



Perceptual-learning evidence for inter-onset-interval- and frequency-specific processing of fast rhythms

Ruijing Ning¹ · Samuel J. Trosman¹ · Andrew T. Sabin¹ · Beverly A. Wright^{1,2,3}

Published online: 28 November 2018
© The Psychonomic Society, Inc. 2018

Abstract

Rhythm is fundamental to music and speech, yet little is known about how even simple rhythmic patterns are processed. Here we investigated the processing of isochronous rhythms in the short inter-onset-interval (IOI) range (IOIs < 250–400 ms) using a perceptual-learning paradigm. Trained listeners ($n=8$) practiced anisochrony detection with a 100-ms IOI marked by 1-kHz tones, 720 trials per day for 7 days. Between pre- and post-training tests, trained listeners improved significantly more than controls (no training; $n=8$) on the anisochrony-detection condition that the trained listeners practiced. However, the learning on anisochrony detection did not generalize to temporal-interval discrimination with the trained IOI (100 ms) and marker frequency (1 kHz) or to anisochrony detection with an untrained marker frequency (4 kHz or variable frequency vs. 1 kHz), and generalized negatively to anisochrony detection with an untrained IOI (200 ms vs. 100 ms). Further, pre-training thresholds were correlated among nearly all of the conditions with the same IOI (100-ms IOIs), but not between conditions with different IOIs (100-ms vs. 200-ms IOIs). Thus, it appears that some task-, IOI-, and frequency-specific processes are involved in fast-rhythm processing. These outcomes are most consistent with a holistic rhythm-processing model in which a holistic “image” of the stimulus is compared to a stimulus-specific template.

Keywords Temporal processing · Perceptual learning · Psychoacoustics

Introduction

Rhythm is defined as the pattern of time intervals in a stimulus sequence (Grahn, 2012). In hearing, rhythm forms the temporal architecture in music (Thaut, 2013) and conveys linguistic and emotional information in speech (Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011; Juslin & Laukka, 2003). Of interest here are rhythms with inter-onset-intervals (IOIs) shorter than 250–400 ms. Included in this short-IOI range are fast beat and sub-beat rates in music (London, 2012) and the typical syllabic rate for spoken English (Miller, Grosjean, & Lomanto, 1984). Rhythms in this short-IOI range appear to be processed by different mechanisms than slower rhythms

(Drake & Botte, 1993; Friberg & Sundberg, 1995; Hirsh, Monahan, Grant, & Singh, 1990; Kohno, 1992; Michon, 1964; ten Hoopen, Berg, Memelink, Bocanegra, & Boon, 2011; ten Hoopen et al., 1994, 1995). In the current study, we evaluated two rhythm-processing models as applied to sequences with short IOIs using a perceptual-learning paradigm.

The idea that fast rhythms are processed separately from slow rhythms initially came from examinations of the perception of isochronous sequences. These sequences comprise a series of sounds equally spaced in time. Such stimuli have the simplest possible rhythmic structure because rhythms are based on regular repeating beats referred to as their underlying isochrony (Matell & Meck, 2000; Merker, Madison, & Eckerdal, 2009; Ravignani & Madison, 2017; Treisman, 1963). In these studies, listeners detected either changes in the IOI of an isochronous sequence (tempo discrimination) or a single temporal displacement in a sequence that was otherwise isochronous (anisochrony detection). For both tasks, temporal sensitivity typically was captured by a constant threshold (Δt) for sound sequences with IOIs < 250–400 ms, but by a constant Weber fraction ($\Delta t/t$) for longer IOIs (~400–1,500 ms) (tempo discrimination: Drake & Botte, 1993; ten Hoopen et al., 2011; anisochrony detection: Hirsh et al., 1990; Michon, 1964; ten Hoopen et al., 1994, 1995; for review, see:

✉ Ruijing Ning
ruijingning2015@u.northwestern.edu

¹ Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

² Knowles Hearing Center, Northwestern University, Evanston, IL, USA

³ Northwestern University Institute for Neuroscience, Northwestern University, Evanston, IL, USA

Friberg & Sundberg, 1995). In addition, the developmental emergence of rhythm processing appears to occur earlier for sequences with short compared to long IOIs, because, in one investigation, 3- to 4-year-old children could tap in synchrony with a metronome at a fast rate (250-ms IOI), but were unable to do so at slower rates (500- and 1,000-ms IOIs) (Kohno, 1992).

It is typically held that rhythms with short IOIs (< 250–400 ms) are processed holistically, with all events in the sequence processed as a whole, and that rhythms with longer IOIs are processed analytically, with each event in the sequence processed separately (Hibi, 1983; Kohno, 1992; Warren & Ackroff, 1976). The results of two previous studies are consistent with this proposed holistic/analytic dichotomy. First, in an electrophysiological study (Sussman & Gumenyuk, 2005), listeners were presented with a sound sequence in which a higher frequency tone occurred every fifth tone in an otherwise fixed-frequency sequence. A mismatch negativity (MMN) was elicited by the “odd-ball” tones in these sequences for slow presentation rates ($\text{IOI} \geq 400$ ms), suggesting that the tones were perceived as single repeating tones. However, for a fast rate ($\text{IOI} = 200$ ms) no MMN was elicited, implying that the five-tone unit was formed in the neural representation. Second, in a memory-related study (Kohno, 1992), on each trial, participants were asked to memorize a multi-syllable nonsense word in which the syllables were separated by a fixed inter-syllable interval (ISI) and were tested on recall after a multiplication interference task. Recall was better when the ISI was short (250 ms) than when it was long (500 ms and 1,000 ms). The authors interpreted this outcome to be consistent with the idea that fast sequences are processed holistically, because multiplication is an analytic or cognitive task and only interfered with memorization when syllable presentation rates were in the putative analytic processing range, leaving the holistically processed fast sequences intact. These findings support the idea that fast sequences are processed holistically. However, neither the MMN study nor the memory study explicitly required rhythm processing. Thus, the processing of short-IOI sequences has yet to be examined with a traditional rhythm-processing task.

Here we evaluated rhythm-processing models in the short-IOI range ($\text{IOI} < 250\text{--}400$ ms) using a perceptual-learning paradigm with a rhythm-processing task: anisochrony detection. We trained listeners on anisochrony detection with a 100-ms IOI marked by 1-kHz tones for seven daily sessions and examined whether any influence of that training generalized to each of four untrained conditions: temporal-interval discrimination with the trained IOI and marker frequency (where “frequency” refers to the spectral frequency of the tonal markers) and anisochrony detection with an untrained IOI (200 ms) or untrained marker frequency (4 kHz and variable frequency). To assess the influence of the training, we compared the change in threshold between pre- and post-training tests for

the trained listeners to that of listeners who completed the pre- and post-training tests but did not receive any training in between (controls). We determined that the training induced learning on the trained condition, or generalization to untrained conditions, if, for a given condition, the threshold change was greater for the trained listeners than for the controls. The predicted pattern of generalization across the untrained conditions differs between the analytic and holistic processing models.

According to the analytic model, anisochrony detection occurs on the level of single events, through the comparison of the individual temporal intervals formed by the successive stimuli in the sequence (Keele, Nicoletti, Ivry, & Pokorny, 1989; Pashler, 2001; Rammsayer & Brandler, 2004; Schulze, 1978). This model holds that on the presentation of each marker in the sound sequence, its temporal distance from the previous marker is registered and analyzed and an anisochrony is detected if that distance differs from previous ones in the sequence. If training modifies an analytic processor, training-induced learning on anisochrony detection with a single IOI and a fixed marker frequency should generalize to the discrimination between the durations of individual temporal intervals (temporal-interval discrimination) for the trained interval, because the same interval comparison has been practiced during training. It also should generalize to anisochrony detection with untrained marker frequencies, because marker frequency does not affect the temporal structure. Finally, it should generalize to anisochrony detection with untrained intervals that are multiples of the trained interval, because such untrained intervals were present in the trained stimuli.

According to the holistic model, anisochrony detection occurs on the level of the stimulus as a whole, through the detection of departures from an isochronous template to which the sequence is matched (Hibi, 1983; Kohno, 1992; Warren & Ackroff, 1976). This model holds that on the presentation of the entire sound sequence, a holistic “image” is formed and compared to an isochronous template based on the standard stimulus, and an anisochrony is detected if the “image” departs from the template. If training modifies a template-based processor, the training-induced learning should not generalize to anisochrony detection with untrained IOIs or marker frequencies, because those sequences would be processed by templates that are matched to their standard IOIs and marker frequencies rather than to the standard IOI and marker frequency used during training. It also should not generalize to temporal-interval discrimination for the trained interval, because the template used during training would be matched to a sequence rather than just two tones.

Compared to controls, trained listeners improved on the trained anisochrony-detection condition, but did not generalize that learning to any of the untrained conditions, suggesting that the training modified a holistic processor of temporal sequences.

Method

Listeners

A total of 16 young adults (11 females) with a mean age of 20.3 years ($SD = 2.7$ years) participated in this investigation. Eight of the 16 participants served as trained listeners and the remaining eight served as controls. All listeners had normal audiograms as assessed during the familiarization session (see below). None had previous experience with psychoacoustic tasks. Listeners reported from 0 to 15 years of musical training ($M = 6.59$, $SD = 4.57$), with no difference between the trained and control groups ($M_{\text{Trained}} = 6.69$, $M_{\text{Control}} = 6.50$, $t(14) = 0.08$, $p = 0.94$). None of the listeners had any experience with percussion or were actively playing an instrument at the time of the study. Further, a cursory analysis suggested that there was no relationship between the amount of musical experience and starting thresholds on the anisochrony-detection task for these listeners. Listeners were paid for their participation. All procedures were approved by the Institutional Review Board at Northwestern University.

Experimental design

Conditions and stimuli

Listeners completed five conditions: four anisochrony-detection conditions and one temporal-interval discrimination condition. For the anisochrony-detection conditions, listeners were asked to discriminate between an isochronous stimulus (the standard) and an anisochronous stimulus (the signal). In the isochronous standard, six 15-ms pure tones were equally spaced from one another in time (t), creating five equal temporal intervals. In the anisochronous signal, one of the middle four tones, selected at random, was shifted forward in time. Thus, in the signal, one of the five intervals was slightly longer than usual ($t + \Delta t$) while the next interval was slightly shorter ($t - \Delta t$), creating a “jitter” in the otherwise regular rhythm. The signal and standard were presented in random order in two-presentation forced-choice trials. Listeners were instructed to indicate which stimulus was anisochronous, that is, which presentation contained the jitter, by pressing a key on a computer keyboard. Feedback (“CORRECT!” or “Wrong”) was provided after every trial.

For the temporal-interval discrimination condition, listeners were asked to discriminate a fixed time interval (t) marked by two 15-ms, 1-kHz pure tones (the standard) from a longer time interval ($t + \Delta t$) marked by the same two tones (the signal). The signal and standard were presented in random order in two-presentation forced-choice trials. Listeners were instructed to indicate in which presentation the two tones were separated by the longer time interval by pressing a key

on a computer keyboard. Feedback (“CORRECT!” or “Wrong”) was provided after every trial.

All conditions are denoted by the task (anisochrony detection or temporal-interval discrimination), the time interval (t) between successive tones in the standard stimulus, and the frequency of the tones that marked those intervals. The conditions included: (1) anisochrony detection 100 ms, 1 kHz; (2) anisochrony detection 200 ms, 1 kHz; (3) anisochrony detection 100 ms, 4 kHz; (4) anisochrony detection 100 ms, variable frequency (the frequency of each tone was randomly selected from a major musical scale); and (5) temporal-interval discrimination 100 ms, 1 kHz.

All stimuli were presented at 86 dB SPL. Each 15-ms pure tone was gated using cosine-squared rise/fall ramps of 5 ms. The time between successive tones (t) was measured from onset to onset. The stimuli were generated digitally using a digital-signal-processing board (TDT AP2; Tucker-Davis Technologies, Gainesville, FL, USA) and delivered to a 16-bit digital-to-analog converter (TDT DD1) followed by an anti-aliasing filter (8.5-kHz low pass; TDT FT5), an attenuator (TDT PA4), and a headphone buffer (TDT HB6) connected to the left earpiece of Sennheiser HD265 headphones in circumaural cushions. Listeners were tested in a sound-attenuated room.

Threshold estimation

For both tasks, Δt was adjusted adaptively using a three-down, one-up rule to estimate the discrimination threshold. Within each block of 60 trials, the Δt decreased after every three consecutive correct responses and increased after each incorrect response. Each trial at which the Δt switched from increasing to decreasing or vice versa was marked as a reversal trial and the Δt values on those trials recorded. The first three reversal values for each block were discarded, and the mean of the largest remaining even number of reversal values was calculated. This procedure yielded an estimate of the Δt value that the listener needed in order to achieve 79.4% correct performance (Levitt, 1971), referred to as the discrimination threshold. No estimate was computed from blocks with fewer than seven total reversals. The step size was 10 ms until the third reversal, and 1 ms thereafter.

Organization of the experiment

Listeners in the experiment were divided into two groups: a trained group and a control group. The trained group completed a familiarization session, a pre-training test, seven daily training sessions, and a post-training test. The familiarization session consisted of a set of two-presentation forced-choice auditory tasks to introduce listeners to the procedure. These tasks yielded detection thresholds for tones at the standard audiometric frequencies (0.25, 0.5, 1, 2, 4, and 8 kHz)

presented in quiet and for 1-kHz tones in forward- and backward-masking protocols using stimuli that were modeled after a previous report (Wright, Lombardino, King, et al., 1997). During the pre- and post-training tests, listeners completed five threshold estimates (300 total trials) per condition for the five conditions in random order: anisochrony detection 100 ms, 1 kHz; 200 ms, 1 kHz; 100 ms, 4 kHz; 100 ms, variable frequency; and temporal-interval discrimination 100 ms, 1 kHz. Each of the seven training sessions consisted of 12 threshold estimates (720 total trials) on the anisochrony detection 100 ms, 1 kHz condition. The control group completed the familiarization session, the pre-training test, and the post-training test, but did not participate in the training phase. Instead, the controls had a gap between the pre- and post-training tests that was approximately equivalent to the length of the training phase. The average number of days between the pre- and post-training tests was 17.6 days for trained listeners and 19.3 days for controls.

Analyses

Data set

Due to technical issues and procedural error, some data are missing. Two listeners (one trained and one control) were not tested at all on the anisochrony-detection 100-ms, variable-frequency condition. Another listener (a control) was not tested in the post-training test on the anisochrony-detection 200-ms, 1-kHz condition, so the pre-training data of that listener on that condition were omitted.

Two types of threshold outliers were excluded from the analyses. First, to reduce the influence of aberrant single-threshold estimates as obtained from a single block of trials, the most aberrant of the five (testing) or 12 (training) threshold estimates from a given listener in a given condition on a given day was excluded if that estimate was more than 20 ms higher or lower than the next nearest estimate in that set. Eleven out of 1,434 threshold estimates were excluded based on this procedure. Second, to reduce the influence of listeners with atypical performance on the pre-training test, all data from a given listener for a given condition were excluded if the mean of the five threshold estimates for that listener and condition at the pre-training test was outside of the 1.5 interquartile range from the median of individual pre-training mean thresholds in that condition. Data from only one listener (a control) in one condition (temporal-interval discrimination 100 ms, 1 kHz) were excluded based on this procedure.

Performance on the pre- and post-training tests

To examine pre-training performance, we performed a one-way repeated-measures ANOVA on the pre-training thresholds (see *Threshold estimation*) with condition as the within-

subject factor, followed by *post hoc* pairwise comparisons. We also performed Pearson correlations of the pre-training thresholds among all of the conditions.

To evaluate overall changes in performance, we compared pre-training thresholds to post-training thresholds both between and within groups, separately for each condition. Between groups, we computed 2-group (trained vs. control) by 2-time (pre- vs. post-training) ANOVAs on the pre- and post-training thresholds, with time as a repeated measure. We concluded that training affected performance if there was a significant interaction between group and time. Within groups, we analyzed the simple effect of time by performing pairwise comparisons. We concluded that performance improved/worsened if there was a significant simple effect of time.

Effect sizes are reported as η^2 for all ANOVAs, and as Hedge's g for all pairwise comparisons (Torchiano, 2017). All pairwise comparisons and correlations were corrected for multiple comparisons (Benjamini & Hochberg, 1995).

To determine whether the magnitude of threshold change was influenced by pre-training performance, for conditions in which the interaction between group and time was significant, we fitted a regression line to the trained-group data with pre-training threshold as the predictor of post-training threshold and compared the slope of the regression line to 1. We concluded that the threshold change was independent of the pre-training threshold if the slope did not significantly differ from 1.

To take individual differences in starting threshold into account, we also performed a between-group (trained vs. control) analysis of covariance (ANCOVA) on the post-training thresholds using pre-training threshold as the covariate, separately for each condition. The ANCOVA assumption of the homogeneity of regression slopes was met for all but one condition: anisochrony detection 200 ms, 1 kHz ($p = 0.02$; all other $p \geq 0.17$). We concluded that training affected performance if there was a significant main effect of group. The results of the ANCOVAs were consistent with the ANOVAs for all conditions, so are not reported in the results section.

Performance from the pre-training test through the training phase

To determine whether the trained group improved significantly on the trained anisochrony-detection condition from the pre-training test through the training phase, we fitted a regression line to the daily mean thresholds of all of the trained listeners against the log of the day number (pre-training test as day 1, training as day 2–8). We used a log transformation for the day number because plotting a power function (the typical shape of a learning curve; e.g., Ritter & Schooler, 2001) on a log scale yields a straight line. We also evaluated the improvement of each listener separately by fitting a

regression line to the daily individual threshold estimates of the listener against the log of the day number. We concluded that the group/listener improved monotonically if the regression line was significant and had a downward slope. For a subset of listeners with V-shaped rather than linearly decreasing learning curves, we segmented the learning curve for each listener into two parts at the day with the best performance during the training phase and performed linear regression on each segment on the log of the day number.

Results

Pre-training performance

The pre-training thresholds of the 16 listeners did not differ across IOI or marker frequency for the fixed-frequency anisochrony-detection conditions. The mean thresholds for anisochrony detection at 100 ms 1 kHz, 200 ms 1 kHz, and 100 ms 4 kHz, ranged from 20.5 to 21.2 ms (pairwise comparisons: all Hedge’s $g \leq 0.14$, $p \geq 0.50$) (Fig. 1). However, thresholds were higher for anisochrony detection 100 ms, variable frequency – mean threshold 32.1 ms – than for all of the other anisochrony-detection conditions (Hedge’s $g \geq 1.10$, $p \leq 0.01$). The threshold for temporal-interval discrimination 100 ms, 1 kHz was 24.4 ms, which was higher than for anisochrony detection 100 ms, 1 kHz (Hedge’s $g = 0.79$, $p = 0.01$), but lower than for anisochrony detection 100 ms, variable frequency (Hedge’s $g = 1.43$, $p \leq 0.01$). All other possible pairwise comparisons were not significant (Hedge’s $g \leq 0.43$, $p \geq 0.53$; all comparisons follow a one-way ANOVA condition main effect: $F(4,48) = 8.06$, $p < 0.01$, partial $\eta^2 = 0.40$). At the individual level, the thresholds in the four conditions with 100-ms IOIs correlated with each other at least

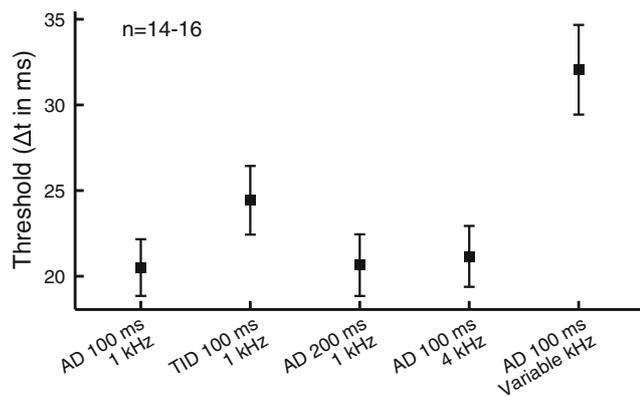


Fig. 1 Pre-training performance: Means. Mean increase in milliseconds in the IOI value required for 79% correct detections (threshold; Δt) for naïve listeners on the four anisochrony-detection conditions [100 ms, 1 kHz ($n=16$); 200 ms, 1 kHz ($n=15$); 100 ms, 4 kHz ($n=16$); 100 ms, variable frequency ($n=14$)] and one temporal-interval discrimination condition [100 ms, 1 kHz ($n=15$)]. The error bars represent ± 1 SEM. AD anisochrony detection, TID temporal-interval discrimination

marginally in five out of six pairs of conditions (all $r \geq 0.54$, $p \leq 0.06$), while the thresholds for anisochrony detection 200 ms, 1 kHz did not correlate with the thresholds for any other condition (all $r \leq 0.42$, $p \geq 0.15$) (Fig. 2).

Performance pre- to post-training

Trained condition

Trained listeners improved more than controls from the pre- to post-training test on the trained condition: anisochrony detection 100 ms, 1 kHz (ANOVA group \times time interaction: $F(1,14) = 5.57$; $p = 0.03$, $\eta_p^2 = 0.28$) (Fig. 3A, column 3). Trained listeners improved from the pre- to post-training test [mean difference (pre- minus post-training threshold) = 5.44, $p < 0.01$, Hedge’s $g = 1.41$], while controls showed no change (mean difference = 1.13, $p = 0.41$, Hedge’s $g = 0.28$) (Fig. 3A, column 4). The magnitude of improvement for the trained listeners was consistent across pre-training threshold (simple regression: $b = 0.91$, $R^2 = 0.69$, $p = 0.01$; comparison to slope of 1: $t(7) = -0.36$, $p = 0.73$) (Fig. 3A, column 5).

Untrained conditions

Learning on anisochrony detection 100 ms, 1 kHz did not generalize to temporal-interval discrimination, generalized negatively to anisochrony detection with the untrained IOI

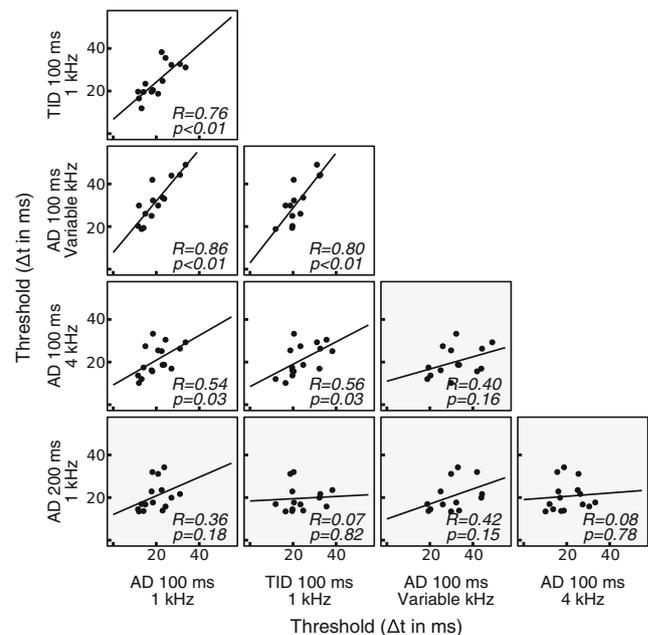


Fig. 2 Pre-training performance: Correlations. Scatterplots of pre-training thresholds (filled circles) between all pairs of conditions (panels). Regression lines are fitted to the data in each panel (solid lines) with Pearson’s R and the corresponding p -values, corrected for multiple comparisons, reported. The background color of each panel indicates whether the correlation is significant (white) or not significant (gray). AD anisochrony detection, TID temporal-interval discrimination

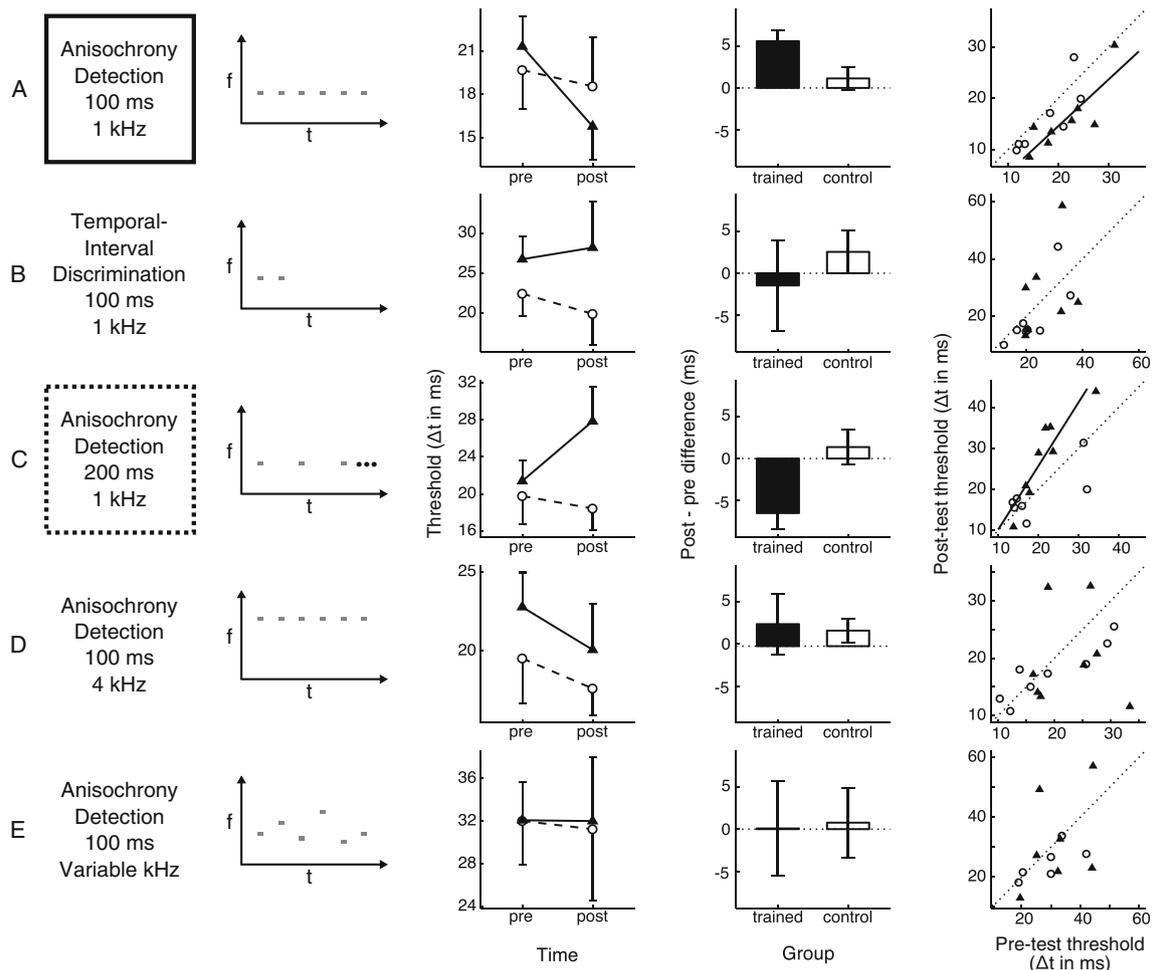


Fig. 3 *Learning and generalization.* **Column 1:** condition name: **A:** the trained condition; **B–E:** untrained conditions. **Column 2:** Schematic spectrograms of the standard stimuli. **Column 3:** Mean pre- and post-training test thresholds for trained listeners (filled triangles) and controls (open circles) in milliseconds. Error bars represent ± 1 SEM (across listeners), only one direction is displayed. **Column 4:** Average amount of improvement from pre- to post-training test for trained listeners (black bars) and controls (white bars) in milliseconds. Positive values indicate a threshold decrease, reflecting an improvement in task performance. Error bars represent ± 1 SEM (across listeners). **Column 5:** Individual

(200 ms, 1 kHz), and did not generalize to anisochrony detection with the untrained frequencies (100 ms, 4 kHz and 100 ms, variable frequency). For the untrained temporal-interval discrimination task, trained listeners performed similarly to controls (ANOVA group \times time interaction: $F(1,13) = 0.49$; $p = 0.50$, $\eta_p^2 = 0.04$), signifying a lack of generalization (Fig. 3B, column 3). On average, neither group improved or worsened from the pre- to post-training test (trained: mean difference = -1.45 , $p = 0.73$, Hedge's $g = 0.09$; control: mean difference = 2.53 , $p = 0.73$, Hedge's $g = 0.34$) (Fig. 3B, column 4). There also was no apparent trend for change among the individuals (Fig. 3B, column 5).

For the untrained IOI (200 ms, 1 kHz), the trained listeners actually showed significant worsening over controls (ANOVA

thresholds for trained listeners (filled triangles) and controls (open circles) for the pre- (x-axis) and post- (y-axis) training tests. The farther the symbols are from the diagonals (dashed lines), the greater the improvement (below the diagonal) or worsening (above the diagonal). For conditions in which the group \times time interaction is significant, the regression line of post-training thresholds on pre-training thresholds is displayed for the trained listeners (solid lines). Significant group \times time interactions are indicated by boxes around the condition names: trained group improved (solid-line box); trained group worsened (dashed-line box)

group \times time interaction: $F(1,13) = 7.43$; $p = 0.02$, $\eta_p^2 = 0.36$), indicating negative generalization (Fig. 3C, column 3). Trained listeners got worse from the pre- to post-training test (mean difference = -6.46 , $p = 0.01$, Hedge's $g = 1.10$), while controls showed no change (mean difference = 1.36 , $p = 0.53$, Hedge's $g = 0.23$) (Fig. 3C, column 4). The magnitude of worsening for the trained listeners was consistent across pre-training threshold (simple regression: $b = 1.55$, $R^2 = 0.83$, $p < 0.01$; comparison to slope of 1: $t(7) = 1.91$, $p = 0.10$) (Fig. 3C, column 5).

For the untrained frequencies (100 ms, 4 kHz and 100 ms, variable frequency), trained listeners performed similarly to controls (ANOVA group \times time interaction: $F \leq 0.04$, $p \geq 0.84$, $\eta_p^2 \leq 0.01$), denoting a lack of generalization (Fig.

3D and E, column 3). Neither group showed a change in threshold from the pre- to post-training test on average (100 ms, 4 kHz trained: mean difference = 2.70, $p = 0.51$, Hedge's $g = 0.25$; control: mean difference = 1.90, $p = 0.51$, Hedge's $g = 0.44$; 100 ms, variable frequency trained: mean difference = 0.08, $p = 0.99$, Hedge's $g < 0.01$; control: mean difference = 0.77, $p = 0.99$, Hedge's $g = 0.07$) (Fig. 3D and E, column 4), or a tendency for change at the individual level (Fig. 3D and E, column 5).

Learning curve

Although trained listeners improved more than controls on the trained condition (anisochrony detection 100 ms, 1 kHz) between the pre- and post-training tests, they did not, as a group, show monotonic improvement on that condition from the pre-training test through the training phase (linear regression over the log of the day number: $b = -0.45$ ms/log(day), $R^2 < 0.01$, $p = 0.72$) (Fig. 4). Only three of the eight trained listeners had learning curves with a significant downward slope (L1–L3; all $b \leq -2.39$, $R^2 \geq 0.17$, $p \leq 0.001$). Of the remaining five listeners, three had V-shaped curves (L4–L6). For these listeners, performance improved from the pre-training test (day 1) to the day with their best performance in the training phase ($b \leq -3.07$, $R^2 \geq 0.09$, $p \leq 0.05$), but then worsened to the last day of training (day 8) ($b \geq 7.60$, $R^2 \geq 0.06$, $p \leq 0.04$). Notably, for each of these listeners, the within-listener standard deviation was smaller on the day with the best mean performance during training than on most of the other training days, mirroring the pattern of the mean results. Of the other two listeners, one (L7) did not improve from the pre-training test to the day with the lowest threshold during training ($b = 0.85$, $R^2 \leq 0.01$, $p = 0.87$), and worsened after reaching the lowest threshold ($b = 9.84$, $R^2 = 0.08$, $p = 0.02$); the other (L8) had flat learning curves (both segments $-2.92 \leq b \leq 3.77$, $R^2 \leq 0.07$, $p \geq 0.11$). The mean pre-training test threshold was higher for the V-shaped group (17.9, 23.8, and 31.2 ms, averaging to 24.3 ms) than for the downward group (14.2, 18.6, and 27.2 ms, averaging to 20.0 ms) and the other listeners (15.1 and 22.6 ms, averaging to 18.9 ms). There were also fewer days on average between the pre- and post-training tests for the downward group (16, 16, and 17 days, averaging to 16.3 days) than for the V-shaped group (15, 17, and 22 days, averaging to 18.0 days) and the other listeners (20 and 21 days, averaging to 21.5 days).

Discussion

The primary purpose of this project was to evaluate two models of rhythm processing in the short-IOI range (IOI < 250–400 ms) using a perceptual-learning paradigm. Listeners who practiced anisochrony detection with a 100

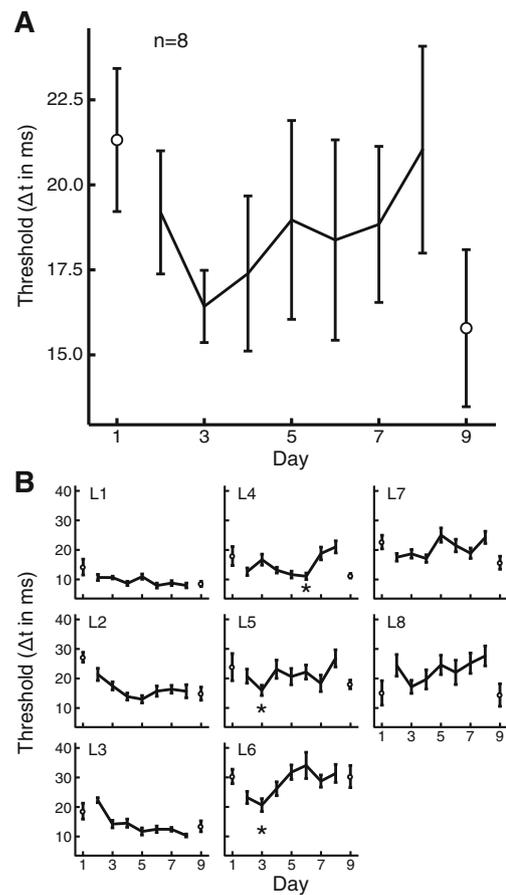


Fig. 4 Learning curves. **A:** Mean thresholds for the trained anisochrony-detection condition (100 ms, 1 kHz) across the testing (open circles) and training (filled circles) sessions. **B:** As in A, but for individual listeners (panels). For listeners with V-shaped curves from day 1 to day 8 (L4–L6), the star represents the day with the lowest threshold in the training phase. Error bars represent ± 1 SEM across subjects (**A**) and within subjects (**B**). Note that the ordinate range differs across **A** and **B**

ms, 1 kHz standard stimulus improved more between pre- and post-training tests than controls who completed the tests but received no intervening training. However, this learning did not lead to better performance on any untrained condition. The learning on the trained condition negatively generalized to the trained anisochrony-detection task with the trained marker frequency but an untrained IOI (200 ms, 1 kHz). It also failed to generalize to an untrained interval-discrimination task with the trained IOI and marker frequency (100 ms, 1 kHz), and to the trained anisochrony-detection task with the trained IOI but untrained marker frequencies (100 ms, 4 kHz and 100 ms, variable frequency). These results suggest that the neural circuitry modified by the training is selective for IOI, task, and frequency, and is competitive between different IOIs. This outcome is most consistent with the holistic model of rhythm processing.

The negative generalization from the trained anisochrony-detection task (100-ms IOI) to the same task with a different IOI (200 ms) (Fig. 3C) suggests that fast rhythms are

processed by an IOI-specific, even IOI-competitive mechanism. This idea gains further support from the observation that pre-training thresholds were correlated among the conditions with 100-ms IOIs (in five out of six cases), but not between those conditions and the condition with the 200-ms IOI (Fig. 2). As in previous investigations, the thresholds for anisochrony detection were constant across different IOIs in the short-IOI range in naïve listeners (~20 ms for 100-ms and 200-ms IOIs). The current data extend the idea that fast (< 250–400-ms IOIs; constant threshold) and slow (~400–1,500-ms IOIs; constant Weber fraction) rhythms are processed separately (see *Introduction*) by demonstrating that even within the short-IOI range the processing of different IOIs is separable.

The present demonstration of negative generalization from the trained 100-ms IOI to the untrained 200-ms IOI adds to a short list of other reports of negative generalization in perceptual learning (Fitzgerald & Wright, 2005; Regan & Beverley, 1985; Sabin, Eddins, & Wright, 2012). One possible explanation for the current case is based on the idea that listeners learned during training that critical information was provided every 100 ms over a period of ~500 ms because there were six tones in the sequence and the temporal position of the anisochrony was randomized. If so, the worsened performance on the untrained 200-ms IOI could have occurred either because the critical information did not occur at half of the time points when it was normally expected (during the first ~500 ms) or continued to occur after the last time point when it was normally expected (during the second ~500 ms) (for an example of the reduced detectability of signals presented at unexpected times, see Wright & Fitzgerald, 2004). Another possible explanation rests on the idea that listeners used different modes of temporal perception for the two IOIs because the 100-ms IOI fell below, but the 200-ms IOI fell within, the temporal limits of beat perception (London, 2012). If so, the worsened performance on the untrained 200-ms IOI could have occurred because learning to perform the task in one mode dampened the ability to perform the task in the other mode.

The lack of generalization from the trained anisochrony-detection task to the temporal-interval discrimination task (Fig. 3B) suggests that fast rhythms and short temporal intervals are processed by separate mechanisms. Consistent with this interpretation, in one report, performance on anisochrony detection and temporal-interval discrimination was uncorrelated (Rammsayer & Altenmüller, 2006). However, in other reports, performance was correlated between these two tasks (see Fig. 2; also, Rammsayer & Altenmüller, 2006; Rammsayer & Brandler, 2004), and the pattern of change in temporal sensitivity as a function of IOI was similar for both fast sequences and short single intervals (Drake & Botte, 1993), suggesting that temporal sequences and single intervals are processed by a common mechanism. It is noteworthy that

the different conclusions about the relationship between sound-sequence and single-interval processing appear to be associated with the level of musical expertise. The evidence that these two types of stimuli are processed by a common mechanism was obtained from listeners with either no (Rammsayer & Altenmüller, 2006; Rammsayer & Brandler, 2004) or only a moderate amount (the current study) of musical training (and from four listeners for whom the extent of musical training was not reported; Drake & Botte, 1993, their Experiment 1). In contrast, the evidence favoring separate mechanisms was obtained from highly trained musicians (Rammsayer & Altenmüller, 2006) and from listeners who underwent concentrated training on a specific sound-sequence task (the current study). This overall pattern raises the possibility that sequences and single intervals are initially processed by a common mechanism but are processed by separate mechanisms following training. Another possibility is that these two stimulus types are processed by separate mechanisms even initially, and that the correlation between them in naïve listeners with minimal to moderate musical training is either coincidental or reflects general aspects shared between the two stimulus types that are learned prior to tapping the stimulus-specific processes.

The lack of generalization from the trained anisochrony-detection task (100 ms, 1 kHz) to the same task with different marker frequencies (100 ms, 4 kHz or 100 ms, variable frequency) (Fig. 3D and E) suggests that fast rhythms are processed by a frequency-specific mechanism. Consistent with this interpretation, pre-training thresholds were higher for the variable-frequency anisochrony-detection condition than for all of the fixed-frequency conditions (Fig. 1). Other evidence for frequency-specific processing of sound sequences comes from reports that changing the pitch of one tone in a sequence of tones of otherwise constant pitch resulted in worsened precision of temporal judgments near the changed tone (Henry & McAuley, 2009; Hirsh et al., 1990; Lake, LaBar, & Meck, 2014). These outcomes support the idea that the temporal dimension is not singled out in rhythm processing, but rather is influenced by the pitch. Interestingly, anisochrony-detection thresholds were correlated between 1 kHz and 4 kHz and between 1 kHz and variable frequencies in naïve listeners (Fig. 2). If these correlations are not coincidental, they suggest that there are frequency-general components of the processing of fast rhythms in addition to the more primary frequency-specific aspects.

Models

Between the analytic and holistic models of rhythm processing, the present data are more consistent with the holistic model. According to the analytic model, anisochrony detection occurs on the level of single events and is based on the comparison of individual intervals. Therefore, learning on

anisochrony detection should have generalized to interval discrimination, and to anisochrony detection with untrained marker frequencies and with an untrained IOI that is a multiple of the trained IOI. None of these predictions was supported by the results. It is also noteworthy that the generalization pattern differed from that previously reported for learning on temporal-interval discrimination: generalization across frequency but not across IOI (e.g., Karmarkar & Buonomano, 2003; Wright, Buonomano, Mahncke, & Merzenich, 1997). This difference further supports the idea that anisochrony detection for the current short-IOI stimuli was not based on single intervals. In contrast, according to the holistic model, anisochrony detection is based on a comparison of a holistic “image” of the sequence with an isochronous template modeled after the standard stimulus. Therefore, the lack of generalization across IOI, frequency and marker number (task), three different departures from the standard stimulus, is as predicted by that model.

Summary and conclusions

- 1) We investigated the processing of fast rhythms (IOIs < 250–400 ms) using a perceptual-learning paradigm with a rhythm-processing task: anisochrony detection.
- 2) Prior to training, thresholds for anisochrony detection were consistent across different IOIs for stimuli with fixed-frequency markers and were higher for a stimulus with variable-frequency markers as well as for a temporal-interval discrimination task.
- 3) Pre-training thresholds were correlated among the conditions with the same IOI (in five out of six cases), but not between conditions with different IOIs.
- 4) Post-training thresholds on the trained anisochrony-detection condition (100 ms, 1 kHz standard stimulus) were lower for trained listeners compared with controls. The amount of improvement was constant regardless of pre-training threshold.
- 5) The learning on the trained anisochrony-detection condition did not lead to better performance on any untrained condition. There was negative generalization to an untrained IOI and no generalization to untrained marker frequencies or to an untrained temporal-interval discrimination task.
- 6) These results are most consistent with the holistic model of rhythm processing, suggesting that sound sequences in the short-IOI range are processed as a whole, and that anisochrony is detected by comparing the holistic “image” of the stimulus to an isochronous template that is selective for IOI, frequency and marker number.

Acknowledgments We thank Paul Reber and Sazzad Nassir for their helpful comments on a preliminary draft of this paper. This work was

sponsored in part by NIH/NIDCD, by the Defense Advanced Research Projects Agency (DARPA) Biological Technologies Office (BTO) TNT program under the auspices of Dr. Doug Weber and Tristan McClure-Begley through the Space and Naval Warfare Systems Center, Pacific Grant/Contract No. N66001-17-2-4011, and by a Northwestern University Undergraduate Summer Research Grant.

Disclaimer Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA) Biological Technologies Office (BTO).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 921–934. <https://doi.org/10.1037/a0021922>
- Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54(3), 277–286. <https://doi.org/10.3758/BF03205262>
- Fitzgerald, M. B., & Wright, B. A. (2005). A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *The Journal of the Acoustical Society of America*, 118(6), 3794–3803. <https://doi.org/10.1121/1.2074687>
- Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *The Journal of the Acoustical Society of America*, 98(5), 2524–2531. <https://doi.org/10.1121/1.413218>
- Grahn, J. A. (2012). Neural mechanisms of rhythm perception: Current findings and future perspectives. *Topics in Cognitive Science*, 4(4), 585–606. <https://doi.org/10.1111/j.1756-8765.2012.01213.x>
- 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), 551–564. <https://doi.org/10.1037/0096-1523.35.2.551>
- Hibi, S. (1983). Rhythm perception in repetitive sound sequence. *Journal of the Acoustical Society of Japan (E)*, 4(2), 83–95. <https://doi.org/10.1250/ast.4.83>
- Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics*, 47(3), 215–226. <https://doi.org/10.3758/BF03204997>
- Juslin, P. N., & Laukka, P. (2003). Emotional expression in speech and music. *Annals of the New York Academy of Sciences*, 1000(1), 279–282. <https://doi.org/10.1196/annals.1280.025>
- Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, 10(2), 141–147. <https://doi.org/10.1101/lm.55503>
- Keele, S. W., Nicoletti, R., Ivry, R. I., & Pokorny, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgements? *Psychological Research*, 50(4), 251–256. <https://doi.org/10.1007/BF00309261>
- Kohno, M. (1992). Two mechanisms of processing sound sequences. In *Speech Perception, Production and Linguistic Structure* (pp. 287–293). Amsterdam, the Netherlands: IOS Press.
- Lake, J. I., LaBar, K. S., & Meck, W. H. (2014). Hear it playing low and slow: How pitch level differentially influences time perception. *Acta*

- Psychologica*, 149, 169–177. <https://doi.org/10.1016/j.actpsy.2014.03.010>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477.
- London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford, UK: Oxford University Press.
- Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays*, 22(1), 94–103.
- Merker, B. H., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45(1), 4–17. <https://doi.org/10.1016/j.cortex.2008.06.011>
- Michon, J. A. (1964). Studies on subjective duration: I. Differential sensitivity in the perception of repeated temporal intervals. *Acta Psychologica*, 22, 441–450. [https://doi.org/10.1016/0001-6918\(64\)90032-0](https://doi.org/10.1016/0001-6918(64)90032-0)
- Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41(4), 215–225. <https://doi.org/10.1159/000261728>
- Pashler, H. (2001). Perception and production of brief durations: beat-based versus interval-based timing. *Journal of Experimental Psychology: Human Perception and Performance*, 27(2), 485. <https://doi.org/10.1037/0096-1523.27.2.485>
- Rammsayer, T. H., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal*, 24(1), 37–48. <https://doi.org/10.1525/mp.2006.24.1.37>
- Rammsayer, T. H., & Brandler, S. (2004). Aspects of temporal information processing: A dimensional analysis. *Psychological Research*, 69(1–2), 115–123. <https://doi.org/10.1007/s00426-003-0164-3>
- Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology*, 8, 1820. <https://doi.org/10.3389/fpsyg.2017.01820>
- Regan, D., & Beverley, K. I. (1985). Postadaptation orientation discrimination. *Journal of the Optical Society of America A*, 2(2), 147–155. <https://doi.org/10.1364/JOSAA.2.000147>
- Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In *International Encyclopedia of the Social and Behavioral Sciences* (Vol.13, pp. 8602–8605). Amsterdam, the Netherlands: Pergamon.
- Sabin, A. T., Eddins, D. A., & Wright, B. A. (2012). Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system. *Journal of Neuroscience*, 32(19), 6542–6549. <https://doi.org/10.1523/JNEUROSCI.5732-11.2012>
- Schulze, HH. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research*, 40(2), 173–181. <https://doi.org/10.1007/BF00308412>
- Sussman, E. S., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *Neuroreport*, 16(13), 1519–1523. <https://doi.org/10.1097/01.wnr.0000177002.35193.4c>
- ten Hoopen, G., Berg, S. V. D., Memelink, J., Bocanegra, B., & Boon, R. (2011). Multiple-look effects on temporal discrimination within sound sequences. *Attention, Perception, & Psychophysics*, 73(7), 2249–2269. <https://doi.org/10.3758/s13414-011-0171-1>
- ten Hoopen, G., Boelaarts, L., Gruisen, A., Apon, I., Donders, K., Mul, N., & Akerboom, S. (1994). The detection of anisochrony in monaural and interaural sound sequences. *Perception & Psychophysics*, 56(1), 110–120. <https://doi.org/10.3758/BF03211694>
- ten Hoopen, G., Hartsuiker, R., Sasaki, T., Nakajima, Y., Tanaka, M., & Tsumura, T. (1995). Auditory isochrony: Time shrinking and temporal patterns. *Perception*, 24(5), 577–593. <https://doi.org/10.1068/p240577>
- Thaut, M. (2013). *Rhythm, Music, and the Brain: Scientific Foundations and Clinical Applications*. New York, NY: Routledge. <https://doi.org/10.4324/9780203958827>
- Torchiano, M. (2017). *effsize*: Efficient effect size computation (R package version 0.7.1). Retrieved from <https://CRAN.R-project.org/package=effsize>
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the internal clock. *Psychological Monographs: General and Applied*, 77(13), 1–31. <https://doi.org/10.1037/h0093864>
- Warren, B. M., & Ackroff, J. M. (1976). Two types of auditory sequence perception. *Perception & Psychophysics*, 20(5), 387–394. <https://doi.org/10.3758/BF03199420>
- Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience*, 17(10), 3956–3963. <https://doi.org/10.1523/JNEUROSCI.17-10-03956.1997>
- Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*, 387(6629), 176–178. <https://doi.org/10.1038/387176a0>
- Wright, B. A., & Fitzgerald, M. B. (2004). The time course of attention in a simple auditory detection task. *Perception & Psychophysics*, 66(3), 508–516. <https://doi.org/10.3758/BF03194897>