
***This is a self-archiving copy and does not fully replicate the published version***

Auditory spectral versus spatial temporal order judgment: Threshold distribution analysis

Leah Fostick¹* and Harvey Babkoff²,³

¹ Department of Communication Disorders, Ariel University, Ariel, Israel
² Department of Psychology, Ashkelon Academic College, Ashkelon, Israel
³ Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel

Running Head: Spectral and spatial TOJ threshold distribution

*Corresponding author:
Leah Fostick
Department of Communication Disorders
Ariel University
Ariel, 40700
Israel
Tel. +972-54-4925354
Email: Leah.Fostick@ariel.ac.il
Auditory spectral versus spatial temporal order judgment: Threshold distribution analysis

Some researchers suggested that one central mechanism is responsible for temporal order judgments (TOJ), within and across sensory channels. This suggestion is supported by findings of similar TOJ thresholds in same modality and cross-modality TOJ tasks. In the present study we challenge this idea by comparing the threshold distributions of the spectral and spatial TOJ tasks. In spectral TOJ, the tones differ in their frequency (“high” and “low”). In spatial (or dichotic) TOJ, the two tones are identical but are presented asynchronously to the two ears and thus differ with respect to which ear received the first tone and which ear received the second tone (“left”/“right”). Although both tasks are regarded as measures of auditory temporal processing, a review of the literature suggests that they trigger different patterns of response. The present study is based on 388 participants in 13 spectral TOJ experiments, and 222 participants in nine spatial TOJ experiments. None of the spatial TOJ distributions deviated significantly from the Gaussian; while all of the spectral TOJ threshold distributions were skewed to the right, with more than half of the participants accurately judging temporal order at very short inter-stimulus intervals (ISI). The data do not support the hypothesis that one central mechanism is responsible for all temporal order judgments. We suggest that spectral TOJ paradigm may provide the opportunity for two-tone masking or temporal integration, which is sensitive to the order of the tones and thus provides perceptual cues that may be used to judge temporal order. This possibility should be considered when interpreting spectral TOJ data, especially in the context of comparing different populations.

Keywords: Temporal order judgment (TOJ), Distribution analysis

Public Significance Statement:

In the present study we found that spectral and spatial TOJ thresholds distributions differed significantly. While spatial TOJ distribution were Gaussian, spectral TOJ distributions were “bowl” shaped. These findings argue against the hypothesis that one central mechanism is
responsible for all temporal order judgments (TOJ) and suggests that different perceptual strategies are employed when performing spectral and spatial TOJ
**Introduction**

Temporal order judgment (TOJ) reflects the individual's ability to perceive the temporal order of at least two short stimuli, presented rapidly one after the other. It has been a measure of temporal processing for more than a century (e.g., Gibbon, & Rutschmann, 1969; Hirsh, 1952, 1959; Hirsh & Sherrick, 1961; James, 1886; Pastore, 1983) and is used widely for measuring temporal processing among a variety of sub-populations, such as: 1) aging adults (e.g., Ben-Artzi, Babkoff, & Fostick, 2011; Fostick & Babkoff, 2013a; Nishikawa, Shimo, Wada, Hattori, Kitazawa, 2015), 2) dyslexic readers (e.g., Ben-Artzi et al., 2005; Chen, Zhang, Ai, Xie, & Meng, 2016; Jaśkowski & Rusiak, 2008; Fostick, Bar-El, & Ram-Tsur, 2012a; Fostick, Eshcoli, Shtibelman, Nechemya, & Levi, 2014a; Laasonen, Service, & Virsu, 2001; Ortiz, Estévez, Muñetón, Domínguez, 2014; Reed, 1989; Tallal, 1980); 3) sleep deprived young adults (Babkoff, Zukerman, Fostick, & Ben-Artzi, 2005; Fostick, Babkoff, & Zukerman, 2014b); 4) students with ADHD (Kleiner, Negbi, Or, Zuaretz, & Fostick, 2011; Liddle, Jackson, Rorden, Jackson, 2009); and 5) patients with aphasia (e.g., Ro, Rorden, Driver, Rafal, 2001; von Steinbuechel, Wittmann, Strasburger, & Szela, 1999; Fink, Churan, & Wittmann, 2006a).

TOJ is measured in different modalities (e.g., Babkoff & Fostick, 2013; Ben-Artzi, Fostick, & Babkoff, 2005; Casarotti, Michielin, Zorzi, & Umiltà, 2007; Fink, Ulbrich, Churan, & Wittmann 2006b; Fostick & Babkoff, 2013a,b; Lee et al., 2013; Miyazaki, Yamamoto, Uchida & Kitazawa, 2006; Ram-Tsur, Faust, & Zivotofsky, 2006, 2008; Rayner, Lee, & Woodruff, 2015; Sambo et al., 2014), as well as cross-modality (Bertelson, & Aschersleben, 2003; Setti et al., 2014; Tinazzi, Fiorio, Bertolasi, & Aglioti, 2004). Some researchers suggest that one central mechanism controls the order judgment, regardless of the stimuli modality or presentation (Eijkman and Vendrik, 1965; Hirsh and Sherrick 1961; Kristofferson, 1967a,b). The main evidence for a central mechanism was provided by the seminal work of Hirsh and Sherrick (1961). They showed that the order of tone pairs of different modalities (auditory, visual, and tactile), as well as the order of cross-modal stimuli
(e.g. auditory-visual, auditory-tactual, visual-tactual), was perceived with stimulus-onset asynchronies (SOA) of 17 msec. Other researchers challenged the idea of one central mechanism being responsible for all temporal order judgments, because such a mechanism would require dividing attention among the different sensory channels. The implied ability of such a central mechanism to perceive the order of information flow from different sensory channels with the same ISI of 17 msec, as well as the known limitations on divided attention between sensory channels, makes the idea of a single central time order processor difficult to accept (Sternberg and Knoll, 1973; Stroud, 1955).

The results of more recent studies also seem to challenge the idea of a single central mechanism of temporal order. Studies of auditory temporal order mainly use auditory spatial (dichotic) TOJ and spectral TOJ (e.g., Ben-Artzi et al., 2005; Fink, Churan, & Wittmann, 2005; Fink et al., 2006a, b; Szymaszek, Szalag, & Sliwowska, 2006; Szymaszek, Sereda, Pöppel, & Szalag, 2009). Spatial TOJ refers to the successive presentation of two tones of the same frequency to the two ears, to which the participant responds as to which ear received the first and which ear received the second tone. Spectral TOJ involves the presentation of two tones of different frequencies, either delivered to both ears simultaneously or to one ear alone; the participant responds as to the order of presentation of the two tones of different frequency.

If all judgments of temporal order are made by the same central processor regardless of their source or origin, both paradigms should provide similar results. Indeed, Hirsh & Serrick (1961) reported finding the same SOA threshold (17 msec) both when the stimuli were high and low tones and when the stimuli were presented asynchronously to left and right ear. However, when examining participants’ response patterns other than just their mean threshold, we reported that each of the paradigms produced a unique response pattern (Fostick & Babkoff, 2013b). For example, very few participants are able to perform spatial TOJ with stimulus onset asynchrony (SOA) as short as 20 msec. However, 46% of the participants succeeded in doing so when performing spectral TOJ (see Fostick & Babkoff, 2013b, Experiment 1). In terms of the overall performance of a group of participants, there is a
gradual increase in the group accuracy of spatial TOJ accuracy with an increase in SOA, from approximately 55% at SOA=20 msec, to almost 100% with SOA=255 msec. In contrast, the group performance in spectral TOJ is not affected much by changes in SOA and begins at around 75% with SOA= 20 msec to around 85% with SOA= 255 msec (Fostick & Babkoff, 2013b, Experiment 1). Furthermore, when performance accuracy is plotted against SOA for spectral TOJ, only 54% of the variance is accounted for by SOA. A similar plot of accuracy versus SOA accounts for 97% of the variance when performing spatial TOJ (Fostick & Babkoff, 2013b, Experiment 2). Differences in response patterns between spectral and spatial TOJ can also be seen when examining data reported in other studies (e.g., Fink et al., 2005, 2006b; Szymaszek et al., 2009).

The debate as to whether there is one central mechanism or different mechanisms for the judgment of auditory temporal order has not been aided by reference to neurophysiological data. It seems that both spectral and spatial TOJ activate areas around the left temporo-parietal cortex and posterior sylvian regions (Bernasconi, Grivel, Murray, & Spierer, 2010; Temple, 2002; Temple et al., 2003). However, additional areas were found to be associated with each of the two paradigms. For example, the left inferior frontal gyrus, the right frontal temporal, and the anterior cingulate gyrus were found to be related to spectral TOJ (Temple, 2002; Temple et al., 2003), while the bilateral fronto-parietal network, including the left prefrontal cortex, left parietal lobules, and left occipito-temporal regions were found to be related to spatial TOJ (Binder, 2015).

The aim of the current study was to challenge the hypothesis of a single central mechanism for judging temporal order, by comparing and evaluating the response patterns of participants when performing spectral and spatial TOJ, and by focusing on the overall threshold distributions of spectral and spatial TOJ tasks. Previous studies did not report the shape of their TOJ threshold distributions (Fink et al., 2005, 2006a, b). However, some researchers reported non-Gaussian threshold distributions as justification for their choice of statistical tests (Fink et al., 2005, 2006b). Indeed, non-Gaussian distributions are usually
considered to be a technical problem, easily solved by using non-parametric tests to determine statistical significance. However, if the spectral TOJ threshold distributions are consistently found to be non-Gaussian across different studies, the likelihood increases that the non-Gaussian form of the distribution may be a characteristic of the task performance rather than a minor technical problem.

We analyzed data collected in 13 spectral TOJ experiments and 9 spatial TOJ experiments in order to determine the threshold distributions across a variety of stimulus parameters. Concomitant with recent discussions in the psychological literature regarding difficulties arising from the exclusive use of the first and second moments of the Gaussian distribution to analyze data (Doherty, Shemberg, Anderson, & Tweney, 2012; Maner, 2014; Perezgonzalez, 2015; Schneider, 2015; Speelman & McGann, 2013; Trafimow, 2014), and in an attempt to examine the entire range of the threshold data, we present and analyze the threshold distributions of 13 spectral and 9 spatial TOJ experiments.
SERIES I: SPECTRAL TOJ

Method

Participants: The characteristics of the participants in each of the 13 experiments are presented in Table 1. All of the experiments included 21 to 53 healthy participants, similar to the number tested in published TOJ studies (e.g., Ben-Artzi et al., 2005; Fink et al., 2005, 2006a, b; Fostick & Babkoff, 2013a, b; Fostick et al., 2012a, b, 2014a, b; Szymaszek et al., 2006, 2009). The ages of the participants ranged between 20 and 35 years, and all were native Hebrew speakers, with no history of learning disabilities, head injuries, neurological or psychiatric disorders, or usage of psychoactive substances. Only participants with hearing levels of at least 20 dB HL for .5, 1, 2, and 4 kHz were included in the experiments, and only those with no musical education or with some rudimentary musical education (6 months to 4 years) in their childhood (more than 10 years before the study). All participants were university students from a number of academic departments.

Stimuli and procedure: As noted, the spectral TOJ task consists of two pure tones of different frequencies presented binaurally (see Table 2 for the frequencies of the pure tones in each experiment). Participants were required to identify the order of presentation of the two tones ("high-low" or "low-high"). The order of the tones was randomly assigned across trials, and there were an equal number of high-low and low-high tone combinations in each session.

All experiments began with a four-stage training session. In the first stage, participants were familiarized with the high and low tones by listening to five examples of each. In the second stage, participants were presented with either a high or a low tone and were instructed to press the appropriate buttons to identify "high" and "low" accordingly. Each response was followed by feedback ("correct" or "incorrect"). In the third stage, the association between the tones and the response keys was tested. Participants were instructed to press the corresponding key for each tone. No feedback was provided. This stage continued until a
criterion of at least 20 correct responses out of 24 trials was achieved. All participants met the
criterion. The fourth stage included pairs of high-low and low-high tones, and the participants
were required to press the appropriate keys corresponding to the order of the tones. Each
response was followed by feedback ("correct" or "incorrect").

The stimulus characteristics for each of the experiments are presented in Table 2.
Experiment 9 was a replication of experiment 5 and therefore has the same stimulus
characteristics. The inter-stimulus intervals (ISI) separating the two tones of different
frequency ranged from 0-300 msec. The low frequency tone was always 1,000Hz, except for
Experiment 1 in which the low frequency tone was 300 Hz. The high frequency tone varied
across experiments and ranged from 600-3,500Hz (Table 2). Tone duration was 15msec for
experiments 1 and 6-13. For experiments 2-5, tone durations ranged from 5-30msec. Tone
intensity was either 60 dB SPL or 40 dB SL.

In experiments 6-13, intensity level (40 dB SL) was determined after measuring the
absolute hearing threshold for each stimulus for each participant in each of the experiments.
The absolute threshold was determined in a two-alternative, two-down-one-up adaptive
procedure. Three blocks were presented for each tone (first three blocks for 1,000 Hz and
following three for 1,800 Hz tone). In each trial, participants were presented with a tone and
were requested to report whether they heard it. Each block was terminated after 10 reversals
and the threshold for each block was calculated as the average of the last eight reversals.

In most of the experiments, spectral TOJ was determined by using the method of
constant stimuli, with ISI ranging between 2-240 msec (8-150 msec in Experiments 1 and 5-
200 msec in Experiment 10, Table 2). In these experiments, each ISI value was repeated 16
times, yielding between 128-288 trials per participant. The number of trials per participant
was determined by the number of ISIs used in the study (ISI values X 2 tone pair
combinations X 16 trials). The trials were presented randomly in all of the constant stimuli
experiments, except for Experiment 13 in which ISI values were presented in blocks.
Experiments 6, 7, 8, and 11 were performed with the 2-down-1-up adaptive procedure, in
which the ISI was reduced following two consecutive correct responses and increased following one error. The initial ISI value was 300 msec and the steps were 25 msec for the ISI range of 100 – 300 msec; 10 msec for 50 – 100 msec; 5 msec for 15 – 50 msec, and 2.5 for an ISI range of 0 – 15 msec. The experiment was terminated after 10 reversals (see also Fostick et al., 2013b, 2014a,c).

When using the constant stimulus method, spectral TOJ threshold was calculated as the ISI corresponding to 75% accuracy, after fitting psychometric functions to the data. When using the adaptive procedure, the threshold was calculated as the average of the last eight reversals (out of 10).

**Apparatus:** Experiments 1-4 were conducted using a 486 personal computer and experiment 5 with a Pentium I personal computer. These computers controlled the stimulus presentation and recorded responses. In experiments 1-5, tones were generated by a sound-generator device (TDT-system II: Tucker-Davis Technologies, Gainesville, FL). The acoustic stimuli in experiments 6-13 were delivered using Samsung NP350U2A laptop computer with Realtek High Definition Audio chip set. In all experiments, the sounds were presented binaurally through TDH49 headphones.

Experiments 1-4 were programmed using C programming language, and 5-13 using Matlab™ software version 6.5 that also generated the acoustic stimuli in experiments 6-13, controlled their presentation, and recorded responses. Screening for hearing sensitivity was performed using Danplex DA64 or Maico Hearing Instruments Ltd MA32 audiometers.

All experiments were performed in a sound-attenuated room, with attenuation of 6 dB.
Results

The results of all 13 experiments are presented in Table 3. Data from Experiments 2-5 were published previously (Fostick & Babkoff, 2013a, b), but included only accuracy data and not thresholds. Thresholds could not be computed for most of the participants in the studies: Of all participants in the 13 experiments (n=388), 49% of the participants (N=190) had either an accuracy level higher than 75% at the shortest ISI value used for that experiment (2, 5, or 8 msec) with the method of constant stimuli, or correct responses at ISI=0 msec with the adaptive procedure. For these participants the threshold was estimated to be the smallest SOA (15 msec in experiments 1, 5-13; 5 msec in experiment 2; 10 msec in experiment 3; and 20 msec in experiment 4). Eighteen percent of participants (N=68) had accuracy levels less than 75% even with the longest ISI value (150, 200, or 240 msec) when using the method of constant stimuli, or had incorrect responses at ISI=300 msec when using the adaptive procedure. For these participants the threshold was estimated to be the largest SOA (165 msec in experiment 1; 245 in experiment 2; 250 in experiment 3; 260 in experiment 4; and 240 in experiments 5-13).

The distributions of spectral TOJ thresholds, i.e., the proportion of participants in each threshold bin, for each of the experiments are shown in Fig. 1. The thresholds of the participants in the first threshold bin (N=190) are shorter than 20 msec (SOA) and constitute the left mode of each of the distributions (Fig. 1). Thresholds that are 220 msec and longer (N=55) constitute the right mode of each distribution. The thresholds of the remaining participants (37%, N=143), whose performance enabled the classic calculation of SOA threshold, were categorized in bins of 20-50 msec; 50-90 msec; 90-150 msec; or 150-220 msec.

Inspection of Table 3 shows that the 13 spectral TOJ experiments produced a large range of mean thresholds (31.95 – 116.13 msec) and standard deviations (33.97 – 130.06 msec). The resultant grand mean of all the SOA thresholds across all 13 experiments (the average of all of the means in Table 3, weighted by their N) is 78.21 msec. Although the mode for most of the experiments was in the < 20 msec category (20-50 msec category for
Experiment 4 that had 20 msec stimulus duration, the mean threshold in each experiment was much longer. Indeed, Kolmogorov-Smirnov and Shapiro-Wilk tests were found to be significant for all experiments, thus rejecting the null hypothesis that the spectral TOJ threshold distributions of these 13 experiments are Gaussian (Table 3).
SERIES II: SPATIAL TOJ

Method

Participants: The characteristics of the participants in each of the 9 experiments are presented in Table 4. Each experiment included 17 to 59 healthy participants, aged 20 to 35, native Hebrew speakers with no history of learning disabilities, head injuries, neurological or psychiatric disorders, or usage of psychoactive substances. Only participants with normal hearing and who had hearing levels of at least 20 dB HL for .5, 1, 2, and 4 kHz were included in the experiments. Participants either had no musical education or had some rudimentary education (6 months to 4 years) in their childhood (more than 10 years prior to the experiment). All participants were university students from a variety of departments.

Stimuli and procedure: Two tones of the same frequency were presented asynchronously to the two ears (one to each ear) (see Table 5 for the frequency of the tones in each experiment). Participants were required to identify the order of the presentation of the tones to the two ears ("right-left" or "left-right"). The order of the stimulus presentation to the two ears was randomly assigned across trials, and there were an equal number of right-left and left-right combinations in each session.

Similar to spectral TOJ, all the experiments began with a four-stage training session. In the first stage participants listened to five examples of the experimental tone in each ear. In the second stage, they were instructed to press the correct key to identify the ear in which they heard the tone, followed by feedback. In the third stage, participants were tested to verify that they knew the association between the ear and the response key, and the fourth stage included pairs of “right-left” and “left-right” tones, to which they were required to respond.

The stimulus characteristics for each experiment are presented in Table 5. Experiments 7, 8, and 9 were replications of experiments 4, 5, and 6, and have the same stimulus characteristics. The two tones presented to the two ears were separated by ISIs that ranged
between 5-400 msec. The tone frequency was usually 1,000Hz, except for experiments 1 and 2 in which tone frequencies were 300 and 600Hz. Tone duration for all the experiments was 15 msec. Tone intensity was usually 40 dB SL, after measuring the absolute threshold for each participants, as described earlier. Tone intensity in Experiments 1 and 2 was 60 dB SPL.

Similar to the spectral TOJ experiments, most of the spatial TOJ experiments were performed using the method of constant stimuli, with ISI ranging between 5-400 msec and the number of trials per participant ranging between 192-512. The ISI values were presented randomly in all of the constant stimuli experiments, except for experiments 5 and 8 in which the ISI values were presented in blocks (randomized by block). Experiments 6 and 9 were performed using the 2-down-1-up adaptive procedure (see also Fostick et al., 2013b, 2014a, c); with an initial ISI value of 150 msec (the same steps and procedure were used as for spectral TOJ). Thresholds were also calculated using the same method described in the spectral TOJ series.

**Apparatus:** Experiments 1 and 2 were conducted using a 486 personal computer and experiment 3 with a Pentium 1 personal computer. In these experiments tones were generated by a sound-generator device (TDT-system II: Tucker-Davis Technologies, Gainesville, FL). The acoustic stimuli in experiments 4-9 were delivered using Samsung NP350U2A laptop computer with Realtek High Definition Audio chip set. Experiments 1 and 2 were programmed using C programming language, and 3-9 using Matlab™ software version 6.5 that also generated the acoustic stimuli in experiments 4-9, controlled the stimulus presentation, and recorded responses. Other apparatus were similar to those described in the spectral TOJ series.

All experiments were performed in a sound-attenuated room, with attenuation of 6 dB.
Results

Spatial TOJ threshold data from all of the experiments are presented in Table 6 and plotted in Fig. 2. Since most of the experiments yielded data in four category bins only (for most of them, there were no thresholds in the <20 msec and ≥220 msec bins), the Kolmogorov-Smirnov and Shapiro-Wilk tests could not be performed. Instead, we present the z values for Skewness and Kurtosis. The values for all experiments were within the range -1.96 < z < 1.96, indicating no significant deviation from the Gaussian distribution.

Data from experiments 1, 2, and 3 were published previously (Ben-Artzi et al., 2005; Fostick & Babkoff, 2013a). However, similar to the spectral TOJ data, these publications included only accuracy data and not thresholds. In contrast to spectral TOJ, all of the participants achieved at least 75% correct at the longest inter-stimulus intervals, so thresholds could be computed in the classic manner for all 222 individuals who participated in the nine spatial TOJ experiments. The proportion of participants in each threshold category is plotted for each of the nine experiments in Fig. 2. For comparison purposes, we used the same threshold category bins as for the spectral TOJ threshold distributions. Note that in contrast to spectral TOJ distributions (Fig. 1), the spatial TOJ distributions have neither a prominent left mode, nor a right mode. Half of the participants in all nine experiments (51%, n=114) had SOA thresholds between 50-90 msec; while 25% (n=55) had thresholds between 90-150 msec. A smaller group (14%, n=30) had spatial TOJ thresholds that ranged between 150-220 msec, and only very few had thresholds ≥220 msec (4%, n=8), 20-50 msec (3%, n=6), or <20 msec (4%, n=9).

In contrast to the spectral TOJ experiments that resulted in mean thresholds that ranged over 100 msec, the nine spatial TOJ experiments produced mean thresholds that ranged over less than 40 msec (56.84-93.23 msec), and much smaller standard deviations (20.32-57.57 msec). The overall spatial TOJ threshold mean of each the experiments is found within the category bin of the mode (50-90 msec) for that experiment, except for Experiment 3 whose mean threshold was slightly longer (93.23 msec). The resultant grand mean of all spatial TOJ SOA thresholds (the average of all of the means in Table 6, weighted by their N) is 78.34 msec. Since most experiments yielded thresholds in
only four category bins, Kolmogorov-Smirnov and Shapiro-Wilk tests could not be computed. However, Z scores for both Skewness and Kurtosis were within ±1.98, thus rejecting the alternative hypothesis that the distributions of all these nine experiments are not Gaussian (Table 6).

Figure 3 presents the threshold distributions for spectral and spatial TOJ across experiments. Comparing these distributions using Chi-square test for goodness-of-fit revealed significantly different distributions ($\chi^2_{15} = 723.124, p<0.0001$).


Discussion

Data from two series of experiments were presented above: 13 spectral TOJ experiments and 9 spatial TOJ experiments. The overall grand mean of the spatial TOJ and spectral TOJ thresholds were approximately the same (78.34 msec and 78.21 msec).

However, all 13 of the spectral TOJ threshold distributions were found to be skewed to the right, with the mode at very short thresholds (mostly < 20 msec) indicating that 49% of all of the spectral TOJ thresholds, expressed as SOA, were less than 20 msec. In contrast, spatial TOJ threshold distributions were always found to be normally distributed, with the mode and the mean in a threshold category bin of 50-90 msec.

In spite of repeated evidence for non-Gaussian distributions of spectral TOJ thresholds (Fink et al., 2005, 2006a, b), the meaning and implications of these skewed distributions have not been discussed. A close inspection of our Figure 1 shows that the skewness of spectral TOJ threshold distributions is mainly due to the large number of participants that have spectral TOJ thresholds shorter than 20 msec. Similar distribution, with a major left mode at the very short SOA was also presented by Fink et al. (2006b, Figure 1). Most researchers have reported that a minimum separation (in terms of either ISI or SOA) is required in order to correctly identify the order of the stimuli presented. The pioneering work of Hirsh (1959) and Hirsh and Sherrick (1961) had suggested that SOA threshold could be as short as 17 msec. Their results seems to be similar to the results we report for the spectral TOJ series of experiments. However, while in the spectral TOJ series the participants were untrained and only about half of them had short thresholds of < 20 msec, Hirsh (1959) and Hirsh and Sherrick (1961) reported their results based on 4 – 5 trained participants, so that no population TOJ threshold distribution is available from that period for comparison.

All of the recent studies have reported longer average spectral TOJ thresholds, approximately 30 – 60 msec (Babkoff and Fostick, 2013; Babkoff et al., 2005; Ben-Artzi et al., 2005; von Steinbüchel et al. 1999; Wittmann & Fink, 2004). Nevertheless, some researchers have reported that a fairly large number of participants can perform spectral TOJ
with very short SOAs separating the onsets of the two tones. Fostick and Babkoff (2013b) reported that 46% of participants (n=68) had accuracy levels higher than 75% with SOA=20 msec (stimulus duration = 15 msec) in a spectral TOJ task. Fink et al. (2006b) reported that more than 20 out of 50 participants had SOA thresholds of 11-20 msec (stimulus duration = 10 msec). High accuracy levels at relatively short SOA values were also reported by others (e.g., Ben-Artzi et al., 2005; Fink et al., 2005, 2006b; Szymaszek et al., 2009).

In comparison, for spatial TOJ, there are almost no reports of extremely short SOA thresholds (Fostick & Babkoff, 2013a; Fostick & Babkoff, 2013; Fink et al., 2005, 2006b). Fostick and Babkoff (2013b) reported only 3% of participants (two participants out of 68) who were able to correctly identify the order of the tones at SOA = 20 msec when judging spatial TOJ. Fink et al. (2006b) also reported that one participant (2% of the 49) had a spatial TOJ threshold that was estimated by the authors to be 11 msec.

Previous attempts to explain the ability of participants to perceive a sequence of short duration sounds were made by Warren and colleagues (e.g., Warren & Ackroff, 1976; Warren & Bashford, 1993). They suggested that participants used two different perceptual strategies to correctly judge the temporal order of the stimuli based upon the task demands: (1) holistic pattern recognition- in which participants perceive the elements in the sequence as one pattern and judge their order by comparing the overall patterns of the different sequences; or (2) direct identification of the elements in the sequence and their temporal order. According to Warren et al., a holistic pattern recognition strategy produces much shorter discrimination thresholds (e.g., SOA=5 msec per item in the sequence) than a direct identification strategy (e.g., SOA=200 msec per item, 600 msec for the entire sequence) (Warren & Ackroff, 1976). Thus, Warren et al. introduced the hypothesis that the judgment of temporal order can be performed using one of two different strategies. This explanation may be relevant to the performance pattern observed in the data of Series I (spectral TOJ): the ability of participants to perform accurately with very short SOAs (<20 msec) may indicate the use of holistic
pattern recognition, especially since the discrimination thresholds are consistent with those reported by Warren et al.

This hypothesis, however, does not fully explain the pattern of spectral TOJ threshold distributions. According to Warren et al., two different perceptual strategies are employed by participants when performing different types of temporal order tasks. Holistic pattern recognition is used by participants for tasks that require “same/different” judgments (when participants are asked to match the temporal pattern of sound sequences); while direct identification is used by participants when they are asked to reproduce the order of the sequence of sound stimuli (“Direct Naming of Order of Components”). However, in the experiments reviewed in the current paper, both for the spectral TOJ and the spatial TOJ tasks, participants were instructed to judge the temporal order of the two tones. Both tasks involved “Direct Naming of Order of Components” experiments and the data were collected in both sets of experiments by both the method of constant stimuli and the adaptive procedure. Consequently, the Warren et al hypothesis cannot explain either the very short thresholds found among 49% of participants in the 13 spectral TOJ experiments, or the absence of such short thresholds in any of the 9 spatial TOJ experiments.

Perhaps the difference in the thresholds found in the two auditory temporal order tasks lies in the differences in the very structure of the tasks. In spatial TOJ, the two tones are similar in all of their characteristics, and are delivered to different ears. Spectral TOJ involves the presentation of two tones of different frequencies. In both paradigms the tones are presented very rapidly. With very short ISIs the two tones can also mask each other and/or be perceived as a singular event. In spatial TOJ, the two tones are equal in all their characteristics. Therefore, if the tone in the right ear masks the tone in the left ear, it sounds the same as when the tone in the left ear masks the tone in the right ear. When the SOA is short, the availability of the temporal cue is eliminated and there is not any additional cue to perceive the order of the tones. Participants then find it difficult to perform the task, as reflected in low accuracy levels with short ISIs in spatial TOJ. However, in spectral TOJ,
since the tones in each pair are different in their frequency, the nature of the inherent masking paradigm may provide different perceptual outcomes when the low tone is presented before the high tone, than when the high tone is presented before the low tone.

One of the well-known phenomena of basilar membrane mechanics is that lower frequency tones mask higher frequency tones more than high frequency tones mask lower frequency tones (Ehmer, 1959; Gelfand, 2005; Mayer, 1894; Small, 1959). In addition, with very short ISIs (< 8 msec) backward masking is more effective than forward masking (Gelfand, 2005; Wilson and Carhart, 1971). Therefore, the high-low frequency pair may provide more efficient (backward) masking than the low-high pair (forward masking). Because of the different masking patterns and relative differences in the strengths of backward and forward masking when SOA < 20 msec, the two orders of tone frequency (low before high and high before low) may be perceived differently and consequently, be correctly identified as different.

Following a similar line of thought regarding the possibility of perceptual cues based on peripheral interactions between the two stimuli in spectral TOJ task, one might suggest that the difference between spectral and spatial TOJ may result from a difference between temporal integration in the same auditory channel (by stimulation of the same basilar membrane) versus temporal integration across the two auditory channels. Studies have shown that when two successive stimuli are presented, they can be combined together into an integrated singular percept by means of temporal integration (Tervaniemi, Saarinen, Paavilainen, Danilova, and Näätänen, 1994; Saija, Andringa, Baškent, and Akyürek, 2014; Yabe et al., 1998). Furthermore, because of frequency differences, temporal integration of the two stimuli in the spectral TOJ paradigm might be more potent than in the spatial TOJ paradigm where the frequency of the tones are the same. The possibility that masking and/or temporal integration may be providing perceptual cues in addition to the perception of temporal order suggests that we should be cautious when interpreting previous spectral TOJ data, especially when comparing the performance of young controls with that of populations
with suspected deficits. If masking cues or temporal integration play a role in the successful performance of spectral TOJ, then differences in performance among different populations must take this possibility into account as well as the possible differences in temporal resolution when attempting to interpret performance differences between populations.

The suggestion of the availability of different types of sequence perception, by the use of masking cues or temporal integration in the spectral TOJ paradigm but not in the spatial TOJ paradigm, argue against the idea of a singular central mechanism responsible for all temporal order judgments. Some researchers have argued against a single central mechanism for all sensory signals because of the limitations on divided attention across sensory channels (Sternberg and Knoll, 1973; Stroud, 1955). However, the results of the present study argue against a single central TOJ mechanism even within the same sensory channel. In the current and previous studies, we, as well as others (Fink et al., 2005, 2006a, b; Fostick & Babkoff, 2013b; Szymaszek et al., 2006, 2009) compared two TOJ tasks within the same sensory channel, the auditory modality. Although no information flow from different sensory channels was required for the judgment of temporal order, different response patterns were found. We interpreted the findings to mean that different stimulus paradigms (e.g., spectral VS spatial TOJ) may provide different types of information or cues that may be used by different central mechanisms to judge temporal order. Methodologically speaking, the results of the current study suggest that comparisons of TOJ data across different tasks and sensory channels without considering stimulus parameters and intra as well as cross-channel idiosyncrasies may not be valid. The suggestion that one central mechanism was responsible for temporal order judgments was based on a series of studies that reported mean TOJ thresholds only. However, examining TOJ threshold distributions in addition to reporting the mean threshold as we did in the current study provided information that questions the hypothesis of a single central mechanism for temporal order judgments.

Understanding the underlying mechanisms of the different response patterns in spectral TOJ performance is critical for the use of this task in auditory temporal processing evaluation
for groups suspected of temporal processing deficit. If normal controls can use different
strategies in performing the task, this information must be taken into account when comparing
the performance of normal populations with populations purported to be deficient in
discriminating temporal order (e.g., aging adults, dyslexic readers). Perhaps, the “deficit”
reflects a difficulty with the use of one performance strategy, but not with another. If
confirmed, this new insight may enable more accurate understanding of the nature of deficits
in at-risk populations.
References


Maner, J. K. (2014). Let’s put our money where our mouth is: If authors are to change their ways, reviewers (and editors) must change with them. *Perspectives on Psychological Science, 9*(3), 343-351.


