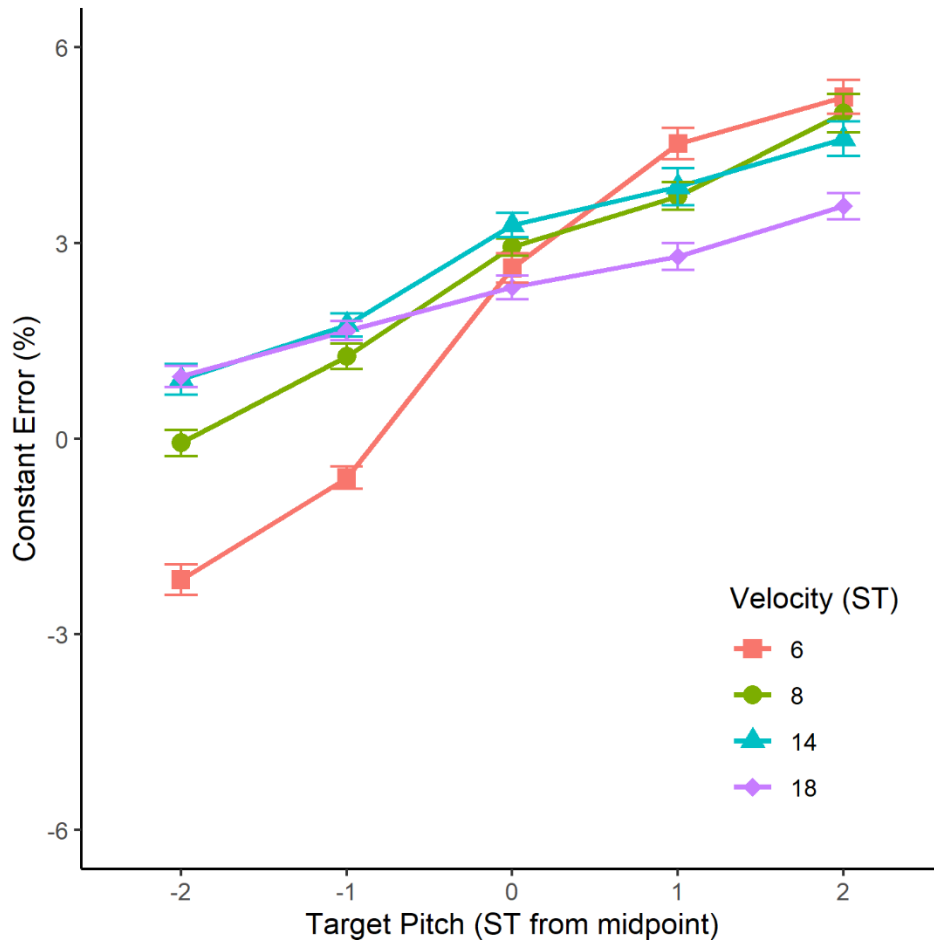


Figure 2 Experiment 1 results showing mean relative CE with standard error bars as a function of target pitch deviation in the four pitch velocity conditions. Kappa effects were larger (steeper slopes) for slower (e.g., 6 ST) compared to faster (e.g., 18 ST) velocity conditions.

A similar linear mixed-effects model on JND showed no effect of pitch but a significant effect of velocity on temporal discrimination threshold ( $\beta = 0.47$ , 95% CI [0.29, 0.66],  $p < .001$ ), with the largest JNDs observed for the fastest velocity condition (18 ST / 728 ms), and no interaction between target pitch and velocity. Thus, discrimination thresholds (JNDs) were not affected by target pitch deviations but were larger for faster pitch velocity conditions. JNDs ranged from 18.82 to 27.97 ( $M = 22.08$ ) across all pitch and velocity conditions and were comparable to previous studies.



## Discussion

In Experiment 1, average CE overall was positive, indicating a general bias toward responding that sequences had a “long-short” timing pattern. Our results also replicated the pitch-based auditory kappa effect found in previous studies: participants showed greater bias to respond “long-short” when the target pitch level was higher compared to when it was 0 or lower. Thus, when the pitch of the target tone was closer to the pitch of tone B than tone A, participants judged the target to be closer in time to B than A. However, in contrast to previous findings, our results showing an interaction between target pitch and velocity demonstrated a larger kappa effect for slower compared to faster velocities when pitch velocity was manipulated by changing only the pitch separation while keeping the time interval between A and B ( $IOI_T$ ) constant. This effect was in the opposite direction of previous results from the study by Henry and McAuley (2009) who found larger kappa effects for faster velocities created by manipulating only  $IOI_T$ . Additionally, temporal discrimination thresholds (JNDs) were generally larger for faster velocities.

### Experiment 2: Auditory Kappa Effect with Inconsistent Pitch Trajectories

Experiment 2 served as a test of the competing *auditory motion* and *auditory grouping hypotheses* by examining whether the auditory kappa effect is present for sequences with unpredictable auditory motion trajectories. In contrast to Experiment 1, ‘A-X-B’ tone sequences did not have pitches presented in increasing-frequency order. Instead, all sequences in Experiment 2 included a change in pitch motion direction, from downward-to-upward or from upward-to-downward, by manipulating the target tone pitch from trial-to-trial such that it was never between the pitch of tone A and tone B. Within a block, these inconsistent pitch motion trajectories were presented in randomized order. Sequences with unpredictable changes in pitch

motion direction should prevent listeners from forming expectations of constant velocity previously proposed to give rise to auditory kappa effects. Thus, the *auditory motion hypothesis* predicts no kappa effect for sequences with sudden changes in pitch motion direction, due to a lack of reliable pitch/time trajectory predictions across sequences. However, the *auditory grouping hypothesis* predicts a kappa effect for these sequences due to participants perceptually grouping together the two tones that are more similar in pitch, and streaming out the third, more pitch-distant, tone.

## Method

*Participants.* Twenty-nine naïve participants recruited from the University of Michigan community took part in the experiment. Twelve participants were excluded from analysis due to accuracy below 80% in the easiest timing conditions, indicating a failure to complete the experiment task as instructed. The final sample included seventeen participants ( $n = 9$  female,  $n = 8$  male) between the ages of 18 and 41 years ( $M = 24.1$ ,  $SD = 7.4$  years). Fifteen participants reported having some formal music training, ranging from 2 to 11 years ( $M = 7.7$ ,  $SD = 3.0$  yrs).

*Design and Stimuli.* Experiment 2 implemented a 8 X 8 within-subjects design. Eight relative target pitch distances ( $\pm 6, 8, 14, 18$  ST) were presented across eight target timing levels ( $\pm 4\%, 8\%, 12\%$ , and  $16\%$ ). The eight target pitch conditions included four target pitch conditions below the pitch of tone A ( $-$  ST) and four above tone B ( $+$  ST). This created two types of pitch contours where half of target pitches introduced a change in pitch motion direction from decreasing-to-increasing frequency and the other half changed direction from increasing-to-decreasing (Figure 3 – top left panel). Thus, when the pitch of X was below that of A ( $-$ ST), the pitch was relatively closer to A than B and vice-versa. Within each relative pitch distance

condition, the pitch was roved by 1 ST. Bounding and target timing conditions were identical to those described in Experiment 1 (Figure 1 – right panel).

Just as Experiment 1, stimuli were three-tone sequences in an ‘A-X-B’ paradigm where tone A and tone B were separated by 728 ms and B was always 6, 8, 14, or 18 ST higher than A, as in Experiment 1. In Experiment 2, however, the three tones were not in increasing-frequency order. Target (‘X’) tone pitches were always lower than tone A or higher than tone B, thus, target pitch conditions introduced a change in the direction of pitch motion within each sequence. Randomized changes in the direction of pitch motion should prevent the formation of expected constant pitch velocities. When the target pitch was lower than A (negative X pitch), it was relatively closer to A than B pitch and when the target pitch was higher than B (positive X pitch), it was relatively closer to B than A. Within a condition, A and B had fixed pitch separation (analogous to pitch velocity in Experiment 1) equal to the magnitude of target relative pitch. The relative target pitch distance is summarized as the difference in absolute values of pitch leap size (in ST) from A to X and from X to B ( $|AX| - |XB|$ ). Thus, relative pitch difference of X is a signed measure characterized by magnitude and direction of pitch proximity to A compared to B. For example, if the bounding tones were 6 ST apart, the target tone, X, was always either 6 ST closer to A than B (- X conditions) or 6 ST closer to B than A (+ X conditions). The proximity of pitch X to pitch A compared to pitch B was always such that the difference in pitch distance from A to X and from X to B was equal to the total pitch displacement between A and B. The sign of X indicates the direction of pitch change: a negative sign (-) signifies that pitch X moved below A, while a positive sign (+) signifies that pitch X moved above B.

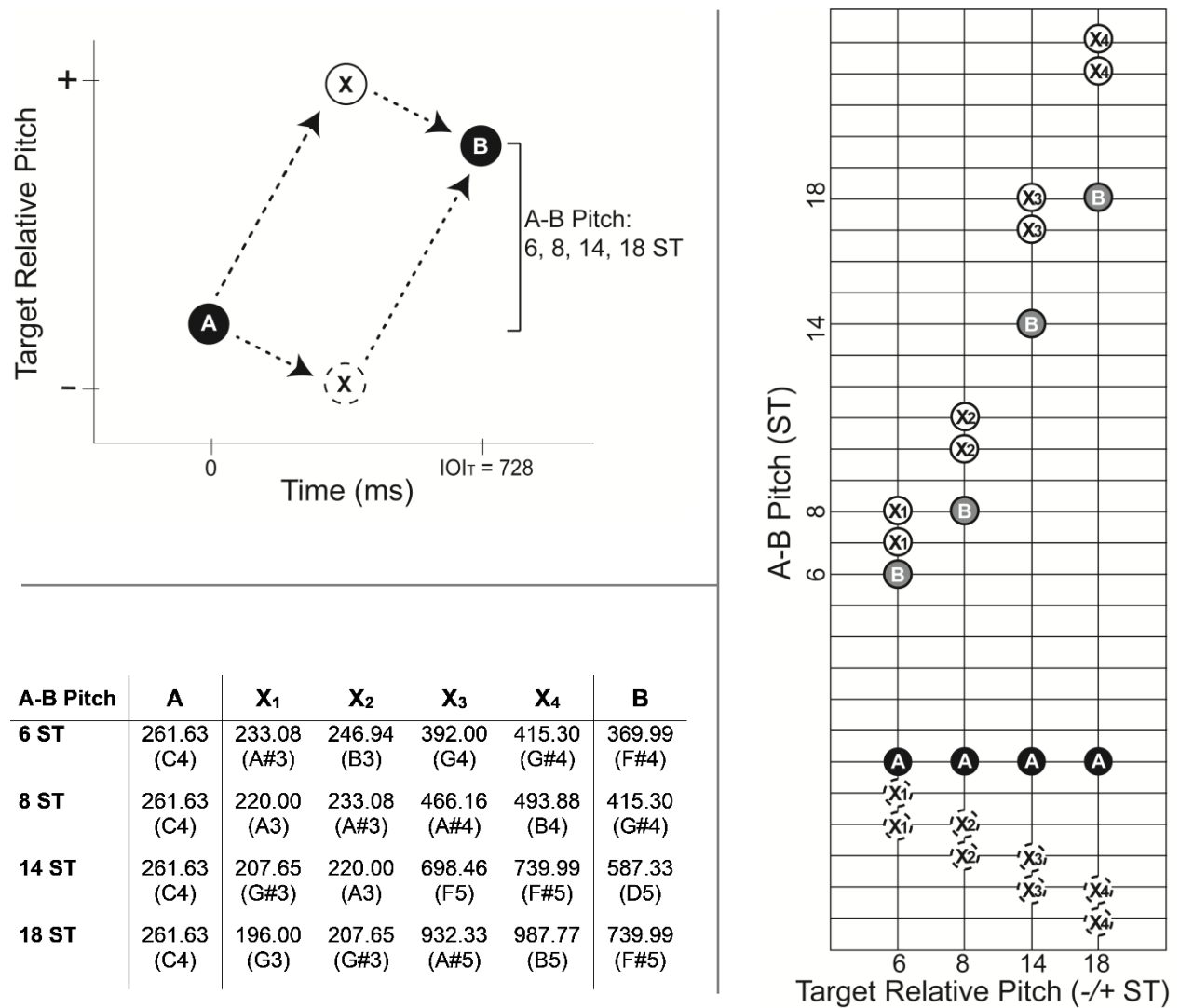


Figure 3. Experiment 2 stimulus schematics. *Top Left*: Inconsistent pitch trajectories formed by altering the pitch of X such that it was never between A and B. The A-B pitch conditions were identical to the pitch velocities from Experiment 1. *Bottom Left*: Table of all tone frequencies (in Hz) with musical pitch names in parentheses. Rows display all possible pitch values for a given A-B pitch condition. Tone A was always 261.63 Hz, and tone B was 6, 8, 14, or 18 semitones (ST) higher than tone A. The pitch of target tone X varied across conditions and was 6, 8, 14, or 18 semitones (ST) closer to A than B (-), or vice-versa (+). X<sub>1</sub> and X<sub>2</sub> are roved values of target pitches closer to A (corresponding to dashed circles in other panels), X<sub>3</sub> and X<sub>4</sub> are roved pitches closer to B. *Right*: Graphical representation of all possible pitch conditions. The relative pitch distance of X ( $|AX| - |XB|$ ) was always equal to A-B pitch distance, where negative pitch distances (dashed circles) represent X pitch closer to A and positive distances (solid circles) represent X pitch closer to B. The X pitch was roved between two values per condition.

There were 2 possible pitch values for X in each pitch distance condition, for a total of 16 possible target pitches. All absolute pitch values for each condition are shown in Figure 3 (bottom left). Because tone A was fixed across pitch distance conditions, we varied the pitch of X by 1 ST such that target pitches were not identical across all conditions. In the +/- 6 ST conditions, X was 1 to 2 ST below A or above B, in the +/- 8 ST condition, X was 2 to 3 ST below A or above B, in the +/- 14 ST conditions, X was 3 to 4 ST below A or above B, and in the +/- 18 ST conditions, X was 4 to 5 ST below A or above B (Figure 3 – right). As in Experiment 1, pitch manipulations excluded same-note or octave-equivalent relations among pitches presented within a sequence. There were 4 blocks of trials corresponding to +/- 6 ST, +/- 8 ST, +/- 14 ST, and +/- 18 ST conditions with 10 runs of 32 trials in each block for a total of 320 trials per block. Each trial run had 1 presentation of each of the 4 possible target pitch and timing offset combinations presented in randomized order.

*Equipment and Procedure.* All equipment and procedures were identical to Experiment 1.

*Data Analysis.* All procedures for computing dependent variables were identical to Experiment 1. Effects of pitch distance on CE and JND were analyzed using linear mixed effects models using the same *R* packages as in Experiment 1.

## **Results**

Figure 4 shows average CE as a function of relative target pitch distance. Upon visual inspection, CE was more negative for more negative relative pitch distances and more positive for larger positive distances, indicating more “short-long” responses when the target pitch was closer to the pitch of tone A than B, and more “long-short” responses when the target pitch was closer to B than A. To test for kappa effects as a function of pitch proximity in sequences with

inconsistent pitch trajectories, a linear mixed-effects model was performed on constant error (CE) with target relative pitch distance (ST) as a fixed effect and participant as a random effect. The effect of target pitch on CE was significant, where CE increased by .22% for every 1 ST shift in target pitch closer to B than A (95% CI [0.18, 0.27],  $p < .001$ ). This significant slope indicates the presence of a kappa effect, where participants showed greater bias to respond “long-short” when the target pitch was closer to the pitch of tone B than tone A. Average CE across all velocity and target pitch conditions was .51%, and a one-sample t-test against 0 was not significant ( $t(271) = 1.50$ ,  $p = .14$ , 95% CI [-0.16, 1.19]).

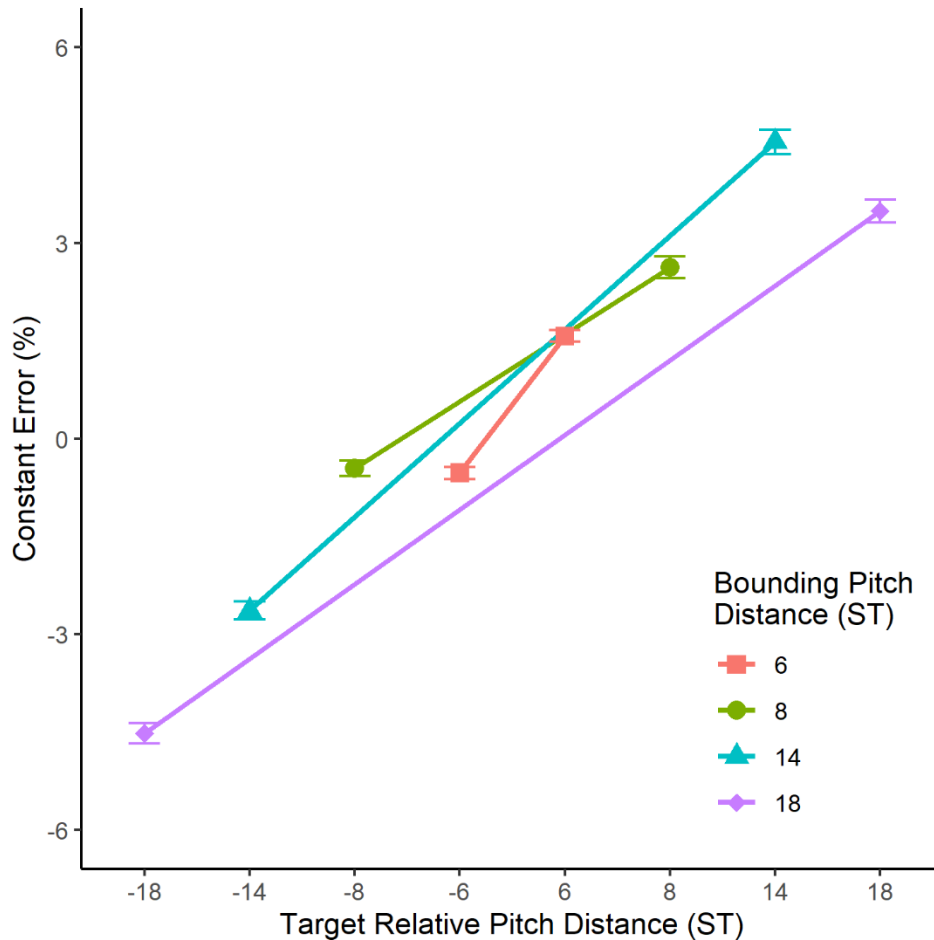


Figure 4. Experiment 2 results showing mean relative CE with standard error bars as a function of relative target pitch distance. Negative relative pitch difference values represent targets ('X') closer in pitch to A than B, and positive values were closer in pitch to B than A. The absolute value of relative target pitch distance was equal to A-B pitch distance (colored lines). Results showed a kappa effect in all conditions where participants were biased to report "short-long" more when X pitch was closer to A (negative relative pitch) and "long-short" when X pitch was closer to B (positive relative pitch).

To consider the possibility that the total pitch motion across the three tones systematically biased timing judgments in this study, we computed the summed absolute value of pitch change across all 'A-X-B' sequences. The absolute change in pitch from A to X and from X to B represents the total pitch distance moved in each sequence, regardless of motion direction (whereas the distance from tone A to tone B is analogous to total pitch 'displacement'). We then

performed a similar linear mixed effects analysis on CE with total pitch change and bounding pitch distance and their interaction as fixed effects and participant as a random effect. The total pitch distance across all 3 tones did not have a significant effect on CE ( $\beta = 0.1$ , 95% CI [-0.23, 0.43],  $p = .56$ ) nor did bounding pitch distance (i.e., pitch displacement) ( $\beta = 0.54$ , 95% CI [-0.29, 1.38],  $p = .20$ ), and the interaction was not significant ( $\beta = -0.02$ , 95% CI [-0.05, 0.01],  $p = .19$ ). Thus, the total change in pitch over each sequence was not associated with a systematic bias in timing, whether considering the pitch change from A to X to B or the pitch displacement from A to B regardless of target pitch X.

Finally, a linear mixed-effects model on JND showed a significant effect of relative pitch distance ( $|AX| - |XB|$ ) on temporal discrimination threshold ( $\beta = 0.19$ , 95% CI [0.11, 0.27],  $p < .001$ ), where discrimination thresholds increased as pitch distance grew more positive. Average JNDs ranged from 17.21 to 25.98 ( $M = 20.88$ ) for this experiment, in line with previous experiments.

## Discussion

The results of Experiment 2 showed a pitch-based auditory kappa effect for sequences of tones with unpredictable (i.e., randomized) changes in pitch motion direction per trial. Participants responded that the timing pattern was “long-short” more when the target tone (X) was relatively closer in pitch to tone B than tone A, and responded “short-long” more when X was relatively closer in pitch to tone A, despite all sequences having inconsistent pitch trajectories. There was no difference in the magnitude of kappa effect between A-B pitch conditions which were analogous to different pitch velocities from Experiment 1. Furthermore, there was no systematic timing bias (i.e., no kappa effect) based on the total pitch motion or pitch displacement of sequence conditions. Taken together, these results support the *auditory grouping*



*hypothesis* over the *auditory motion hypothesis* as an underlying explanation of auditory kappa effects.

### **Experiment 3: Auditory Spatial Kappa Effect**

Experiment 3 tested how the locations of sequential sounds moving through auditory space in a free-field setting bias listeners' judgments of temporal onsets. This experiment tested for an auditory spatial kappa (ASK) effect by varying the spatial proximity of the target sound relative to the bounding sounds in an 'A-X-B' paradigm similar to Experiments 1 and 2. This experiment extends previous studies on visuospatial kappa effects to the auditory-spatial domain. Previous studies on ASK effects have produced mixed results. One study using simulated or 'virtual' space failed to show kappa effects for 8-tone sequences (Sarrazin et al., 2007), and two studies using pure-tones played from loudspeakers at varying distances (i.e., in "real" space) showed either the inverse of the kappa effect (i.e., shorter durations associated with longer distances) with 2-tone sequences (Roy et al., 2011) or mixed results across conditions with 4-tone sequences (Grondin & Plourde, 2007). Because pure-tones are difficult to accurately localize, we tested for this effect using narrowband noise (NBN) bursts in three differing sound fields, each spanning 90°, within the front hemifield to examine how varying auditory spatial resolution as a function of sound source location moderated temporal judgment effects (Hartmann, 1983; Fostick & Fink, 2021; Shigeno & Oyama, 1983; Stevens & Newman, 1936). Furthermore, we tested two frequency bands to compare contributions of interaural level and timing cues for sound source location on the proposed ASK effect, as lower frequencies (below ~1500 Hz) have more reliable timing cues and higher frequencies (above ~3000 Hz) have more reliable level cues for localization (for a review see: Middlebrooks & Green, 1991).

*Participants.* Twenty-eight naïve participants recruited from the University of Michigan community took part in the experiment. Six participants were excluded from analysis due to accuracy below 80% in the easiest timing conditions, indicating a failure to complete the experiment task as instructed. The final sample included 22 participants ( $n = 12$  female,  $n = 9$  male,  $n = 1$  non-binary) between the ages of 20 and 57 years ( $M = 29.4$ ,  $SD = 8.3$  years). Twenty participants reported having some formal music training, ranging from 3 to 18 years ( $M = 7.6$ ,  $SD = 3.7$  yrs).

*Design and Stimuli.* Experiment 3 implemented a  $2 \times 3 \times 5 \times 8$  within-subjects design. Two frequency bands (500 and 4000 Hz center) and three sound fields (left, front, and right  $90^\circ$ ) were crossed with five target locations ( $-30$ ,  $-15$ ,  $0$ ,  $+15$ ,  $+30^\circ$  azimuth) and eight target timing levels ( $\pm 4$ ,  $8$ ,  $12$ , and  $16\%$ ). Target timing levels were the same as previous experiments in terms of percent deviation from the temporal bisection point (Figure 1 – Right panel).

Sequences of three 200 ms narrowband noise (NBN) bursts were presented in an ‘A-X-B’ paradigm similar to Experiment 1 and 2. Bounding (A and B) sounds had fixed spatial and temporal separation of  $90^\circ$  azimuth and 1000 ms IOI<sub>T</sub>, respectively. The target (X) location was always between A and B, with five target location conditions that were spatial analogues to target pitch conditions from Experiment 1. Target sounds were presented at the spatial midpoint between bounding sounds as well as  $15^\circ$  and  $30^\circ$  closer in space to sound A or B. Target locations are represented relative to the spatial midpoint (0), where negative spatial deviations represent target sounds closer in space to A and positive deviations were closer to B. Each sound field spanned  $90^\circ$  between A and B, and the three sound fields covered the front half of the speaker array such that no sounds were ever presented behind the participant (Figure 5). Sound field conditions were blocked, with the order of fields counterbalanced across participants according

to a Latin-squares scheme. Thus, sounds within a block were always presented in a consistent direction, with sounds in the right and left fields starting directly in front of the participant ( $0^\circ$  azimuth) and moving to the side, ending at  $90^\circ$  (right) or  $270^\circ$  (left), and in the front field always moved from left to right, starting at  $315^\circ$  and ending at  $45^\circ$  azimuth. NBN sounds had a bandwidth of one-octave and either 500 ('low') or 4000 Hz ('high') center frequency.

The eight target timing and five target location conditions were fully crossed and presented in randomized order once per run, with 8 runs of 40 trials for each sound field for a total of 320 trials per block. There was a short mandatory break (1+ min) between each block before moving on to the next sound field condition. Participants completed all 960 trials in the high or low frequency condition in one session, then returned 1 to 49 days later ( $M = 11$ ,  $SD = 15$ ) to complete the other frequency condition, with the order of frequencies counterbalanced across participants.

*Equipment.* This experiment took place in a hemi-anechoic chamber equipped with a 24-channel 360-degree surround loudspeaker system under the control of a dedicated PC using a custom MATLAB program. Sounds were presented over Orb Audio loudspeakers through an Antelope Orion 32+ D/A multi-channel converter and Crown CT8150 amplifier for audio rendering and output. Loudspeakers were positioned every  $15^\circ$  in the azimuthal plane (Figure 5) and 44 inches vertically, corresponding to roughly ear-height when seated. Responses were collected via a wireless Bluetooth numeric keypad with printed labels on two buttons displaying "short-long" and "long-short" response options.

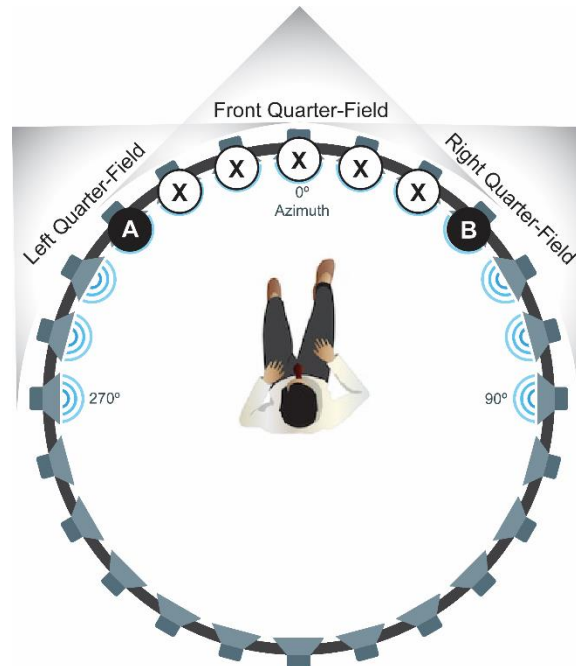


Figure 5. Schematic showing an overhead view of participant seated in center of 360° azimuth speaker array with 3 sound fields labeled. All fields spanned 90° (i.e., a “quarter-field”) between sounds ‘A and B’, with sound sequences moving from 315° to 45° in front, from 0° to 270° on the left, 0° to 90° on the right. Example of A-X-B locations for the front field are displayed with five possible target X locations shown in open circles.

*Procedure.* Participants listened to ‘A-X-B’ NBN burst sequences while seated in the center of the 360° speaker array and facing a fixation ‘+’ positioned at 0° azimuth. As in experiments 1 and 2, they heard each sequence and responded whether the pattern of time intervals between tones was ‘long-short’ or ‘short-long’ by clicking a labeled button on the handheld keypad. Participants were explicitly told to ignore the spatial location of the sounds and only respond based on the timing. Prior to starting the experiment, participants completed practice trials with feedback following the same procedure as Experiment 1 and 2. Participants then completed three blocks of 320 experimental trials, one for each sound field condition, for a total of 960 trials per session, and each session lasted around 2.5 hours in total. All other aspects of the practice and testing procedure for this task were identical to previous experiments.

## Results

Figure 6 shows average CE as a function of target location for all sound fields in low (left panel) and high frequency band (right panel) conditions. Upon visual inspection, CE was lower for negative location deviations and higher for positive locations, indicating relatively more “short-long” timing responses when the target sound was closer in space to sound A, and more “long-short” responses when the target sound was closer to B. To test for kappa effects as a function of spatial location in each sound field and frequency in this experiment, a linear mixed-effects model was performed on constant error (CE) with target location (degree deviation from midpoint), sound field, frequency, and their two-way and three-way interactions as fixed effects and participant as a random effect. The effect of target location on CE was significant, where CE changed by .20% for every one-degree shift in target location in the azimuthal plane (95% CI [0.14, 0.26],  $p < .001$ ). Significant slopes indicated the presence of a kappa effect, where participants showed greater bias to respond “long-short” when the target sound location was closer to sound B. The effect of sound field on CE was also significant. Average CE for the right quarter-field was on average 2.1% higher than the front quarter-field (95% CI [0.36, 3.78],  $p = .02$ ), and average CE for the left quarter-field was on average 1.9% higher than the front quarter-field (95% CI [0.20, 3.62],  $p = .03$ ). There were no effects of frequency or significant two-way or three-way interactions (all  $p > .063$ ). Average CE across all locations, sound fields and frequencies was 4.73%, and a one-sample t-test against 0 was significant ( $t(659) = 13.94$ ,  $p < .001$ , 95% CI [4.06, 5.40]), indicating a general response bias.

To investigate temporal discrimination thresholds across conditions in this experiment, a similar linear mixed-effects regression model was performed on JND scores as in previous experiments. This analysis showed no significant effects or interactions (all  $p > .061$ ). Thus, we

did not find an effect of target location, sound field condition, or frequency band on discrimination thresholds in this experiment. Average JNDs ranged from 21.72 to 34.47 ( $M = 27.05$ ) across all conditions, comparable to previous experiments.

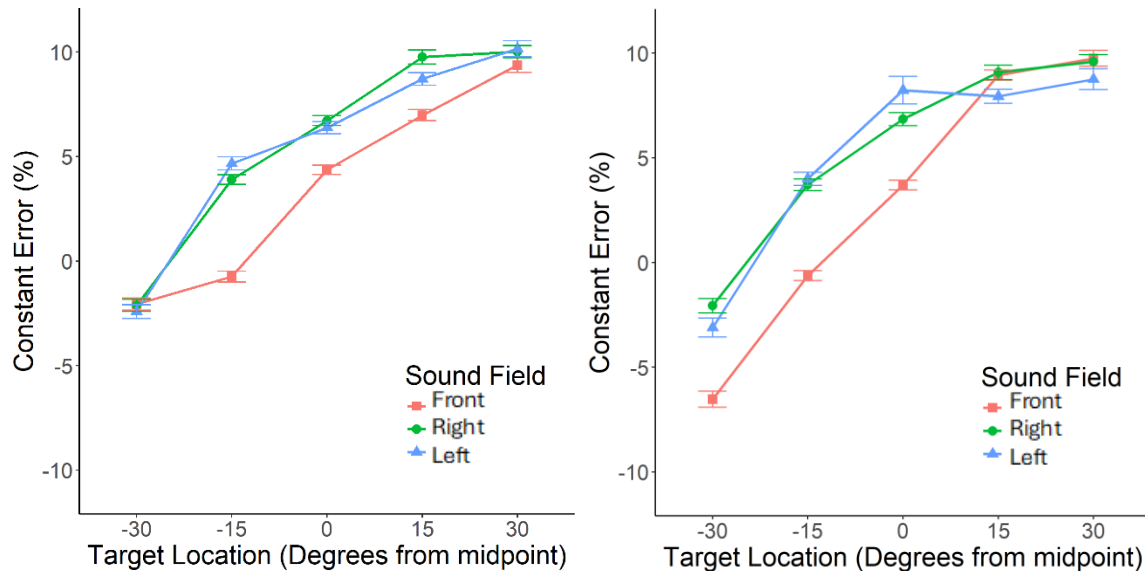


Figure 6. Experiment 3: Mean relative CE with standard error bars as a function of target location deviation in the three sound field conditions with low frequency (left panel) and high frequency (right panel) NBN sounds. Results showed ASK effects (significant positive slopes) in all location and frequency conditions.

## Discussion

Results of Experiment 3 provided strong evidence for an auditory spatial kappa (ASK) effect using NBN sequences moving through space (with consistent center frequencies). In general, participants showed a relative bias to respond “short-long” more when the target location was closer to A, and a bias to respond “long-short” more when the target X spatial location was closer to B than A. Similar to Experiment 1, average CE overall was positive, indicating a general bias toward responding that sequences had a “long-short” timing pattern in

this experiment. A significant effect of sound field was also found, where point estimates for CE were higher for the right and left quarter-fields compared to the front, indicating that this general “long-short” response bias was greater for sounds moving from directly in front of participants to the right or left side compared to sounds moving within the front quarter field (from left to right) crossing the sagittal midline. No significant effects of frequency were found, such that ASK effects were present for NBN sounds in lower and higher frequency bands (centered around 500 and 4000 Hz, respectively). In sum, these results support the generalizability of kappa effects from visual space and pitch space to auditory space.

### General Discussion

The primary goals of the present study were to test the prevailing *auditory motion hypothesis* against an *auditory grouping hypothesis* as a parsimonious alternative explanation for auditory kappa effects and to test for a reliable auditory spatial kappa (ASK) effect with sounds moving through space. Three experiments were conducted in which three sequential sounds (A-X-B) changed in pitch or spatial location and participants reported the relative durations of the first (A-X) and second (X-B) inter-onset interval (IOI) by responding that the timing pattern was “short-long” or “long-short.” Experiment 1 tested the prediction of the *auditory motion hypothesis* that tone sequences with faster pitch velocity cause larger-magnitude kappa effects than slower velocities for differing velocity conditions created by manipulating the overall pitch separation (rather than overall temporal separation) among tones. Experiment 2 tested the prediction of the *auditory grouping hypothesis* that kappa effects would be observed for tone sequences with inconsistent pitch trajectories, regardless of non-uniform pitch motion. Experiment 3 tested for ASK effects using sequences of three narrowband noise (NBN) bursts moving through auditory space in three sound field and two frequency band conditions.

Experiment 1 replicated and extended previous findings on pitch-based kappa effects with one key difference in the pattern of our results. Kappa effects were larger for sequences with slower velocities when pitch velocity was manipulated by changing the overall pitch separation of tones. The interaction between target pitch level and velocity condition demonstrated that CE slopes were steeper for slower velocity conditions compared to faster velocity conditions, contradicting previous studies showing the reverse pattern of results when manipulating pitch velocity by changing the overall temporal separation between tones (Henry & McAuley, 2009). These results contradict the predictions of the *auditory motion hypothesis*, supporting instead the *auditory grouping hypothesis* as the pitches of sequential tones in slower velocity conditions (i.e., 6 ST/ 728 ms) were closer together, supporting grouping of tones more than in faster velocity conditions (i.e., 18 ST/ 728 ms) which had tones that were more likely to be perceptually segregated into independent streams (Moore & Gockel, 2002; Vliegen et al., 1999). For example, when target tone X was shifted 2 semitones (ST) closer to A relative to the pitch midpoint in the slowest velocity condition, the pitch separation between AX was 1 ST, and between XB was 5 ST, supporting grouping of A and X into one stream and segregation of B into a separate stream. However, the analogous pitch shift (-2 ST) in the fastest velocity condition resulted in a pitch separation between AX of 7 ST and between XB of 11 ST. Thus, relative pitch proximity of the target tone provided a weaker grouping cue in the faster compared to the slower pitch velocity conditions. While results of this experiment may appear to contradict the findings of Henry and McAuley (2009) that faster velocity sequences elicit stronger kappa effects, auditory streaming models have demonstrated that both greater pitch separation and smaller temporal separation among tones more reliably induces stream segregation (Bregman, 1990; van Noorden, 1975). Thus, the overall reduction of IOIs between all tones in their faster velocity



sequences would predict a stronger segregation of the more distinct pitch, resulting in a greater duration overestimation for gaps across streams.

The results of Experiment 2 provided further support for the *auditory grouping hypothesis* by showing an auditory kappa effect based on relative pitch proximity for sequences of tones with inconsistent pitch trajectories. As in previous studies of pitch-based kappa effects, participants were biased to report that the two A-X-B tones closer in pitch were closer in time. However, in our experiment, each sequence had non-uniform motion such that there was a change in direction of pitch change from downward-to-upward or from upward-to-downward that was randomized across trials to prevent the formation of a stable velocity referent. Thus, this kappa effect cannot be explained by violations of target pitch expectations based on imputing average constant pitch velocity across sequences. When we considered only the overall distance of pitch motion across all three tones (absolute value of pitch change from A to X to B) or the pitch displacement from A to B (regardless of X), no systematic difference in timing bias was observed as a function of pitch velocity, where  $V = \Delta p / \Delta t$ .

Finally, the results of Experiment 3 showed robust kappa effects for A-X-B sounds presented from different speaker locations, confirming the generalizability of visuospatial kappa effects to the auditory spatial domain. When tested previously with tone sequences, results showed mixed support for ASK effects with some studies showing null results (Sarrazin et al., 2007) or an inverse kappa effect (Roy et al., 2011), and only one study showing mixed evidence with a kappa effect present for a subset of duration conditions (Grondin & Plourde, 2007). Our results demonstrated an ASK effect for NBN sequences moving through auditory space, where participants were biased to report that the target ‘X’ sound was closer in time to the bounding ‘A’ or ‘B’ sound that was closer in space. This supports our hypothesis that previous studies failing

to show this effect were due to insufficient cues for spatial location with pure-tone stimuli, which may have prevented space-based temporal biases (Fostnick & Fink, 2021; Shigeno & Oyama, 1983; Stevens & Newman, 1936). Our results showing an ASK effect for NBN sounds presented in both low (500 Hz) and high (4000 Hz) frequency bands indicates that this is a general effect of spatial location, regardless of dominant interaural level or interaural timing cues. Interestingly, the magnitude of ASK effect was equivalent across sound field conditions, although localization performance tends to be better for sounds presented in front of the listener than directly to the side (Middlebrooks & Green, 1991; Stevens & Newman, 1936). One possibility is that our smallest spatial separation between sequential sounds of 15° was large enough to be readily discriminated (i.e., suprathreshold) regardless of sound field condition.

When considered with previous studies on auditory kappa effects, results of the present experiments support the streaming-based *auditory grouping hypothesis* over the *auditory motion hypothesis* as an explanation for these effects. Our findings are in line with a long-standing literature on auditory streaming paradigms that have demonstrated perceptual grouping based on features including pitch, spatial location, spectral content, intensity, and temporal structure (Bizley & Cohen, 2013; Micheyl et al., 2013; Moore & Gockel, 2012; Vliegen et al., 1999). In general, auditory features that serve as cues for grouping or segregating stimuli into objects or streams also bias time perception among sequential events. This applies not only to kappa effects, but also to previously-established temporal ordering effects where listeners fail to judge the order of rapid events across streams although able to accurately judge sequential order within a stream (Barsz, 1991; Bregman & Campbell, 1971; Micheyl & Oxenham, 2010). More broadly, feature binding that contributes to object perception and scene analysis is well-established based on heuristics and Gestalt principles of figure-ground organization across sensory domains. This

may form a common neurological and cognitive basis for feature-based perceptual time illusions (i.e., dilation or contraction) in the presence of temporal ambiguities.

The current study has broad implications for models of stream segregation and our understanding of how different stimulus features influence time perception. Several models of stream segregation are based on temporal coherence, suggesting that cross-feature temporal alignment facilitates grouping of sound sources (Krishnan et al., 2014; Lu et al., 2024; Shamma et al., 2011; Szabo et al., 2016; Teki et al., 2013). However, these models must also consider bidirectional interactions, including how non-temporal features affect temporal judgments. The confluence of evidence for kappa across various stimulus attributes and tasks suggests these time-perception phenomena are influenced by perceptual context effects. Future research should test current models' ability to account for dynamically-changing contextual features to improve time perception models.

### **Acknowledgements**

We would like to thank Dr. Nori Jacoby and Ilhan Onder for helpful discussions at the start of this research, research assistants Zhiqing Yin and Deshaun Omane for their help with programming and data collection for Experiments 1 and 2, and Dr. Ruth Litovsky for valuable comments and feedback during the planning and conceptualization of Experiment 3. This project was supported by grants NIH T32DC00011 and F32DC022162 (CK).

### References

- Alards-Tomalín, D., Leboe-McGowan, L.C. & Mondor, T.A. Examining auditory kappa effects through manipulating intensity differences between sequential tones. *Psychological Research* 77, 480–491 (2013). <https://doi.org/10.1007/s00426-012-0438-8>
- Barsz, K. Auditory pattern perception: The effect of tone location on the discrimination of tonal sequences. *Perception & Psychophysics* 50, 290–296 (1991). <https://doi.org/10.3758/BF03206752>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package ‘lme4’. *Convergence*, 12(1), 2.
- Bausenhart, K.M., Quinn, K.R. (2018). On the interplay of visuospatial and audiotemporal dominance: Evidence from a multimodal kappa effect. *Attention, Perception, & Psychophysics* 80, 535–552. <https://doi.org/10.3758/s13414-017-1437-z>
- Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, 14(10), 693-707. <https://doi.org/10.1038/nrn3565>
- Boltz, M. G. (1998). The processing of temporal and nontemporal information in the remembering of event durations and musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1087–1104. <https://doi.org/10.1037/0096-1523.24.4.1087>

- Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *44*(3), 400 - 413. <https://doi.org/10.1037/h0084255>
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249. <https://doi.org/10.1037/h0031163>
- Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1953). A new phenomenon in time judgment. *Nature*, *172*(4385), 901-901. <https://doi.org/10.1038/172901a0>
- Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1954). Interdependence of temporal and auditory judgments. *Nature*, *174*(4431), 642-644. <https://doi.org/10.1038/174642a0>
- Eitan, Z. (2013). How pitch and loudness shape musical space and motion. In Tan, S. L., Cohen, A., Cohen, A. J., Lipscomb, S. D., & Kendall, R. A. (Eds.), *The psychology of music in multimedia*, (pp. 165-191). Oxford University Press
- Ewert, S. D. (2013). AFC – A modular framework for running psychoacoustic experiments and computational perception models. In German Acoustical Society (DEGA) (Ed.), *Proceedings of the International Conference on Acoustics AIA-DAGA 2013* (pp. 1326–1329). Merano, Italy: German Acoustical Society (DEGA).
- Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics*, *16*(3), 522-528. <https://doi.org/10.3758/BF03198581>

Fostick, L., & Fink, N. (2021). Situational awareness: The effect of stimulus type and hearing protection on sound localization. *Sensors*, *21*(21), 7044.

<https://doi.org/10.3390/s21217044>. PMID: 34770351; PMCID: PMC8587889.

Gescheider, G., (1985). *Psychophysics: Method, theory, and application* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grondin, S., & Plourde, M. (2007). Judging multi-minute intervals retrospectively. *Quarterly Journal of Experimental Psychology*, *60*(9), 1303-1312.

<https://doi.org/10.1080/17470210600988976>

Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, *74*(5), 1380-1391. <https://doi.org/10.1121/1.390163>

Henry, M. J., & McAuley, J. D. (2009). Evaluation of an imputed pitch velocity model of the auditory kappa effect. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 551. <https://doi.org/10.1037/0096-1523.35.2.551>

Henry, M. J., McAuley, J. D., & Zaleha, M. (2009). Evaluation of an imputed pitch velocity model of the auditory tau effect. *Attention, Perception, & Psychophysics*, *71*, 1399-1413. <https://doi.org/10.3758/APP.71.6.1399>

Huang, Y.L., Jones, B. (1982). On the interdependence of temporal and spatial judgments. *Perception & Psychophysics*, *32*, 7–14. <https://doi.org/10.3758/BF03204862>

Hubbard, T. L. (2017). Momentum in music: Musical succession as physical motion. *Psychomusicology: Music, Mind, and Brain*, *27*(1), 14–30.

<https://doi.org/10.1037/pmu0000171>

- Johnston, H. M., & Jones, M. R. (2006). Higher order pattern structure influences auditory representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 2–17. <https://doi.org/10.1037/0096-1523.32.1.2>
- Jones, B., & Huang, Y. L. (1982). Space-time dependencies in psychophysical judgment of extent and duration: Algebraic models of the tau and kappa effects. *Psychological Bulletin*, 91(1), 128–142. <https://doi.org/10.1037/0033-2909.91.1.128>
- Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology*, 53(1), 59-96. <https://doi.org/10.1016/j.cogpsych.2006.01.003>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Krishnan, L., Elhilali, M., & Shamma, S. (2014). Segregating complex sound sources through temporal coherence. *PLoS Computational Biology*, 10(12), e1003985. <https://doi.org/10.1371/journal.pcbi.1003985>
- Marty, N., Marty, M. & Pfeuty, M. (2021). Relative contribution of pitch and brightness to the auditory kappa effect. *Psychological Research*, 85, 55–67. <https://doi.org/10.1007/s00426-019-01233-y>
- 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(4), 1029–1039. <https://doi.org/10.1037/a0017601>



- Micheyl, C., Schrater, P. R., & Oxenham, A. J. (2013). Auditory frequency and intensity discrimination explained using a cortical population rate code. *PLoS Computational Biology*, 9(11), e1003336. <https://doi.org/10.1371/journal.pcbi.1003336>
- Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, 42, 135–159. <https://doi.org/10.1146/annurev.ps.42.020191.001031>
- Moore, B. C., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88(3), 320-333.
- Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 919-931. <https://doi.org/10.1098/rstb.2011.0355>
- Lu, K., Dutta, K., Mohammed, A., Elhilali, M., & Shamma, S. (2024). Temporal-coherence induces binding of responses to sound sequences in ferret auditory cortex. *bioRxiv*, 2024-05. <https://doi.org/10.1101/2024.05.21.595170>
- Price-Williams, D. R. (1954). The kappa effect. *Nature*, 173(4399), 363-364.
- R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
- Roy, M., Kuroda, T., & Grondin, S. (2011). Effect of space on auditory temporal processing with a single-stimulus method. In Strumillo, P. (Ed.), *Advances in sound localization*, (pp. 95-104). BoD–Books on Demand.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114-123.

<https://doi.org/10.1016/j.tins.2010.11.002>

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli.

*Perception & Psychophysics*, *40*, 9–19. <https://doi.org/10.3758/BF03207588>

Shigeno, S., & Oyama, T. (1983). Localization of speech and non-speech sounds. *Japanese*

*Psychological Research*, *25*(2), 112–117. <https://doi.org/10.4992/psycholres1954.25.112>.

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream

segregation. *Psychological Bulletin*, *133*(5), 780. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-2909.133.5.780)

[2909.133.5.780](https://doi.org/10.1037/0033-2909.133.5.780)

Stevens SS, Newman, EB. (1936). The localization of actual sources of sound. *The American*

*Journal of Psychology*. *48*(2), 297–306. <https://doi.org/10.2307/1415748>

Suto, Y. (1952). The effect of space on time estimation (S-effect) in tactual space. *Japanese*

*Journal of Psychology*, *22*(3), 189-200.

Szabó, B. T., Denham, S. L., & Winkler, I. (2016). Computational models of auditory scene

analysis: A review. *Frontiers in Neuroscience*, *10*, 524.

<https://doi.org/10.3389/fnins.2016.00524>

Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex

acoustic scenes based on temporal coherence. *Elife*, *2*, e00699.

<https://doi.org/10.7554/eLife.00699>

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy.

*Developmental Psychology*, 25(1), 122. <https://doi.org/10.1037/0012-1649.25.1.122>

van Noorden, L. P. A. S. (1975) Temporal coherence in the perception of tone sequences.

Institute for Perception Research, Technical Univ. Eindhoven.

<https://doi.org/10.6100/IR152538>

Vliegen, J., Moore, B. C., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *The Journal of the Acoustical Society of America*, 106(2), 938-945. <https://doi.org/10.1121/1.427140>