Perception, 1993, volume 22, pages 1467-1482

Auditory looming perception: Influences on anticipatory judgments

Lawrence D Rosenblum, A Paige Wuestefeld, Helena M Saldaña
Department of Psychology, University of California, Riverside, CA 92521, USA
Received 12 December 1992, in revised form 16 June 1993

Abstract. Several studies in the auditory-perception literature hint that listeners may be able to anticipate the time of arrival of an approaching sound source. Two experiments are reported in which listeners judged the time of arrival of an approaching car on the basis of various portions of its auditory signal. Subjects pressed a computer key to indicate when the car would have just passed them, assuming that the car maintained a constant approach velocity. A number of variables were tested including (a) the time between the offset of the signal and the virtual time of passage, (b) duration of the signal, and (c) feedback concerning judgment accuracy. Results indicate that increasing the time between signal offset and virtual time of passage decreases judgment accuracy whereas the actual duration of the signal had no significant effect. Feedback significantly improved performance overall.

1 Introduction

While a good amount of research has been conducted on the topic of visual looming perception (eg McLeod and Ross 1983; Schiff and Detwiler 1979), significantly less has been designed to test issues in auditory looming. It is known that blind individuals can interact successfully with approaching sound sources, in judging when it is safe to cross a busy intersection (eg Winer 1980). The extent of this ability is illustrated by blind athletes who can catch sound-emitting football, basketballs, and softballs, which allows them to participate actively in these sports (M Rosenblum, personal communication). Yet little empirical research has been conducted to determine how listeners determine when a sound source will reach them.

Of the few studies that have explored auditory looming, two have tested listeners' accuracy in judging the moment of passage of a simulated sound source (Rosenblum et al 1987; Warlick 1978). Rosenblum and his colleagues (1987) asked subjects to press a key at the moment they heard a simulated sound source just passing them. One outcome of this study was the suggestion that subjects were able to anticipate the moment of passage on the basis of the available acoustic information *before* the time of passage actually occurred. Rosenblum et al (1987) speculated that there might be prospective information in the earlier portions of a signal to specify the later time of passage.

This intuition has recently been given some support by Schiff and Oldak (1990). In one condition, Schiff and Oldak tested whether subjects could make anticipatory judgments of sound-source passage based on information from earlier portions of a looming trajectory. They edited a number of different recorded looming events so that the time between the offset of the stimulus and time of passage varied between 1.5 and 6.5 s. Subjects were asked to press a key when they thought that the sound source would have reached them, assuming a constant velocity. The results indicated that subjects could perform this task with some accuracy although most judgments were underestimated (ie subjects judged time of passage as being sooner than actual time of passage). In addition, as has been found with visual looming judgments (McLeod and Ross 1983; Schiff and Detwiler 1979), accuracy generally declined as the time between stimulus offset and (ostensive) time of passage increased. Finally, while Schiff and Oldak (1990) found that sighted subjects were generally better at visual than auditory looming judgements, congenitally blind subjects were as accurate in their auditory judgments as sighted subjects were with visual stimuli. This last finding suggests that experience might benefit auditory looming perception.

The fact that listeners can make anticipatory judgments of an auditory looming event suggests that there may be prospective time-to-arrival information in the acoustic signal similar to that thought to exist for visual looming. There is evidence that many anticipatory behaviors such as avoiding and catching looming objects are controlled by visual time-to-contact information (Lee 1976; Lee et al 1983; Savelsbergh et al 1991; Schiff 1965; Schiff and Detwiler 1979). This time to contact (τ_c) is thought to be specified in the inverse of the relative rate of optical dilation of a looming surface (Lee 1976, 1980) and is thus available during earlier portions of a looming event. The efficacy of the τ_c variable has been demonstrated in behaviors as diverse as the diving gannet timing its wing folding (Lee and Reddish 1981), the timing control of a runner traversing a cluttered terrain (Warren et al 1986), and the timing of a baseball batter's swing (DeLucia and Cochran 1985).

In a particularly thorough examination of visual τ_c-based behavior, Whiting and his colleagues (Whiting and Sharp 1974; Whiting et al 1970) tested the minimum visual information needed for catching a looming ball. Using various techniques, Whiting and his colleagues manipulated the duration and period during which the ball's trajectory was illuminated. They found that the ability to catch the ball successfully is directly related to the amount of visible trajectory (Whiting et al 1970). However, some subjects are still quite good when shown only the first half of the flight trajectory; this attests to the prospective nature of τ_c information. In another series of experiments, Whiting and Sharp (1974) asked whether the critical variable was the overall amount of visible trajectory or the time between cessation of illumination and initiation of ball-capture behavior—the occluded period (Whiting and Sharp 1974). To test this issue, Whiting and Sharp varied the occluded period between 0 and 320 ms while maintaining a constant period of illumination (80 ms). They found that successful ball catching increased as the occluded period decreased from 320 to 160 ms but then decreased as the occluded period varied from 160 to 0 ms. This suggests that more than just the amount of visible trajectory is important in τ_c judgments.

It has also been shown that similar visual information exists for perceiving the time at which a looming object will *bypass* an observer along near-miss paths (Young, as cited in von Hofsten and Lee 1985). Thus, it could be that time to contact (τ_c) is better construed as a more general time to arrival (τ_a) (Schiff and Oldak 1990).

To return to auditory looming, although the Schiff and Oldak study was an admirable first attempt at testing auditory τ_a, many questions are still unanswered. For example, it is unclear how much of an auditory looming trajectory listeners need to hear in order to perform accurate τ_a judgments. Furthermore, it has not been determined whether actually hearing the moment of passage helps in judgment accuracy. Finally, Schiff and Oldak's findings with blind individuals suggest that experience might help in this ability. However, no research has been conducted on whether short-term experience or judgment feedback can improve accuracy. In a series of two experiments, we examine these issues of auditory looming perception by implementing a methodology similar to that of Schiff and Oldak (1990) and that of Whiting and his colleages (Whiting and Sharp 1974; Whiting et al 1970).

To generate our auditory looming stimuli, we edited two recordings of a looming car so various amounts and segments of the signals were available to listeners. As in the Schiff and Oldak (1990) study, subjects were asked to press a key when they thought the car would be passing them, given constant velocity. In the first experiment, three questions concerning auditory looming information were explored. Specifically, we were interested in the influence on judgment accuracy of

(1) the occluded-period duration, (2) the audible signal duration, and (3) hearing the actual moment of passage.

The second experiment was designed to test the effect of feedback on listeners' judgment accuracy. We were interested in feedback for two reasons. First, as stated, Schiff and Oldak (1990) found that the experience of blind individuals enabled them to be significantly more accurate than sighted subjects at auditory τ_a judgments. It can be assumed that at least one way in which blind individuals fine-tune their skill is through feedback. Clearly, self-obtained feedback would be far too dangerous a basis for calibrating looming judgments of an oncoming vehicle. However, blind individuals might improve this skill through the verbal feedback provided by mobility trainers or by transferring experience from less threatening contexts in which obtaining (haptic) feedback would be feasible. Secondly, many successful visual τ_a judgments would seem to involve some form of feedback. Thus, the ball catching of Whiting's experiments, as well as the batting and running tasks mentioned earlier, are likely fine-tuned through haptic and visual feedback.

2 General recording and editing of stimuli

Recordings of an approaching car were made in the free field with a single unidirectional microphone (Sennheiser) and a high-quality monaural cassette recorder (Marantz #PMD221). (The events were recorded monaurally because of constraints on the computer-editing system described below.) The recordings were made from the edge of the road so that the car passed about 3 feet (0.9 m) from the microphone. The microphone was angled so that it was nearly parallel with the road, facing the oncoming car. More precisely, the angle formed by the microphone's direction and road was about 6° so that it pointed at a location about 30 feet from where the car actually passed the microphone. Recordings were made at both 15 and 25 mph(1) and the driver monitored that this velocity stayed constant during the approach. When the car traveled at 15 mph it traversed approximately 140 feet and when the car traveled at 25 mph it traversed approximately 200 feet. Twenty recordings were made and the two best recordings for each speed were chosen based on overall quality. Each of these two recordings was then sampled by a Compaq 386 computer at a sampling rate of 10 kHz. These recordings were low-pass filtered at 5 kHz.

The sound spectrograms of both signals are shown in figure 1. Each signal included the entire trajectory. The signal did not stop at the point where the car reached the microphone, but included a portion of the signal that came after passing the microphone. Including a signal portion after the time of passage ensured that, when hearing a trial which included the moment of passage, subjects' button-press judgments would not be based simply on hearing the signal end. The 15 mph signal was 6328 ms long while the 25 mph signal was 5578 ms long. In order to test anticipatory judgments, each of the recordings was divided into three temporally equal parts (all cut at zero crossings). This allowed us to easily deliver different signal portions and different combinations of these signals to subjects. This strategy of dividing the signal into thirds has been used successfully to determine the salient portions of a ball's visual trajectory in a baseball-batting task (DeLucia and Cochran 1985). Dividing the signal into thirds resulted in three stimulus portions of about 2109 ms each for the 15 mph recording and 1859 ms for the 25 mph recording.

The point at which the car reached the microphone (time of passage) was determined by the peak intensity of the signal and was checked against measurements of elapsed time performed during event recording. The time of passage for both signals is indicated on the spectrograms in figure 1. For the 15 mph signal the time of

(1) 1 mile h^{-1} (mph) ≈ 1.6 km h^{-1}.

passage was 714 ms from the onset of the third portion, and for the 25 mph signal it was 709 ms from the onset of its third portion. These edited stimuli acted as the basic building blocks for the stimulus presentations.

Dividing the signals into thirds enabled us to present seven different event types (per speed condition) to subjects (see figure 2). Types 1, 2, and 3 each contained only the first, second, and third portion (respectively) of the entire signal. Types 4, 5, and 6 each contained two thirds of the entire signal. Type 4 involved the first two thirds, Type 5 involved the last two thirds, and Type 6 involved the first and last third with the middle portion replaced by silence. Lastly, Type 7 was the entire signal.

The stimuli were presented diotically to subjects directly from the computer over high-quality headphones. The decision to present the stimuli over headphones rather than through a loudspeaker was based on a number of considerations. First, informal pilot work suggested that listeners found it somewhat easier to concentrate on the stimuli and task if headphone presentations were used. Second, although Schiff and Oldak (1990) used free-field, loudspeaker presentations for their experiment,

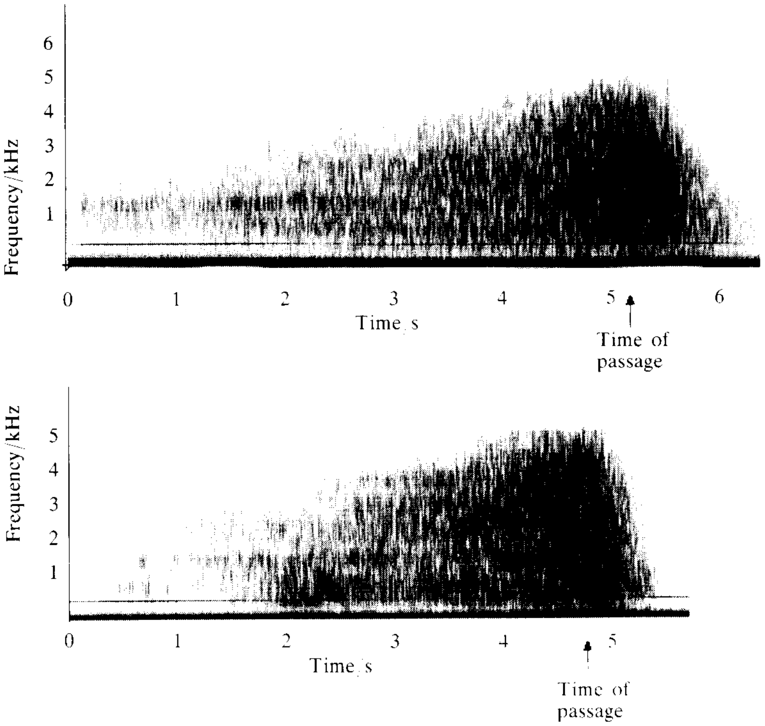

Figure 1. Sound spectrograms of the 15 mph signal (top) and 25 mph signal (bottom).

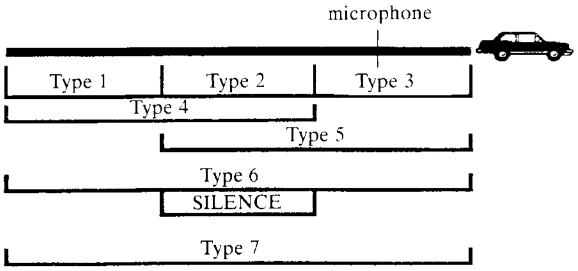

Figure 2. A pictorial representation of the stimuli used in both experiments (see text for details).

Rosenblum et al (1987) used headphone presentations and obtained systematic looming-judgment data. Third, presenting stimuli through a speaker would mean that, before reaching the subject's ear, the acoustic signal would be structured by the experiment room in addition to the environment under which the stimuli were recorded. It has been found that one way in which listeners determine sound-source (and, potentially, changing sound-source) distance is through the relative reflectances provided by a normal echoic environment (Mershon and Bowers 1979). Given that the echoic environment of our experiment room would modify the signal, using headphone presentations ensured that listeners heard a signal structured mainly by the acoustic environment present during stimulus recording.

The peak stimulus intensity at the headphones was 75 dB SPL for the 25 mph stimuli and 74 dB SPL for the 15 mph stimuli.

3 Experiment 1

The first experiment was designed to test listener τ_a accuracy without the benefit of feedback. To determine the salient aspects of the signal, three issues were tested: (1) does the duration of the occluded period (time between signal offset and ostensive time of passage) influence accuracy; (2) does hearing the actual time of passage improve judgment accuracy; and (3) does the duration of the audible looming signal influence accuracy?

3.1 *Method*

3.1.1 *Subjects.* Twelve undergraduate students participated in this study as partial fulfillment of a class requirement. All subjects reported having good hearing.

3.1.2 *Procedure.* Subjects were seated at a table in front of a Macintosh computer. They were told that they would be listening to recordings of a car coming towards them and that their task was to indicate at what point the car would have reached them. The experimenter then demonstrated with a small model car how they could imagine themselves standing by the side of the road facing an approaching car. Subjects were told that some of the recordings they would hear would be full recordings, whereas others would be partial recordings. Subjects were asked to respond in the same manner to both types of recordings and simply indicate at what point the car would have reached them, assuming a constant speed. Subjects started each presentation by pressing a key on the computer, and then indicated their judgment by pressing the key a second time. (This second key press did not stop the stimulus.) The computer was programmed to perform millisecond timing between key presses. After the judgment was made, there was a short delay in which the computer recorded the time between key presses. The subject was then able to initiate the next presentation. No feedback concerning accuracy was given during any part of this experiment.

The stimuli were delivered to subjects in blocks—each containing all seven signal types of a given approach velocity (15 or 25 mph). Blocking the trials in this way was based on memory constraints of the computer system. Within each block, each signal type was presented five times, yielding 35 trials which were randomized within the block. Two blocks were generated for each of the 15 and 25 mph stimulus series. There was counterbalancing for block-presentation ordering such that half of the subjects heard a block of 15 mph stimuli first (15, 25, 15, 25), and the other half received a 25 mph stimulus block first (25, 15, 25, 15). In all, subjects received 10 trials of each of the 14 conditions producing a total of 140 trials (7 signal types × 2 velocities × 10 trials each). Subjects received 6 practice trials before beginning in order to get used to the stimuli and the task. The 6 practice trials consisted of condition types 7, 4, and 6 from both the 15 and the 25 mph looming events. Subject participation lasted for about one hour.