IDENTIFICATION OF GATED ENVIRONMENTAL SOUNDS

Christiane Spanik and Kathleen Pichora-Fuller
University of British Columbia, 5804 Fairview Ave., Vancouver, BC, V6T 1Z3

INTRODUCTION

Environmental sounds are constantly present in our everyday lives, yet relatively little is known about how humans perceive such sounds and ascribe meaning to them (McAdams, 1993). Better understanding of how humans identify environmental sounds may have important implications for the design of machines that recognize and respond to specific auditory objects in complex scenes.

The gating paradigm has been very useful in research concerning the time course of word recognition (Grosjean, 1980, 1996). A meaningful sound sample is truncated into *gates*, or sub-samples, of varying duration. The first gate, or shortest sample, is presented and the listener attempts to identify the sound. Responses to each gate are recorded. The gate size is incremented until the sound is accurately and confidently identified. Misidentification errors for shorter gates are analyzed to determine the nature of the confusions that arose when the sample was insufficient for the listener to resolve perceptual ambiguities. By using this paradigm, the on-line auditory processing of environmental sounds could be compared to previous work on language processing.

METHOD

Participants. Sixteen normal-hearing listeners aged 20 to 35 years were paid to participate in the study. All had lived in the Greater Vancouver Region for at least two years.

Materials. The stimuli were eight recordings of natural environmental sounds selected from the Vancouver Soundscape Project library at Simon Fraser University (Truax, 1996). Four of the soundfiles consisted of high-context sequences of discrete or rapidly changing sound; for example, one soundfile consisted of the sound of the bus approaching and decelerating, braking, door opening, person stepping on steps, putting change in farebox, bus doors closing and engine revving as the bus drives away. The other four soundfiles consisted of low-context slowly changing or repetitive sounds; for example, one soundfile consisted of fizzling, crackling, and rumbling as the fire begins to burn with a slow increase in intensity as the fire burns more strongly. An auditory object occurring mid-soundfile was selected as the target to be identified (roughly analogous to a target word being selected from the middle of a sentence; e.g. Wingfield, 1996). For the four high-context files, the targets were: 1. change dropping into the bus farebox, 2. skytrain warning chimes, 3. computer drive booting up, 4. dot matrix printer printing. For the four low-context files, the targets were: 1. revving of motor cycle (Harley Davidson) engine, 2. ducks taking off from water, 3. fire crackling, 4. waves on a gravel shore. Pilot tests confirmed that the target auditory objects were easily identified in the intact soundfiles. A soundfile of squeaking door hinges was used for practice.

The smallest gate was a 400 ms sample centered on the target auditory object. Gate size could be incremented in either the preceding and following direction by progressively adding another 400 ms of the soundfile from the respective portion of the intact soundfile. Once all of the preceding gates were added, gate size continued to be incremented by adding the following gates until the entire soundfile was presented. Similarly, once all of the following gates were added, gate size continued to be incremented by adding the preceding gates until the entire soundfile was presented. The total duration of the intact soundfiles ranged from 10 to 40 seconds, with the average duration for the high-context sounds being 33 seconds and the average for the low-context sounds being 21 seconds.

Stimuli were prepared using Soundworks on a NeXT computer and converted from a 44 to a 20 kHz sampling rate for presentation using CSRE 4.5 (1995) software on a TDT system.

Conditions. Listeners were tested individually. Each listener attended two sessions, each lasting one to two hours. Hearing screening and the practice condition were completed before test conditions were administered. At each session, four soundfiles were presented, two high-context and two low-context. For each context type, one soundfile per session was presented with gates incrementing in the preceding direction and the other with gates incrementing in the following direction. The order of presentation of the eight soundfiles and the direction of the gate increments was counter-balanced such that each soundfile in each gating direction (8 x 2) was heard at least once in each order by one of the 16 listeners.

Procedures. Stimuli were presented binaurally at an average level of 70 dB SPL over TDH 39P earphones in a double-walled IAC sound-attenuating booth. After the presentation of each gate, the listener described what they thought they had heard. Participants were encouraged to guess and to give as much detail as possible. They also rated their confidence in their response on a scale from 1 to 10. Responses were recorded in writing by the experimenter throughout the experiment. Testing for a soundfile continued until the entire intact soundfile had been presented or until the listener identified the target correctly on five consecutive trials with confidence rated as 7 or greater.

RESULTS

Accuracy of sound identification. Four listeners identified all 8 targets correctly; ten listeners identified 7 correctly; one identified 6 correctly; one identified 5 correctly. Accuracy of identification of the sounds varied depending on the soundfile and gating direction (Figure 1). All listeners correctly identified high-context soundfiles 2 and 4. High-context soundfile 1 and low-context soundfile 4 were correctly identified by all listeners who heard preceding gates first, and low-context soundfile 1 was identified correctly by all listeners who heard following gates first. Performance was worse for other soundfiles. Performance was generally better for high-context compared to low-context soundfiles and for following compared to preceding gating, but the exceptions to this pattern suggest the importance of considering the unique properties of each soundfile that might have contributed to its identification.

Number of gates for identification. Considering each soundfile and each direction of gating, the time course of identification was considered for those listeners who accomplished correct identification. The median number of gates for accurate and confident identification are shown in Figure 2.

Figure 1: Number of listeners who correctly identified each the 8 environmental sound targets including 4 high-context targets and 4 low-context targets. Bars indicate the direction relative to the target in which the gates were incremented, dark bars for the preceding direction and light bars for the following direction.

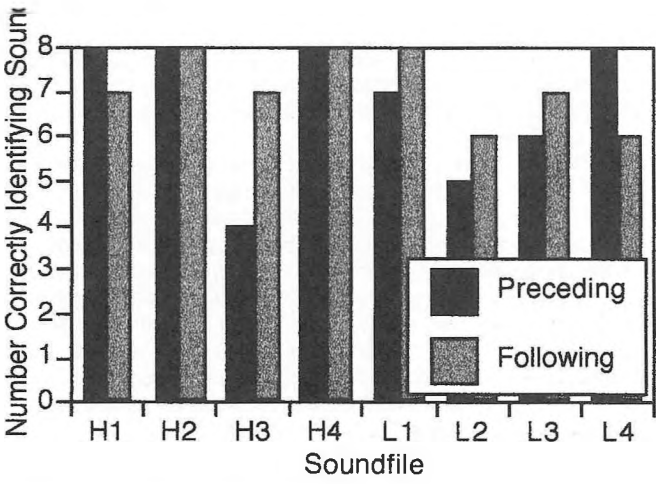
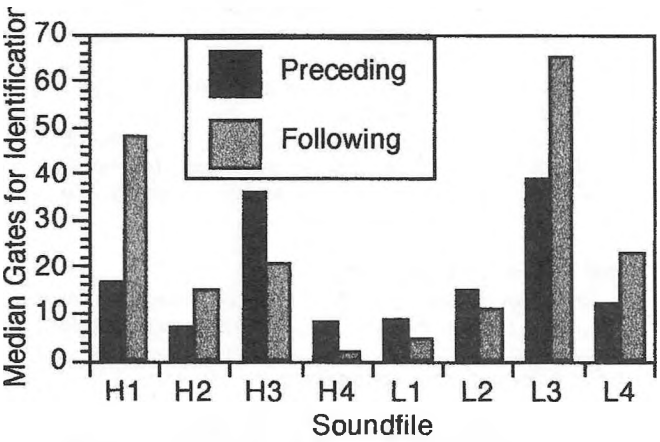


Figure 2: Median number of gates required for correct and confident target identification.



Overall, it is not surprising that few gates were required for listeners to identify the two high-context soundfiles that had been correctly identified by all listeners for both gating directions: 2 (skytrain chimes) and 4 (dot matrix printer). Likewise, few gates were required for listeners to identify the three soundfiles that were correctly identified by all listeners in one of the gating directions: high-context soundfile 1 (change in farebox) in the preceding direction, low-context soundfile 4 (waves on gravel) in the preceding direction, low-context soundfile 1 (motorcycle) in the following direction. However, few gates were also required for listeners to identify some of the files that were not correctly identified by all listeners: low-context soundfile 1 gated in the preceding direction, and low-context soundfile 2 (ducks) gated in both directions.

Error Analysis. The nature and frequencies of misidentifications were analyzed by listing all responses given for each target and counting the number of participants providing each response (Table 1). A large number of different responses were generated for each soundfile, with the majority of being idiosyncratic and with a much smaller set being generated by 3 or more listeners. The misidentifications were sometimes partially correct; for example, the listener identified ducks but did not specify the correct action of the ducks. Sometimes the misidentifications shared general semantic features with the target auditory object; for example, many subjects mentioned water for low-context soundfile 4 but did not mention waves on a beach. Sometimes misidentifications seemed to be acoustically rather than semantically based; for example, the most common misidentification for low-context soundfile 3 was 'rain' instead of 'fire'.

Table 1: Number of misidentifications for each soundfile.

Sound	Number of Different Incorrect Responses		
	> 3 Listeners	2-3 Listeners	1 Listener
H1	1	2	43
H2	3	8	21
H3	5	8	33
H4	2	7	17
L1	3	5	14
L2	5	9	17
L3	6	9	40
L4	5	13	18

DISCUSSION

The ability of listeners to identify environmental sounds increases as the number of gates are increased, but neither the amount nor the type of acoustical context surrounding the target sounds tested were related to identification of the target sound in a straightforward fashion. As suggested by Ballas (1993), performance seems likely to have been influenced by a variety of variables in different domains other than acoustics including the auditory objects' frequency, typicality, context independence, familiarity, and the availability of suitable linguistic labels. The pattern of misidentifications is reminiscent of the 'cohort' of words that listeners generate over the time course of word recognition (Marslen-Wilson & Tyler, 1980).

REFERENCES

Ballas, J.A. (1993). Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: Human Perception and Performance, 19(2), 250-267.

CSRE (4.2) (1995). Computer Speech Research Environment. London, Ontario: AVAAZ Innovations Inc.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. Perception & Psychophysics, 28(4), 267-83.

Grosjean, F. (1996). Gating. Language and Cognitive Processes, 11(6), 597-604.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71.

McAdams (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), Thinking in sound: The cognitive psychology of human audition, (pp. 146-198). Oxford: Clarendon Press.

Truax, B. (1996). Soundscape, acoustic communication and environmental sound composition. Contemporary Music Review, 15(1), 49-65.

Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. Journal of the American Academy of Audiology, 7, 175-182.