Effect of Event Variations and Sound Duration on Identification of Everyday Sound

James A. Ballas
Naval Research Laboratory
Code 5513
Washington, D. C. 20375-5337
Email: ballas@aic.nrl.navy.mil

Abstract

Experimental data on sound perception are presented on two issues: what aspects of a complex sound with several transients are important for accurate identification, and how much of a sound must be heard to identify it. The results of both experiments provide further insight into the properties of everyday sound that are related to good identification performance. The results of these studies support the notion that selection of sounds should be a careful design process.

1 Introduction

A synthetic sound ought to be unambiguous whenever it is used in an interface as an onomatopoeic device [1]. Perhaps the best way to guarantee this is to follow Rettig's [2] advice that professionals trained in creating sound and music be involved in the design process. Additionally there are empirical methods of assessing the ambiguity of a candidate sound [1, 3]. In the psychology of perception literature, there is a long history of research that follows the strategy of examining in detail selected sounds and determining the important features that people use for recognition [4]. Unfortunately, this research has been done for only a few sounds. Because the set of everyday sounds is immense, it is not likely that we will ever have a research base large enough to cover all the sounds of potential interest. What option is there to this pessimistic view? One option is to perform research on selected sounds with the intention of focusing on issues that might provide general insights. That is the motivation for the two experiments reported here.

2 Experiment One

The first experiment was conducted to examine further what parts of a sound contribute to accurate identification, or conversely, confusion and ambiguity. The sounds used were a light switch and stapler and were chosen because they include a pattern of impacts that is somewhat similar. Each sound was thought by some listeners in a previous experiment to be produced by the other cause. Multiple exemplars of each sound were produced by changing the instruments and circumstances of the event. The listeners were presented with each example and asked to identify it as a switch or stapler.

2.1 Method

2.1.1 Subjects

There were twenty participants all with reported normal hearing. All reported that they had heard the sounds of a stapler, pull-chain switch, and push dimmer switch.

2.1.2 Stimuli

Sixty stimuli were used: thirty stapler sounds and thirty switch sounds. One of the stapler sounds was used as a practice stimulus. The sounds were obtained under the conditions listed in Tables 1 and 2. Two of the sounds—switch #6 and stapler #17—were taken from a previous experiment [3]. The sounds were digitized at a 20 kHz sampling rate through a low pass filter set at 10 kHz. A tape was made by generating the sounds with a DAC set at a 20 kHz sampling rate through a low pass filter set at 10 kHz. The order of the stimuli was random. The tape recorder had a frequency response of $\pm$ 3 dB from 30 Hz to 15 kHz.

2.1.3 Procedure

Participants were told that they would hear a series of sounds each of which would be either a stapler or a light switch, the latter being either of the pull-chain or push-dimmer type. The staplers and light switches that had been used to produce the sounds were in view but the participants were not allowed to handle the objects, nor were any sounds produced by these objects prior to or during the experiment. The participants were informed that half of the sounds would be staplers and half would be light switches. Upon hearing each sound, they were asked to identify each sound using a 6-point scale that included a confidence rating:

Event:	Switch	Switch	Switch	Stapler	Stapler	Stapler
Confidence:	certainly	probably	possibly	possibly	probably	certainly
Scale code:	1	2	3	4	5	6

2.2 Results and Discussion

The average ratings for each of the stimuli are listed in Tables 1 and 2 together with standard deviations. The ratings for the two types of events were, on the average, significantly different ($t(58) = 5.49, p < .001$), although as shown in Figure 1, there was overlap in the ratings for the two types of events. The switches, especially the push dimmer switches, were sometimes thought to be stapler sounds. The stapler ratings were less dispersed than the light switch ratings. Whereas only three stapler stimuli were rated at two standard deviations from the mean of stapler sounds, 11 light switch stimuli were rated two standard deviations from the mean of switch sounds. These 11 outliers include all but two of the push-dimmer switch stimuli. Sounds produced by the smallest stapler were rated more like switch sounds than sounds from the two larger staplers.

The ratings for switch #6 and stapler #17 were examined in detail (see Table 3) because these two sounds had been used in a previous experiment with other sounds [3]. The average rating for switch 6 was closer to the appropriate end point of the scale than was stapler 17. This would mean that switch 6 was rated as a switch with greater confidence that was stapler 17 rated as a stapler. This outcome is consistent with the difference in accuracy found in the other experiment: switch 6 was identified more accurately as a switch than was stapler 17 identified as a stapler. However, other data for these two sounds are more similar. Both were identified in about 6 seconds, rather slowly, in the previous experiment. Both had about the same level

Sound No.	Stapler Type[a]	Production Properties[b]	Duration (ms)	Categorization Mean	SD
1	1	A	823	4.1	2.0
2	1	A	377	4.3	1.9
3	1	B	657	4.0	2.0
4	1	B	296	3.7	1.8
5	1	AC	667	5.3	1.0
6	1	AC	303	4.6	1.6
7	1	D	276	4.6	1.4
8	1	D	699	5.2	1.4
9	1	DC	642	4.5	1.4
10	1	DC	243	5.2	0.9
11	1	BF	507	3.4	1.7
12	2	A	923	3.9	1.7
13	2	A	357	4.6	1.4
14	2	C	703	5.0	1.6
15	2	C	567	4.3	1.8
16	2	D	648	4.2	1.7
17	2	D	282	3.7	1.8
18	2	C	684	5.0	1.1
19	2	C	349	4.9	1.5
20	2	E	465	4.9	1.6
21	3	A	974	2.8	1.8
22	3	A	207	3.4	1.9
23	3	C	706	2.7	1.5
24	3	C	421	3.7	1.8
25	3	D	848	3.7	1.8
26	3	D	259	4.6	1.3
27	3	C	779	3.7	1.5
28	3	C	272	4.8	1.5
29	3	BG	599	2.8	1.8
30	3	BG	183	2.9	1.9
Mean				4.15	
Median				4.25	1.65
Interquartile range				.65	.28

Table 1: Production Characteristics and Response Categorization for 30 Stapler Sounds. [a]Stapler Type 1 was a medium sized plastic-cased stapler, (15.2 cm by 5.1 cm); Type 2 was a metal stapler, (20.3 cm by 7.6 cm); and Type 3 was a small metal stapler, (10.2 cm by 5.1cm). [b]Production characteristics: A = Press on wood desk; B = Press in hand; C = With paper; D = Press into foam; E = Press on metal wall; F = No base; G = With base.

Sound No.	Switch Type[a]	Production Properties[b]	Duration (ms)	Categorization Mean	SD
1	1	A	1478	2.6	1.9
2	1	A	319	1.8	1.0
3	1	B	679	4.0	1.9
4	1	B	604	3.9	1.7
5	2	A	707	1.4	0.7
6	2	A	360	1.2	0.7
7	2	C	701	1.6	1.4
8	2	C	317	1.9	1.3
9	3	A	754	1.5	0.9
10	3	A	295	1.4	0.8
11	3	C	694	2.2	1.6
12	3	C	326	1.6	1.1
13	4	A	524	4.9	1.7
14	4	A	225	1.7	1.5
15	4	C	453	2.2	1.3
16	4	C	246	2.8	1.5
17	5	A	955	2.8	1.9
18	5	A	410	1.5	1.1
19	5	C	608	2.7	2.0
20	5	C	599	2.3	1.7
21	6	D	779	4.4	1.9
22	6	D	674	3.8	2.0
23	6	DE	748	3.8	1.9
24	6	DE	572	3.7	1.7
25	6	DF	390	3.4	1.8
26	6	DF	304	2.4	1.4
27	6	DFG	775	4.5	1.4
28	6	DFG	290	4.4	1.7
29	6	DFH	870	4.2	1.6
30	6	DFH	239	2.8	1.6
Mean				2.78	
Median				2.65	1.60
Interquartile range				.93	.33

Table 2: Production Characteristics and Response Categorization for 30 Switch Sounds. [a]Switch types 1-5 were pull-chain, switch type 6 was a push-dimmer switch. [b]Production characteristics: A = Down pull; B = Straight out pull; C = Side pull; D = Handheld press; E = Press on edge of knob; F = Placed with back on wood; G = Placed with side on wood; H = Placed with edge on wood.

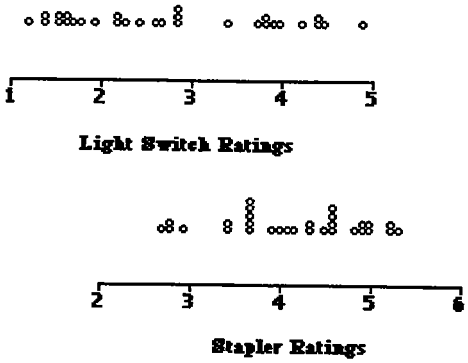

Figure 1: Distribution of average ratings for switch and stapler events.

	Switch 6	Stapler 17
Average rating	1.2	3.7
Class average	2.78	4.15
Deviation from class average	1.58(+)	.45(−)
Deviation from correct scale end	.2	2.3
Identification time[a] (ms)	6022	6055
Causal uncertainty[a]	4.40	4.65
Accuracy[a]	.18	.06
Ecological freq[a]	3	5
Familiarity[a]	2.1	2.2
Identifiability[a]	2.8	2.8

Table 3: Perception data for wwitch 6 and stapler 17. [a]Data from Ballas [3].

of causal uncertainty, and ratings of familiarity and identifiability. These results suggest that certain aspects of identification performance (i.e., accuracy and forced choice classification) relate to the properties of the particular sound, while other aspects (i.e., identification time and causal uncertainty without particular alternatives given) relate to the class of sounds that are exemplified and the number of alternative events that can produce this class of sounds. In other words, there are two levels of analysis that may be necessary in selecting and designing a sound for accurate identification. At one level, one should be concerned with the details of the sound. At another level, the concern is the general type of sound and what other events it can signify. Good design of the details can reduce the ambiguity, but in some cases ambiguity may not ever be eliminated simply because two different events can legitimately produce the same acoustic pattern.

Wave form and spectral analyses of the sounds suggested that stapler stimuli were identified on the basis of low-frequency components in the first transient that are still present but become attenuated in the second. Misidentified stapler sounds had a spectral shift from the first to the second transient. Correctly identified switch sounds were marked by harmonic components probably produced by resonance of the ceramic base. The onset of the sound had a repeated impulse pattern produced by the rolling chain. Misidentified switch sounds, particularly the push-dimmer switches, had a sharp onset in the amplitude envelope, which was more typical of stapler sounds.

3 Experiment Two

The second experiment was conducted to examine how much of a sound is needed for its identification. Even though the complete sound will usually be generated, it would be useful to know when sufficient information has been presented for identification. At the very least it would be useful to know whether the complete sound must be heard. A gated stimulus paradigm was used, one that has been employed in studies of word perception to investigate the continuous acoustic analysis that occurs in the recognition of individual words [5, 6].

3.1 Method

3.1.1 Subjects

There were 63 participants, all with reported normal hearing. They were students from an undergraduate psychology course and participated as part of a class demonstration. The students were tested in two sessions. In the first session, a classroom-quality tape player was employed. In the second session, an improved tape player with a frequency response of ± 3dB from 40 Hz to 15 kHz was used. Results from both sessions were comparable and combined in the results below.

3.1.2 Stimuli

41 sounds were presented for identification with the duration of the sound successively incremented in 50 ms steps. The sounds were those which had been used in previous studies of everyday sound identification [3]. They were sounds of everyday events taken mostly from sound effects libraries. The sounds were digitized at a 20 kHz sampling rate through a low pass filter set at 10 kHz. A tape was made by generating the sounds with a DAC set at a 20 kHz sampling rate through a low pass filter set at 10 kHz.

3.1.3 Procedure

An ABX discrimination procedure was combined with the gating procedure. Participants were told that they would hear two sounds and then one of the two would be repeated, starting at very short duration, and gradually increasing in length. They were told to choose between the two initial sounds as soon as they had heard enough of the test sound to make this decision. The alternative chosen for each test sound was one that had somewhat similar acoustics.

3.2 Results and Discussion

The average number of 50 ms increments that was needed for correct discrimination is listed in Table 4 for each sound. The listing is sorted according to the average gated duration. Since the sounds varied somewhat in duration, the proportion of the needed duration to total duration was also calculated and is shown. The results show that listeners needed to hear very little of the sound to make accurate decisions about which of the pair was repeated. Accuracy was above 70% for all but three of the 41 test sounds. In general less than half of the sound needed to be heard, and in many cases less than 20% of the sound was needed for an accurate response.

Since the distribution of these durations is skewed (see Figure 2), a longer portion is needed for only some of the sounds. Many of the signaling sounds could be discriminated with only a brief duration. Several of the modulated noise sounds were discriminated at a brief duration probably on the basis of pitch differences, pattern differences, or both (e.g., the powersaw was a "whine"

Test Sound	Alternate	Average Number of 50 ms Intervals	Duration (ms)	Sound Proportion
Drip	Lighter	1.048	0.052	0.185
Telephone ring	Automatic rifle	1.097	0.055	0.102
Power saw	Lawnmower	1.129	0.056	0.088
Church bell	Door bell	1.21	0.061	0.095
Car horn	Bugle	1.274	0.064	0.1
Doorbell	Bellbuoy	1.274	0.064	0.1
Lighter	Drip	1.306	0.065	0.508
Touch tone pulse	Car horn	1.5	0.075	0.172
Cork pop	Hammering	1.525	0.076	0.141
Sawing	Bacon frying	1.6	0.08	0.125
Sub dive horn	Fog horn	1.613	0.081	0.138
Fog horn	Sub dive horn	1.633	0.082	0.128
Lawnmower	Power saw	1.645	0.082	0.128
Tree chop	Cork pop	1.65	0.083	0.249
Electric lock opened	Jail door closed	1.656	0.083	0.13
Clock tick	Light switch	1.726	0.086	0.134
Boat whistle	Car horn	1.726	0.086	0.134
Hammering	Door knock	1.71	0.086	0.187
Bellbuoy	Doorbell	1.774	0.089	0.139
Bugle	Doorbell	1.919	0.096	0.15
Telephone hung up	Automatic rifle	1.932	0.097	0.291
Water bubbling	Oar rowing	1.984	0.099	0.155
Car ignition	Power saw	2.131	0.107	0.167
Bacon frying	Hand sawing	2.236	0.112	0.182
Stapler	Light switch	2.541	0.127	0.331
Gunshot indoors	Boat whistle	2.613	0.131	0.394
Rifle shot outdoors	Rifle shot indoors	2.642	0.132	0.397
Light switch	Stapler	2.639	0.132	0.286
Door closed[a]	Stapler	2.788	0.139	0.302
Door knock	Foot step	2.923	0.146	0.228
Oar rowing	Water bubbling	3.056	0.153	0.285
Clog footsteps	Door knock	3.153	0.158	0.247
Automatic rifle	Car ignition	3.37	0.168	0.252
Car backfire	Rifle shot outdoors	3.442	0.172	0.305
Door opened	Jail door closed	3.49	0.175	0.342
Jail door closed	File cabinet closed	3.519	0.176	0.299
Footstep	Telephone hung up	4.283	0.214	0.348
Door latched	Stapler	4.522	0.226	0.384
Toilet flush	Oar rowing	4.85	0.242	0.378
Fireworks[a]	Gunshot outdoors	5.235	0.262	0.409
File cabinet closed[a]	Door latched	5.947	0.297	0.464

Table 4: Duration needed to identify a sound as one of two alternatives. [a]Average accuracy for all subjects was less than 70% for this sound.

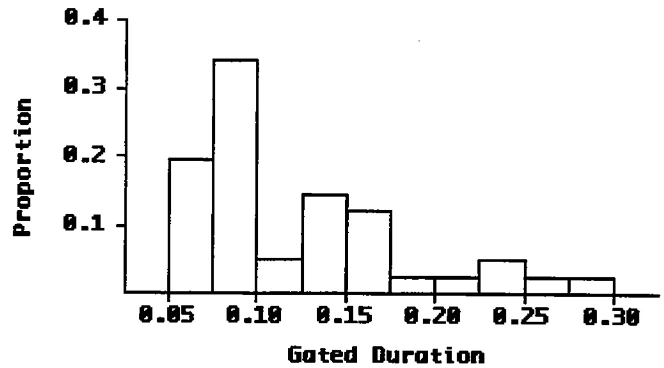

Figure 2: Frequency distribution of gated duration needed for correct discrimination response.

that steadily increased in pitch; the auto rifle was a series of low-pitched bursts). In general, the sounds that were discriminated only at a longer duration had very similar acoustics throughout (e.g., fireworks and outdoor gunshot) or included a complicated pattern of transients (file cabinet being closed).

There are several caveats to these results. First, the incremental gap duration of 50 ms was too long for many of the sound pairs since discrimination was highly accurate with the first interval presented. Thus, further studies might utilize a shorter initial duration. Second, the listeners had just heard the two possible sounds and would have detailed and accurate acoustic images for these sounds to use in deciding upon the test sound. Thus the task was much easier than the task of identifying a sound on its first presentation. Third, the two alternatives were not subjected to a rigorous acoustic similarity analysis and in some instances were quite dissimilar. Very little acoustic information would be needed to decide in these instances.

There is a dilemma that must be faced in designing a study of this type. On the one hand, if the alternatives are precisely alike acoustically, then discrimination would have been chance at any duration. On the other hand, if the alternatives are quite dissimilar, then the initial sound onset may hold sufficient information to discriminate the sounds. Thus the outcome (how much sound is needed) will be highly dependent on the sound pairs. My interest was in discovering whether the gating paradigm would provide useful information about how much of a sound might be needed and so the sound pairs varied in acoustic similarity. Future studies on this issue should address acoustic similarity in greater detail. In doing so, these studies would provide information about the type of acoustic information that enhances rapid discrimination and identification.

Conclusion

These two studies provide some further insights into the factors that are important in the accurate perception of everyday sound. The first study is unique in examining in detail how the production characteristics of two dissimilar events affect the identification of a sound. Depending on the production characteristics, the two dissimilar events were either aurally distinguishable or not. Together with other results, this study supports the notion that acoustic design of everyday sounds must be concerned with the acoustic details as well as the general type of sound being utilized. It may be that certain classes of sounds should be avoided because they inherently do not support unique interpretations.[1] The type of sound used in the first experiment could be described as a phrase with two transient impacts. A number of events (a stapler, pull chain light switch, pen click, camera shutter, coin dropping in machine) can produce this general type of sound, and without careful attention to the detailed design of the sound and analysis of the contextual conditions in which it will be used, perceptual ambiguity can result. Good design of the details can reduce the ambiguity, but in some cases ambiguity may not ever be eliminated simply because two different events can legitimately produce the same acoustic pattern.

The second experiment provided information that perceptual discriminations can be made for many sounds when only the initial segments have been heard. Although in the usual case, one would generate the complete sound, the results showed that listeners generally needed to hear very little of the sound to make accurate decisions. In general less than half of the sound needed to be heard, and in many cases less than 20% of the sound was needed for an accurate response. If so little of the sound is really needed, then one wonders whether the complete presentation of the sound will be perceived positively, unless it has some inherent aesthetic appeal.

4 Acknowledgments

This research was supported by the Naval Research Laboratory and the Office of Naval Research.

References

[1] Ballas, J. A. "Delivery of Information Through Sound." *Auditory Display: Sonification, Audification and Auditory Interfaces*, edited by G. Kramer, 79–94. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[2] Rettig, T. "Appendix I: Comments on ICAD 92." *Auditory Display: Sonification, Audification and Auditory Interfaces*, edited by G. Kramer. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[3] Ballas, J. A. "Common Factors in the Identification of an Assortment of Brief Everyday Sounds." *J. Exper. Psychol.: Human Percep. & Perfor.* **19** (1993): 250–267.

[4] Handel, S. *Listening: An Introduction to the Perception of Auditory Events.* Cambridge, MA: MIT Press, 1989.

[5] Luce, P. A. "A Computational Analysis of Uniqueness Points in Auditory Word Recognition." *Percep. & Psychophys.* **39** (1986): 155–158.

[6] Warren, P., and W. Marslen-Wilson. "Continuous Uptake of Acoustic Cues in Spoken Word Recognition." *Percep. & Psychophys.* 41 (1987): 262–275.

[1] A good example is the ubiquitous computer "beep." It is probably not useful to spend effort fine tuning the 'beep' sound because it is associated with a variety of interpretations.