4aPPb6. Time-to-arrival discrimination of multiple sound sources

Michael S. Gordon*, Darlene Edewaard and Matthew Pacailler

*Corresponding author’s address: Psychology, William Paterson University, 300 Pompton Road, Wayne, NJ 07470, gordonm10@wpunj.edu

Previous research in auditory detection of time-to-arrival (TTA) has tended to focus on single sound sources approaching a listener. However, visual studies of TTA have suggested that there may be different perceptual strategies employed with respect to single versus multiple objects on approach. The current research is designed to directly address the capacity and informational support of listeners to make determinations on the TTA of multiple sound sources. Initial experimentation evaluated the capacity for discrimination between a stationary and moving sound source. Results suggest the difficulty discriminating between sound sources, with a slight advantage for more concussive spectra.

Published by the Acoustical Society of America through the American Institute of Physics
Time-to-arrival discrimination of multiple sound sources

The auditory system is able to serve as an early warning detector that can direct perceptual resources and attention to salient aspects of one’s environment. In that capacity the auditory system must distinguish between sounds within a cluttered acoustic scene for behaviorally-relevant information. Specifically one task of the auditory system may be to determine which of a set of sounds might be moving on a collision path towards a listener. A typical environment might contain a variety of acoustic stimuli, from vehicle engines to bird calls to train whistles. While anyone of these sounds might be relevant, it has been hypothesized that listeners are particularly sensitive to the rising intensity and change in the spectrum of sounds threatening collision (e.g., Neuhoff, 2001; Rosenblum, Carello, & Pastore, 1987). However, many sounds have inherent properties that cause them to have high levels of perceptual and behavioral urgency. Spectra that are louder, contain higher frequency energy, or fluctuate rapidly over time tend to be perceived as higher urgency, and consequently may draw attention (Arrabito, Mondor, Kent, 2004; Edworthy, Loxley, & Dennis, 1991). The current research was designed as an initial investigation into human sensitivity to approaching sound sources within a cluttered acoustic scene.

There have been a number of previous audio and visual studies on time-to-arrival (TTA). Past research has clarified how perceivers respond to a variety of velocities, angles of trajectory, and stimulus durations – in addition to examining the underlying visual and acoustic information that might support detection of TTA (Gordon & Rosenblum, 2005; Schiff & Oldak, 1990; Tresilian, 1991, 1999). Acoustic research on TTA has suggested the particular importance of rising intensity relative to other spectral factors to suggest the arrival of a stimulus (Lutfi & Wang, 1999; Neuhoff, 2001). With respect to the extant research, there is data that describes how listeners attend to various spectral properties of two approaching acoustic stimuli. It was found that amplitude was most influential to this judgment, although this finding was velocity dependent (Lutfi & Wang, 1999; Rosenblum et al., 1987). While this research is important at suggesting how listeners detect acoustic TTA, it does not address specifically how listeners are able to evaluate among a set of competing stimuli which sounds are approaching relative to other stimuli. The current research more directly addresses this issue.

In part the theoretical interest of attending to multiple sound sources in a TTA task relates to visual research on this issue. Previous visual studies of TTA have suggested that there may be different perceptual strategies employed with respect to single versus multiple objects on approach (DeLucia, 2005; DeLucia & Novak, 1997). While the related auditory experiments have not been tested, the reliance on similar and partially interchangeable sources of visual and auditory information in support of TTA (Gordon & Rosenblum, 2005) may indicate a similar alteration in perceptual strategies when listeners are asked to attend to multiple approaching sound sources. To test these ideas field recordings were captured of sound sources moving directly towards a microphone, as well as static recordings. Those recordings were paired together and played to participants. In the initial experiment listeners simply determined which sound source was moving towards them and which was static.
Methods

Participants. A total of 21 participants were recruited for this experiment from William Paterson University with an average age of 19.6 (SD = 3.2). Participation was either purely voluntary or completed for partial course credit. All participants reported normal hearing.

Materials & Apparatus. Acoustic stimuli were calibrated and presented using a Dell Optiplex 790 computer with an Asus XonarXD soundcard to Sennheiser HD280 pro headphones. Calibration was completed using an Extech HD600 sound level meter. Stimuli were presented using OpenSesame version 0.26 (Mathôt, Schreij, & Theeuwes, 2012).

To create the stimuli, sound samples were collected and normalized to a standard RMS level. The sound samples included: (1) 12-talker Babble, (2) Peacock, (3) Turkey, (4) Alarm Clock, (5) Repeating Gun Shot, (6) Loon, (7) Synthesized Shooting Star, (8) Ambulance Siren, (9) Thunder, (10) Fantasy Ting, and (11) Train horn. A sample of these spectra can be seen in FIGURE 1a-d.

![Figure 1](image)

**Figure 1.** Spectrograms of captured sounds: (a) 12-talker Babble, (b) Gun, (c) Peacock, and (d) Ting. Time is represented across the x-axis (approximately 5-sec each) and frequency on the Y-axis (0-5kHz). Regions of darker colors demonstrate higher energy portions of the spectra.

The sound samples were presented for a series of field recordings from a Pyle Pro PWMA60U portable address system that projected at 107 dBA from 1 meter. The amplifier was strapped to the front of a Suzuki GZ250 motorcycle at a height of approximately 0.8 meters. The motorcycle was positioned for stationary recordings at 2.5, 5, and 12 meters and at angles of 0°, 45°, and 135° relative to the recording device. For the moving TTA recordings the motorcycle coasted (with the engine off) towards the recording device at approximately 10, 25, and 40 km/h, and at an angle of 0°. The recordings were captured using a Neumann KU100 binaural dummyhead system to a Zoom H4n solid-state digital recorder.

After capturing these stimuli they were digitally edited to five seconds and set at a standard RMS level. Each moving stimulus was paired with a static 12-talker babble recorded at the three stationary distances. In addition, each static stimulus was paired with the babble recorded on a TTA trajectory. The babble was used as the standard of the dyad because it had the broadest spectra and was the only stimulus that used the human voice, making it somewhat distinctive.
Procedure. The participants initially completed an informed consent that described the TTA task. They were instructed to judge whether the babble (12-people talking simultaneously) or another sound was approaching them. Before the onset of the critical task participants completed a set of six test trials with feedback about whether they had correctly identified the approaching stimulus. The critical task consisted of 180 trials, including 18 of each sound paired with the babble. Judgments were indicated by pressing the “1” for babble or “0” for the other sound. Sounds were presented binaurally through the headphones and responses were recorded onto the computer. Total experiment duration was approximately 30 minutes per participant. At the end of the experiment, participants were debriefed as to the purposes of the experiment.

Results

To evaluate these TTA judgments, a 2x3 anova was completed using the two sound conditions (babble vs. other) and three speed conditions (10, 25, 40 km/h of the moving stimulus). Following the anova, an analysis determined the d’ (sensitivity) using each of the 10 sounds that were paired against the babble in trials. Prior to completion of the anovas, and with respect to the variability in performance abilities and biases, each participant’s data was converted to z-scores to give them a standard range and central tendency.

In the 2x3 anova directed at the moving sound’s speed: F(2, 40) = 3.71, p < .05, η² = .16; indicating that as speed increased participants were less accurate determining which of babble or other stimuli had been moving towards them. There was no reliable difference between whether the babble or another sound was moving: F (1, 20) = 0.49, p > .4, η² = .02. No interaction was found between speed and sound type: F (2, 40) = 1.72, p > .1, η² = .079. See Figure 2 of the speed x sound type to view these data.
Figure 2. Performance at determining which sound was approaching across the three speeds and when either the babble or the competing sound was dynamic. Z-scores near 0 were at chance levels.

The $d'$ analysis was computed for the 10 sounds tested against the babble: Alarm, Gun, Loon, Peacock, Shooting Star, Siren, Thunder, Ting, Train, and Turkey. Hits were analyzed as each judgment that categorized these sounds as moving when they did so. False alarms were each time these sounds were judged as moving when they were stationary and the babble was moving. The following Table 1 shows the $d'$ scores for each of these sounds.

Table 1. Sensitivity to the motion of the sounds tested against the babble as generated by the $d'$ analysis. The table shows $d'$ (overall sensitivity) and $\beta$ (bias) to judge these stimuli as moving.

<table>
<thead>
<tr>
<th></th>
<th>Alarm</th>
<th>Gun</th>
<th>Loon</th>
<th>Peacock</th>
<th>Shooting Star</th>
<th>Siren</th>
<th>Thunder</th>
<th>Ting</th>
<th>Train</th>
<th>Turkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d'$</td>
<td>0.03</td>
<td>0.3</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.21</td>
<td>0.21</td>
<td>0.17</td>
<td>0.28</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.05</td>
<td>-0.14</td>
<td>-0.04</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.15</td>
<td>0.08</td>
<td>0.01</td>
<td>-0.03</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen in this table, sensitivity to the movement of these sounds was fairly modest, and several of the sounds (Alarm, Loon, Train) could not be differentiated at better than chance levels. Performance was slightly better with the more concussive sounds (Gun, Peacock call, Ting) which tended to have very distinct onset/offset to their spectra. The bias ranged close to 0 for the sample.

Conclusions

This research suggests the difficulty participants had judging which of a pair of sounds was moving towards the listener and which was stationary. Each trial paired one of 10 unique (and non-speech) sounds against a 12-talker babble. Despite the practice trials and consistent exposure to the babble, performance was found to be near chance levels. There was a modest improvement among trials in which the moving sound was presented at the slowest speed (10 km/h) – irrespective of whether the babble or another sound was moving. This finding, while intriguing, is at odds with previous TTA studies which have shown that faster speeds support more accurate judgments (Gordon & Rosenblum, 2005; Schiff & Oldak, 1990).

There are several variables that may have contributed to these results. All stimuli were presented as approaching from the same trajectory (0°) allowing only limited and temporary spatial separation between sounds. In addition, while the babble was semantically unique (i.e., the only speech-based sound), it was also the broadest spectrum stimulus and thus produced consistent energetic masking across the range of included sounds. Hence the low signal-to-noise ratio of the approaching sound, and lack of spatial separation, seems to have produced a highly difficult set of trials. Future experiments will manipulate the signal-to-noise ratio between the stimulus pair and introduce a wider-range of trajectories. In addition, the amount of energetic masking will be manipulated to investigate how this variable specifically limits TTA sensitivity.
With respect to the d’ analysis and sensitivity to the TTA stimulus there were some notable differences between the stimuli. In particular it was found that more punctuated and concussive stimuli were easier to detect than others. This finding is indicative that it may be the temporal and dynamic changes in the stimulus – rather than the spectral composition of frequencies – that is most critical to determining TTA.

In conclusion, this research provided some initial data on the difficulties on discriminatory TTA judgments. As suggested with the visual stimuli (e.g., DeLucia, 2005; DeLucia & Novak, 1997), it may be that adding additional static or slower-moving sound sources may alter the perceptual strategies for TTA detection and success. However, the additional difficulty of energetic masking between multiple auditory stimuli makes TTA in this modality particularly challenging.

References