ARL

Parameterizing Sound: Design Considerations for an Environmental Sound Database

by Kelly Dickerson, Jeremy R Gaston, and Savannah McCarty-Gibson

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5067

Parameterizing Sound: Design Considerations for an Environmental Sound Database

Kelly Dickerson and Jeremy R Gaston
Human Research and Engineering Directorate, ARL

Savannah McCarty-Gibson
Science and Engineering Apprentice Program, Harford Community College

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
April 2015	Final	1 June 2014–31 July 2014

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Parameterizing Sound: Design Considerations for an Environmental Sound Database	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Kelly Dickerson, Jeremy R Gaston, and Savannah McCarty-Gibson	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
US Army Research Laboratory ATTN: RDRL-HRS-D Aberdeen Proving Ground, MD 21005-5067	ARL-TR-7198

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Primary author's email: < kelly.dickerson5.civ@mail.mil>.

14. ABSTRACT

This report describes the work underway in establishing a database of environmental sounds for research at the US Army Research Laboratory. This report reviews the current status of research on environmental sound perception as well as the challenges associated with characterizing such a broad stimulus class. One of the goals of this report is to establish what properties of environmental sounds are important to document for the purposes of a database. We end this discussion by providing a path forward for implementing the criteria highlighted here. The current document provides a summary of high-impact research on environmental sound perception as well as documentation on the sound source acquisition process. A list of environmental sounds used in previous research is included with this report, and many of the sound files are available by request.

15. SUBJECT TERMS

environmental sounds, design criteria, complex source categorization, acoustics, psychophysics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Kelly Dickerson
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	UU	28	410-278-5979

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

INTENTIONALLY LEFT BLANK.

1. Introduction

This report provides the background, rationale, and documentation for a project completed during the summer of 2014 as part of a SEAP (Science and Engineering Apprenticeship Program) student project. While there has been a recent push for increased focus on environmental sound research, this thrust of research has revealed that unlike speech and music no standards for research-quality stimuli been established. Further, there are no widely accepted definitions or normative data for documenting "common environmental sounds". This technical report reviews the recent research on environmental sound perception as well as provides basic information on user search behavior and database design. Also included is a description of how the sounds included in the pilot version of the sound library were obtained Some sounds that are representative of sources present in many everyday environments were not available in the public domain and were recorded at the Environment for Auditory Research (EAR) facility. Therefore, quality and measurement standards for those recordings are included here. It is important to point out that creation of an environmental sound database is a complex, multifaceted problem. To move from conceptualization to implementation, experts in psychology, acoustics, linguistics, software engineering, and user experience must work together to make this problem tractable.

2. Environmental Sound Perception

Environmental sounds are ubiquitous, but like music and art, ubiquity does not always lend itself to well-defined structure or definition. From an ecological perspective, VanDerveer (1979) defines environmental sounds as both causal and meaningful, where meaning is derived from their causality. That is, environmental sounds are not meaningful as a collection of individual descriptive acoustic features but are defined by the listener in the context of the event that produced the sound they are hearing. This listener-centric definition is intuitive and descriptive; however, it does little to facilitate an understanding of what aspects of environmental sounds are important for perceptual decisions such as recognition, identification, and discrimination.

Work in psychophysical acoustics has revealed basic low-level properties that influence perceptual decisions. For example, changes in frequency are perceived as changes in pitch, changes in sound level are perceived as changes in loudness, and changes in harmonic structure are perceived as changes in timbre (Moore 2012). All of these examples are fundamental aspects in the perception of any sound. However, complex sounds are more than combinations of single discrete properties. For more than a century, much of the auditory perceptual work has focused on simple and easily defined laboratory-generated stimuli. This has helped to establish a very

good understanding of basic human hearing, but the generalization of these types of simple relationships to the complexity that is relevant to real-world listening is unclear. Indeed, there are many examples of how perception qualitatively changes with increased complexity. In the discrimination of tone sequences, simply adding tones to the sequence and/or varying the serial position of each tone in the sequence can dramatically affect discrimination performance (Watson et al. 1975). Even changing the spectral makeup of nonsimultaneous contextual sounds can cause significant shifts in perceptual functions. This is true of both leading contextual sounds (Ronken 1972; Holt 2006), and trailing contextual sounds (Massaro 1975; Pastore, Gaston, et al. 2008; Pastore and Gaston 2012). Further, when listeners are asked to categorize frequency glides that vary along two dimensions, different listeners adopt qualitatively different categorization strategies (Holt and Lotto 2006). All of these examples represent only very modest increases in stimulus complexity; real-world sounds are generally much more complex.

Human speech perception probably represents the most extensively studied class of complex sounds, and much of speech research has been psychophysical in nature (Pisoni and Remez, 2005). As a result of this extensive research, the state of understanding for speech perception is fairly mature, especially for the perception of segmental speech (i.e., phonemes). Segmental speech stimuli are complex and vary along multiple dimensions, but importantly, no single dimension is invariant. Rather, listeners capitalize on patterned variability such that variability serves as a cue to accurate categorization decisions most of the time (Raphael 2005; Cleary and Pisoni 2008). Generally, accurate categorization requires use of multiple cues, and like tonal stimuli, different listeners tend to use different cues for categorization (Raphael 2005).

Compared with segmental speech perception, the body of environmental sound perception research is quite modest, and has not had the benefit of such an extensive history. Even so, there are indications that similar general principles hold true across these classes of sounds. For example, Pastore, Flint, et al. (2008) modeled individual listener performance for judgments about the sounds of human walkers and found that no single acoustic property could predict the high levels of performance exhibited by some listeners. Rather, the high performance observed required the use of multiple acoustic cues. Moreover, modeling of individual listeners' use of information demonstrated that use of specific acoustic cues varied widely across individuals, and individuals would often use suboptimal cues.

Even though similar general principles may apply across speech and environmental sound perception, the information relevant to each sound class can be very different. Although very complex, the range of speech sounds is constrained by the articulation of the vocal tract (Denes and Pinson 1993); thus, despite the availability of numerous cues, the available information is still somewhat constrained. In contrast, environmental sounds as they are defined here include naturally occurring nonspeech, nonmusic sounds, and thus represent much wider variation in the types of acoustic cues that may be important for perception, while a supporting literature investigating the possible cues to environmental sound perception remains sparse.

3. Environmental Sound Perception Research Methods

To study perceptual decisions under realistic conditions, researchers have accepted that many sources of variability are uncontrollable and unquantifiable. Thus, much of the research in environmental sound perception has high ecological validity, but little ability to generalize given the lack of experimental control over experimental variables. One way to improve experimental control while continuing to use ecologically valid stimuli and methods involves applying psychophysical approaches to environmental sounds. This approach involves determining what physical properties map onto perception, and ultimately how that mapping influences performance on behavioral tasks. For example, Gygi et al. (2007) conducted a broad sound classification study based on listener-generated similarity judgments and examined the correlation of acoustic cues to those judgments. In this study, listeners were presented with all possible pairs of 100 environmental sounds, representing 10,000 observations for each participant. For each pair of sounds, listeners rated how similar they thought the samples were, on a scale of 1 to 7. These similarity ratings were then subject to a multidimensional scaling (MDS) analysis. MDS is a multivariate statistical technique that provides a Euclidean distance mapping of an input distribution (in this case based on stimulus dissimilarities) that can be used to estimate perceptual space (Young 1987); this mapping can be constructed using single or multiple dimensions, though in practice, solutions using more than 3 dimensions can be very difficult to interpret. Similarity, then, is based on Euclidean distance between stimulus items in the solution.

Pair-wise ratings are not the only way to generate similarity matrices for analysis using MDS. Aldrich et al. (2009) found that listeners generated very similar MDS maps for similarity matrices generated via sorting tasks as matrices generated by sequential pair-wise comparisons. The characterization of subjective similarity among items is only one of several important aspects to consider in determining the relationship between complex signals such as environmental sounds and how they are perceived. Gygi and colleagues measured the acoustic properties of each of the 100 sounds included in their study to support the interpretation of the relevant dimensions in the MDS solution. They found that the similarity among items in their sound set could be explained by a 3-dimensional MDS solution where dimension one mapped onto pitch salience and modulation spectrum and dimension two was primarily spectral with spectral centroid and deviation accounting for the greatest percentage of variability. Dimension three captured spectral-temporal complexity, specifically total duration and envelope shape (e.g., bursts, autocorrelation peaks, and standard deviation). The approach of combining objective acoustic measures with subjective similarly ratings can also be applied such that it is predictive of performance on identification, discrimination, or other behavioral tasks. Gaston and Letowski (2012) demonstrated that estimates of listener perceptual space using MDS were predictive of listener recognition of weapon type from the sounds of small-arms fire.

While these examples provide important insights into the relationships between the physical world and perception, this information is incredibly labor intensive to generate. In addition to the general task of designing a targeted, balanced, and well-controlled behavioral study, each environmental sound study requires careful sound source selection, stimulus norming, and stimulus editing (sometime professional-level sound editing expertise is needed). At one level, there is a clear need to reduce the burden of generating environmental sounds for research. At another level, a common well-documented database of environmental sounds for research would help provide consistency and generalizability across research studies.

4. Scope of the Current Project

Environmental sound researchers have accepted that everyday sounds are complex and there is significant uncontrolled variability. One way to overcome this issue is to document the variability present within a complex signal. Across the last few decades, prominent environmental sound researchers have suggested that a database of environmental sounds would be a powerful research tool, particularly in addressing the concern of complex and highly variable signals. Despite this overwhelming awareness, little work has been done to develop such a tool. Some researchers have established websites to share their sound resources; however, with these stimuli little if any documentation on their origin, construction, or quality is provided. Gygi and Shafiro (2010) proposed a set of guidelines to be implemented in the DESRA (database for environmental sound research and application). Yet, nearly 5 years later Gygi and Shafiro have not made any of their resources publicly available or published a searchable database of environmental sounds.

This is not to say that such a broad and complex task should be achieved in 5 years' time; the purpose of this statement is to simply demonstrate that this task has not yet been undertaken or achieved. Gygi and Shafiro (2010) suggest that at a minimum, a database of environmental sounds should include information on the sound familiarity or prevalence in "everyday" listening environments. The exact form this metric takes is a matter of debate. Ballas (1993) and others have developed measures similar to word frequency for environmental sounds; however, other measures such as identifiability or familiarity (Bonebright 2001) may also be appropriate. Other qualitative aspects of environmental sounds may also be appropriate metrics for sound classification, particularly when building a database. As user query behavior is likely driven by some of the same qualitative information as classification tasks in the auditory domain, other qualitative aspects of environmental sounds may also be appropriate metrics for sound classification, particularly when building a database.

In addition to the qualitative/subjective features, physical/objective measures also need to be documented. These measures should at a minimum include waveform statistics such as sample and bit rates, peak and RMS amplitude, duration, number of channels, as well as information

about the microphone used, and its relationship to the sound source (i.e., distance and orientation). Further, contextual information, such as where and under what environmental conditions the recording occurred would be very useful. Documentation of these details would provide consistency of resources across research groups, and save significant time, in addition to being consistent with best practices and American National Standards Institute (ANSI) standards (ANSI S1.13 2005). Contextual information provides more than acoustic details; it may also facilitate in the future functionality of the database in terms of search and usability. Users do not search without purpose; search is initiated by information need (Ruthven et al. 2003). In the case of a sound database for research, information need would be reflected in a user's desire to find a large set of sounds belonging to a particular subclass or a user's need to find a sound that fits into a particular background or context. Thus, the organization and classification of such a broad class of environmental sounds is critical. Environmental sounds can be classified at numerous levels, from contextual hierarchies to feature-level descriptions. Gaver (1993) points out that there is no consistent classification scheme in place in commercially available sound libraries; the typical sound effects library contains a range of descriptive levels. To maximize successful classification but also optimize search results in terms of relevance and appropriateness, sounds within a database should be consistently characterized in terms of contextual, hierarchical, and dimensionally based (featural) information.

5. Sound Library Documentation

The purpose of this section is to document the procedures for collecting and measuring the sounds included in the pilot version of the database.

5.1 Sample Selection and Collection

A literature review was conducted to make some initial determination of the extent to which work related to environmental sound perception was being conducted. Seven different research groups and 10 recent publications, including a book with a companion CD, were selected for evaluation. These selections were based on the availability of sound samples and access to normative data (see Appendix A for a list of the studies included in this report). While this is clearly not an exhaustive list, it does represent a wide range of methods and selection techniques, as well as a large sample of environmental sounds. Most of these sounds were available for download, and interested parties can obtain them by request to the author of this report (see Appendix B for the full list of sounds).

5.2 Inclusion Criteria

Each of the studies included in our survey list specify sample selection criteria that reflect the needs of that research (see Appendix A for brief descriptions of criteria). Additionally, across these studies, there are several variations of a definition for environmental sounds, making

establishing broadly applicable selection criteria for a database somewhat difficult. As an entry point, the definition of environmental sounds used for this project is fairly conservative and consistent with the definitions given by VanDerveer (1979), Ballas (1993), and Gaver (1993). We define environmental sounds as 1) sounds specifically associated with, or produced by, a physical event or human activity and 2) sound sources that are common in the environment. Reproductions or sound effects (e.g., Foley sounds) are not included. The current sound library includes several human and animal vocalizations (e.g., male and female speech, dogs barking, bird calls), while these sounds are representative of many environments it is an open debate in the literature whether they are "environmental sounds" or if these vocalizations should be considered separately. Additionally, there are several samples of sounds generated by various musical instruments, while not uncommon in many environments, this distinct subclass of sounds has received substantial consideration apart from the environmental sound literature, and it is debatable whether these types of sounds should be included in a catalog of environmental sounds.

Not all of the sounds discovered during the literature review process were available for inclusion in the current sound library. Some researchers did not make their samples public, while others did not respond to requests to provide their samples. Only a few researcher groups provided their samples for this project. Samples from Melissa Gregg and Brian Gygi were provided through personal correspondence, and Marcell and colleagues published their catalog of sound samples on the web (http://marcellm.people.cofc.edu/confrontation%20sound%20naming/zipped.htm) as did Hocking et al. (2013) (http://www.imaging.org.au/Nessti). The sound library associated with this report includes all of the samples that were accessible at the time of this report. Of the 482 samples listed across the 10 studies, 310 samples were collected for inclusion in the sound library. Forty-seven additional sounds were recorded in the EAR facility and are also included in the library. Thus, the total number of samples in our sound library stands at 357. The quality standards for the recorded samples are documented in the next section.

5.3 Documenting the Samples

Consistent with the recommendations of Gygi and Shafiro (2010), documented technical details include sample and bit rate information, number of channels in the original recording, file type, and any available microphone and contextual information (where and under what conditions the sample was recorded) that is important for characterizing complex sources, such as environmental sounds. A catalog for the sound library that lists the available technical and acoustic details available for each sample is currently in development and is available by request. This information will be useful to researchers interested in environmental sounds as it provides baseline data about the quality and origin of the sound samples included in this library. Many of the sound samples described in this report are available by request, and a full listing of the sounds evaluated is provided in Appendix B.

5.4 Recording Samples

All recordings were conducted in a hemi-anechoic chamber. All samples were captured at 44.1 kHz, 16 bit, using Adobe Audition 3.0, running on a laptop PC. The samples were captured using a G.R.A.S. sound and vibration 1/2-inch free-field microphone. Calibration of the microphone was accomplished using a G.R.A.S. 42-AA calibrator set (250 Hz at 114-dB sound pressure level).

5.5 Quality Standard for Recorded Samples

For all recordings, the primary concern was to eliminate, or at least minimize, any background noise or other sounds irrelevant to the sound source of interest. Thus, all of the sounds recorded at the US Army Research Laboratory were done in a hemi-anechoic chamber. Often it is not possible to record under these conditions. Rather, it is often necessary to make recordings in uncontrolled natural environments. In these cases, care should be taken to limit irrelevant noise and reflections as much as possible. If outdoors, large open grassy fields are essentially anechoic and can make ideal locations to record sound sources, especially during quiet times of the day. Recordings outdoors should never be performed when the wind is in excess of 15 mph, and a windscreen should be used to minimize any potential wind noise (ANSI S1.13 2005). Use of cardioid or hypercardioid microphones can also be used to help control background noise because they have a directional response pattern that limits sounds outside the cardioid response fields. These can be useful in reducing background noise as well as sound reflections that fall outside of the response field of the microphones.

6. Discussion

One aspect of environmental sound perception that is noticeably absent from this discussion and from the current instantiation of the database is quantifying top-down influences such as semantic- or expertise-related effects on perception. This lack of focus on semantic or categorization issues was intentional and related to how the forthcoming database and its search algorithms will be structured. We allude to the important contribution of top-down factors such as semantic information, user goals, and source familiarity in our brief discussion of factors influencing the initiation and satisfaction of search. There are 2 important aspects in describing the top-down influences; what are the influences for the end user and what are the influences for the listeners presented with sources obtained from the database. There is some recent evidence that the user-generated (researcher-generated) categories are highly similar to the categories generated by listeners in environmental sound perception studies (Aldrich et al. 2009).

Each source has a file name that provides a semantic- or category-level label. When these samples are transitioned from their list into the database, each file will have several semantic tags, where high frequency words (e.g., keywords) synonymous with the file name will be listed. This will enable researchers who use the database to search for samples based on a fairly broad conceptualization of what that particular sample represents.

7.   Conclusions and Recommendations

We expanded on the work of Gygi and Shafiro (2010) by exploring the notion of a database and development of criteria and quality standards for such a resource. Developing a sound database for research is a complex multifaceted task that involves measurement and classification of sounds used in previous research as well as the capture of samples that could not be found or that have not been included in studies up to this point but are clearly representative of the environment we are attempting to represent with this catalog.

Current project efforts have led to the development of a set of sound files associated with the Excel catalog and both are available by request (see Appendix B for the full list of sounds). However, this media is neither dynamic nor in a format that is easy to distribute. The next step in this project is to develop a searchable database where sounds are defined by multiple attributes at multiple levels. Finally, we would like to make this database and its established quality standards publicly available such that the research community could download any of the samples, or upload samples that meet the quality standards but are absent from the database.

8. References

Aldrich KM, Hellier EJ, Edworthy J. What determines auditory similarity? The effect of stimulus group and methodology. The Quarterly Journal of Experimental Psychology. 2009;62(1):63–83.

ANSI S1.13-2005 (R2010). Measurement of sound pressure levels in air. Melville (NY): Acoustical Society of America; 2005.

Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: Human Perception and Performance. 1993;19(2):250.

Bonebright TL. Perceptual structure of everyday sounds: a multidimensional scaling approach. In: Hiipakka J, Zacharov N, Takala T, editors. ICAD 2001. Proceedings of the 7th International Conference on Auditory Display; 2001 Jul 29–Aug 1; Espoo, Finland; Atlanta (GA): Georgia Institute of Technology International Community on Auditory Display; c2001.

Cleary M, Pisoni DB. Speech perception and spoken word recognition: research and theory. In: Goldstein EB, editor. The Blackwell handbook of sensation and perception. Malden (MA): John Wiley & Sons; 2008. p. 499–524.

Denes PB, Pinson EN. The speech chain, 2nd edition. New York (NY): Macmillan; 1993.

Gaston JR, Letowski TR. Listener perception of single-shot small arms fire. Noise Control Engineering Journal. 2012;60(3):236–245.

Gaver WW. How do we hear in the world? Explorations in ecological acoustics. Ecological Psychology. 1993;5(4):285–313.

Gygi B, Kidd GR, Watson CS. Similarity and categorization of environmental sounds. Perception & Psychophysics. 2007;69(6):839–855.

Gygi B, Shafiro V. From signal to substance and back: insights from environmental sound research to auditory display design. In: Auditory display. Heidelberg (Germany): Springer Berlin; 2010. p. 306–329.

Hocking J, Dzafic I, Kazovsky M, Copland DA. NESSTI: norms for environmental sound stimuli. PloS one. 2013;8(9):e73382.

Holt LL. Speech categorization in context: joint effects of nonspeech and speech precursors. The Journal of the Acoustical Society of America. 2006;119(6):4016–4026.

Holt LL, Lotto AJ. Auditory categorization: cue weighting and dimension bias. The Journal of the Acoustical Society of America. 2006;116(4):2624–2624.

Massaro DW. Backward recognition masking. The Journal of the Acoustical Society of America. 1975;58(5):1059–1065.

Moore BC. An introduction to the psychology of hearing. 6th ed. Bingley (UK): Emerald Publishing Group; 2012.

Pastore RE, Flint JD, Gaston JR, Solomon MJ. Auditory event perception: the source—perception loop for posture in human gait. Perception & Psychophysics. 2008;70(1):13–29.

Pastore RE, Gaston JR, Berens MS. Backward recognition masking as a general type of interference in needed poststimulus processing. Perception & Psychophysics. 2008;70(6):1104–1116.

Pastore RE, Gaston JR. A cognitive concept with relevance to speech perception? In: Zonneveld W, Quene H, Heeren W, editors. Sound and sounds: studies presented to M. E. H. Schouten on the occasion of his 65th birthday. Utrecht (Netherlands): Institute of Linguistics OTS; 2012.

Pisoni DB, Remez RE, editors. The handbook of speech perception. Malden (MA): Blackwell Publishing Ltd; 2005.

Raphael LJ. Acoustic cues to the perception of segmental phonemes. The handbook of speech perception. 2005. p. 182–206.

Ronken DA. Changes in frequency discrimination caused by leading and trailing tones. The Journal of the Acoustical Society of America. 1972;51(6B):1947–1850.

Ruthven I, Lalmas M, Van Rijsbergen K. Incorporating user search behavior into relevance feedback. Journal of the American Society for Information Science and Technology. 2003;54(6),529–549.

VanDerveer NJ. Confusion errors in identification of environmental sounds. The Journal of the Acoustical Society of America. 1979;65(S1):S60–S60.

Watson CS, Wroton HW, Kelly WJ, Benbassat CA. Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals. The Journal of the Acoustical Society of America. 1975;57(5):1175–1185.

Young FW. Multidimensional scaling: history, theory, and applications. Hamer RM, editor. New York: Lawrence Erlbaum Associates; 1987.

Appendix A. Inclusion Criteria by Study for 10 Studies Included in Report

The 10 studies mentioned in this report represent a sample of high-impact publications on environmental sound perception. The primary research questions for each of these studies are different. Thus, the source inclusion and evaluation criteria are different for each study. Further, these studies span 2 decades of research, and as such, the source measurement techniques span a wide range of standards and technologies. The purpose of this appendix is to summarize the inclusion criteria for the sound samples present in each study to provide the reader with some metric of the quality standards applied to sounds included in the report's sound library. Further, while a broad and extensive literature review is beyond the scope of this paper, some background on the methods of environmental sound research may be helpful to some readers.

A.1 Ballas (1993)[1]

Forty-one sounds were selected as subjectively good representations of the events causing the sound and to be either easy or hard to identify. The sounds included signals, sounds characterized by some sort of modulated noise, and sounds involving multiple mechanical transients and sounds of discrete impacts. Sounds were tested for discriminability in ABX (same-different) task and were 99.8% discriminable. Several previous studies reported a link between average spectral properties and perceptual performance. Ballas computed average acoustic properties including duration, average magnitude, peak magnitude, power, fast Fourier transform (FFT) spectrum and 1/3 octave bands. Moments of the FFT spectrum were also computed. Further acoustic analyses were conducted based on examination of sound spectrographs. Ballas examined the spectrographs to find spectral-temporal properties that would be related to identification performance. Prospective properties included harmonics, continuous spectral bands, spectral similarity of bursts, spectral width of bursts, and spectral shifts within bursts. Temporal properties were also calculated, specifically, the number of bursts in the sound file, the durations of the bursts, and the ratio of burst duration to total duration. The bursts were defined by envelope modulation rather than gap duration because several sounds had perceptually distinct bursts even though the amplitude envelope did not include distinct gaps of silence.

A.2 Bonebright (2001)[2]

Seventy-four sounds made by common objects were selected for inclusion in this study. Forty-one of these were used previously in Ballas[1]. The remaining 33 sounds were selected from a list generated by 5 independent raters as representative environmental sounds. The vast majority of the samples were recorded and digitized by Bonebright's laboratory staff. However, some sounds were selected from compact disk sound effects libraries. The specific sound effects library is not listed in this report.

---

[1] Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: Human Perception and Performance. 1993;19(2):250.

[2] Bonebright TL. Perceptual structure of everyday sounds: a multidimensional scaling approach. In: Hiipakka J, Zacharov N, Takala T, editors. ICAD 2001. Proceedings of the 7th International Conference on Auditory Display; 2001 Jul 29–Aug 1; Espoo, Finland; Atlanta (GA): Georgia Institute of Technology International Community on Auditory Display; c2001.

Bonebright reports that the sounds included in the 2001 study were sampled at 8 bits, 22.3 kHz, and the mean sample duration was 1,539 ms (range = 75.5–4714 ms). Bonebright's group conducted several acoustic analyses including average intensity (measured by energy flux density divided by the duration of the sound), changes in frequency (measured as differences between the upper and lower frequency in the sound), and changes in time (duration of the sound in seconds). There were several measures of intensity, specifically, changes in intensity, measured as "intensity in Hz from one end of the sound to the other", amplitude ceiling (highest amplitude level in the sound), dynamic range (the difference in amplitude floor and ceiling), peak intensity, measured as "maximum intensity/Hz" and peak frequency (the frequency at which the highest amplitude occurs).

A.3 Gregg and Snyder (2012)[3]

Fifteen common environmental sounds, including human and nonhuman animal vocalizations, were included in this study. All of these sounds were matched for mean amplitude, filtered for noise, and on/off ramps were added to avoid abrupt transients. Sample rate and duration information is also available for this set. No additional normative data are included.

A.4 Gregg and Samuel (2008)[4]

Eighteen common environmental sounds gathered from various online sources were included in this study. The specific web site sources were not mentioned. Speech from a female and a male speaker was recorded as they produced a single sentence and a sentence where all the syllables were replaced with "ma". The sentence that the speech samples were extracted from is not listed in this report. Speech samples were recorded in a sound-attenuated chamber.

All sound samples were digitized to 44.1 kHz and filtered using a noise reduction procedure custom generated for the specific spectral envelope of each stimulus. All samples were truncated to 1,000 ms and included a 10-ms linear on/off ramp to avoid abrupt onset and offset. The stimuli were matched for RMS (root mean square) amplitude to roughly equate for loudness differences.

A.5 Gregg and Samuel (2009)[5]

Eighty-eight samples were initially considered for inclusion in this study—specifically, 4 exemplars (tokens) for each of 22 common categories of environmental sound. For example, "dog" was included as a category and there were 4 acoustically distinct tokens for the dog

---

[3] Gregg MK, Snyder JS. Enhanced sensory processing accompanies successful detection of change for real-world sounds. Neuroimage. 2012;62(1):113–119.

[4] Gregg MK, Samuel AG. Change deafness and the organizational properties of sounds. Journal of Experimental Psychology: Human Perception and Performance. 2008;34(4):974.

[5] Gregg MK, Samuel AG. The importance of semantics in auditory representations. Attention, Perception, & Psychophysics. 2009;71(3):607–619.

category. A subjective similarity rating study was used to select the tokens for each category that were maximally dissimilar. This procedure yielded a down-selected set of 24 sound sources—specifically, 12 highly dissimilar token pairs.

All stimuli were digitized to 44.1 kHz and filtered using a noise reduction procedure custom generated for the specific spectral envelope of each stimulus. All samples were truncated to 1000 ms and included a 10-ms linear on/off ramp to avoid abrupt onset and offset. The stimuli were matched for RMS amplitude to roughly equate for loudness differences. Acoustic distinctiveness was measured via analysis using Praat developed by Boersma and Weenink[6]. Measures of harmonicity (mean about of acoustic periodicy in the signal) and pitch (f0) were also included. The f0 measurement was the spectral mean and for sounds that were more aperiodic, f0 was computed by averaging the f0 measurement through the duration of the signal.

A.6 Gygi, Kidd, and Watson (2007)[7]

Fifty sounds, down-selected from 70 used in an earlier Gygi et al. study,[8] were included in this study. This sample of environmental sounds is described by the study authors as "nearly perfectly identifiable". The authors go on to say that efforts were made to create a representative sampling of different types of meaningful sounds encountered during everyday listening and that this effort was based in part on the work from Gaver[9]. The particular sound categories included in this sample include nonverbal human sounds, animal vocalizations, and machine sounds. These sounds were of various weather conditions and some sounds were generated by human activities. Two tokens for each source event were selected. To reflect the range of sounds associated with a given source even or event class, tokens were selected to be maximally acoustically distinct. Sound sources were obtained from a "high-quality commercial sound effects recordings (Hollywood Edge and Sound FX The General)". Sounds were sampled at 44.1 kHz and roughly equated for loudness using RMS amplitude normalization. The mean duration of sound samples was 2,300ms (range = 579–3,945 ms).

The purpose of this study was to map out similarity space for a set of listeners, thus, subjective similarity ratings were paired with a suite of acoustic analyses. Gygi et al.[7,8] included several envelope measures (long-term RMS/pause-corrected RMS, number of peaks, number of bursts, burst duration, total duration, and roughness; measured as burst duration over total duration). As well as autocorrelation statistics (number of peaks, maximum peak, mean peak and the standard deviation [SD] of the peaks); these autocorrelation statistics capture periodicities in the

---

[6] Boersma P, Weenink D. Praat: doing phonetics by computer (Ver. 5.1. 05). Amsterdam (NL): University of Amsterdam; [accessed 2009 May 1]. http://www.praat.org.

[7] Gygi B, Kidd GR, Watson CS. Similarity and categorization of environmental sounds. Perception & Psychophysics. 2007;69(6):839–855.

[8] Gygi B, Kidd GR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. The Journal of the Acoustical Society of America. 2004;115(3):1252–1265.

[9] Gaver WW. How do we hear in the world? Explorations in ecological acoustics. Ecological Psychology. 1993;5(4):285–313.

waveform, correlogram-based pitch measures (based on Slaney,[10] including mean pitch, median pitch, SD pitch, maximum pitch, mean pitch salience, and maximum pitch salience). Moments of the spectrum were also computed (mean-centroid, skew, and kurtosis). Additionally, RMS energy in octave wide bands from 62 to 16,000 Hz, spectral shifts in time, cross-channel correlations, spectral flux, and modulation spectrum statistics were measured. Gygi's[7,8] stimuli are by far the most extensively documented samples, which is important considering that their sound set was drawn from a commercially and widely available library.

A.7 Hocking, Dzafic, Kazovsky, and Copland (2013)[11]

The purpose of this study was to provide normative data for a large set of environmental sounds. Hocking et al. included the subjective/behavioral measures of response latency, identification accuracy, categorization, familiarity, confidence, token representativeness, various affective ratings, and imageability or concreteness. This study included 110 sounds downloaded from <www.sounddogs.com> and <www.freesound.org>. The sample included equal numbers of living and manmade sources from 9 conceptual categories. All sounds were normalized to 1,000 ms and 16-bit 44.1-kHz sample rate. Sounds were normalized using Audacity, which is open source sound editing and analysis software comparable to Adobe Audition. These sound files are available for download from <http://www.imaging.org.au/Nessti>. Also included at this web address are measures of concept-based frequency measures, specifically Hyperspace Analogue to Language frequency (HAL) norms as well as counts of the number of phonemes, syllables, and the harmonics-to-noise ratio measures. These sound samples are also included in the sound library embedded in this report. Hocking et al. provides the most comprehensive subjective norming of all the studies included in this report; however, their samples, while available, do not include sufficient acoustic analyses.

A.8 Houix, Lemaitre, Misdariis, Susini, and Urdapilleta (2012)[12]

This study provided a detailed lexical analysis of 60 environmental sounds. Based on free identification, sorting, and subjective rating tasks, Houix et al. produced a taxonomy of environmental sounds similar to the original taxonomy proposed by Gaver[9]. The Houix taxonomy had 4 primary categories: liquid, solid, gasses, and machines. Further, listener ratings revealed that the temporal patterning—specifically, the impulsivity or continuousness of a sound—influenced its position within the taxonomy. All sounds were presented at an ecologically adjusted level, such that they were heard at a level that was expected and familiar. Further, all sounds were sampled at a 16-bit resolution at 44.1 kHz.

---

[10] Slaney M. Auditory toolbox: A MATLAB toolbox for auditory modeling work. Cupertino (CA): Apple Computers; 1995. Apple Tech Report No. 45.

[11] Hocking J, Dzafic I, Kazovsky M, Copland DA. NESSTI: norms for environmental sound stimuli. PloS one. 2013;8(9):e73382.

[12] Houix O, Lemaitre G, Misdariis N, Susini P, Urdapilleta I. A lexical analysis of environmental sound categories. 2012;18(1):52–80.

A.9 Marcell, Borella, Greene, Kerr, and Rogers (2000)[13]

Marcell et al. reviewed the clinical and experimental literature and developed a list of 80 previously used sounds. They sourced an additional 40 sounds from sound effect libraries and recordings of their own daily activities. They listed their guidelines to include clarity, realism, and potential identifiability when presented in isolation, that is, without a "supportive context". The sound set is described to represent a wide variety of acoustic events, such as sounds produced by animals, people, musical instruments, tools, transportation, signals, and liquids. Marcell et al. point out that their inclusion criteria are consistent with VanDerveer's definition of environmental sounds, which are defined as "non-speech sounds representing a potentially audible acoustic event which is caused by motions in the human environment".[14] Many of the sounds were edited from the original samples including truncating length, or increasing length, reducing, increasing, or normalizing volume, removing extraneous noise, and applying on/off ramps. Marcell et al. did not include specific details on the editing techniques applied to their samples; however, their entire list of 110 sounds is available on the web at <http://marcellm.people.cofc.edu/confrontation%20sound%20naming/zipped.htm>. These samples are also included in the sound library embedded in this report. Subjective norming is a strength of this sample, but, as with Hocking et al.,[11] no meaningful acoustic analysis has been conducted to provide a mapping from subjective to objective measures.

A.10 Truax (2001)[15]

The disc included with the 2001 publication Acoustic Communication 2nd edition contains 158 sounds that are described at various points within the text. These sounds are meant to serve as examples for different types or classes of environmental sounds. Unfortunately, there is limited information on the origin or the acoustics of many of these samples. The 158 samples associated with the Truax text are available by request.

---

[13] Marcell MM, Borella D, Greene M, Kerr E, Rogers S. Confrontation naming of environmental sounds. Journal of Clinical and Experimental Neuropsychology. 2000;22(6):830–864.

[14] VanDerveer NJ. Confusion errors in identification of environmental sounds. The Journal of the Acoustical Society of America. 1979;65(S1):S60–S60. p. 16.

[15] Truax B. Acoustic communication. 2nd ed. Westport (CT): Greenwood Publishing Group; 2001. Vol. 1.

Appendix B. Sound Catalog

Accordion		Car backfire		Crushing a metal can	
Aerosol can		Car crash		Crushing a tin can	
Alarm clock		Car ignition		Crushing egg shells	
Alloette flyby		Cards shuffled		Cutting a slice of bread	
Automatic rifle		Cash register		Cutting paper with scissors	
Baby crying		Cat		Diesel motor	
Bacon frying		Chain		Dishes	
Bagpipes		Chalkboard erased		Dishwasher running	
Ball turning in casino wheel		Chalkboard written on		Doctor scanner	
Balloon		Chant		Dog	
Banjo		Chewing		Dog barking	
Basketball		Chicken		Door bell	
Bell		Child coughing		Door closed	
Bells chiming		Chimes		Door knock	
Bike		Chopping wood		Door latched	
Bike pump		Church bell		Door opened	
Birds		Cigarette lighter		Door squeal	
Blender		Circular saw		Door, a cupboard closing	
Blinds		Clap		Door, lock turning	
Blowing nose		Clearing throat		Drawer opening on a track	
Blowing up a paper bag		Clicking with a mouse		Drill on concrete	
Boat horn		Clock ticking		Drinking glass plink	
Boat whistle		Clog footsteps		Dropping metallic lid on ground	
Boiling pot		Closing an old door.		Drumming	
Bongos		Closing the door of a microwave		Duct tape	
Book		Coat hangers dropped		Dump truck pass-by	
Bottle top		Coffee perking		Eggs beaten in a bowl with a whisk	
Bowling		Coffee pot whistling		Elastic (snap)	
Bread cutting		Coin dropping		Electric drill	
Brushing teeth		Coin in glass		Electric lock	
Bubbles		Coins falling		Electric saw cutting	
Bugle		Coins shook		Falling stone	
Burp		Comb		Fan	
Bus		Combination lock		Female speaking	
Bus air break		Cooking with fat		Ferry	
Bus stop and go		Cuckoo clock		Ferry horn	
Camera		Corduroy		Fireworks	
Can crush		Cork popping		Fog horn	
Can opener		Cotton tearing		Fog horns	
Can opening		Cow		Folding a wood chair	
Cans in a bag		Crickets		Food processor	
Car accelerating		Crumpling paper		Footsteps	

Fridge		Metal pan, scraping		Ring binder (3-ring binder)	
Garbage closing		Metal tape measure		River	
Gas stove		Metal trash can		Rocking in a rocking chair	
Gas stove grill on, gas and release		Microwave on (running)		Rooster	
Gas stove turning on		Microwaves beeps		Rubber band	
Glass bowl and spoon being place on a table		Monkey		Rubbing finger on a balloon	
Glass breaking		Mosquito		Running	
Glass ding, crystal champagne flute toast		Motorcycle		Salt	
Glass that is moved		Mouse trap		Salt grinder, single grind	
Grating carrots with hand grater		Music box		Sand paper	
Gun shot in doors		Nail file		Sawing	
Gun shot out door		Nails dropping		Saxophone	
Hair brush		Oar rowing		Scales	
Hair dryer		Ocean (waves)		School bus	
Hammering		Opening a beer can		Scotch tape	
Helicopter		Opening a new plastic bag		Scratching interior of iron pot	
Hitting cymbals		Opening a plastic bag		Scream	
Hollow object falling		Opening a screen door		Single prop fly by	
Hollow object rolling		Opening a Zippo lighter		Single squeeze near empty	
Honking		Opening the latch on a suitcase		Sink draining	
Horse neighing		Organ		Sink flowing and stopping	
Horse running		Owl		Siren blaring	
Ice clicks in glass without Liquid		Paint brush		Scissors	
Ice dropping into glass		Plastic container, unscrewing cap		Sleigh bells	
Jacket snap		Plates		Small breaker switch	
Jackhammer		Police siren		Small pulley of metal turning	
Jail door closed		Pouring beer in glass		Sneeze	
Jail house door close		Pulling and tearing a paper towel from the roll		Snoring	
Jar lid		Pulling the top off a bunch of carrots		Soda can	
Key lock		Purse snap		Sonar	
Keys		Putting an empty bucket on the floor		Spiral notebook being torn	
Laughing		Rain		Splash	
Lawn mower		Ratchet		Spraying polish on table	
Light switch		Rattlesnake		Stapler	
Lighting a match		Record scratching		Stirring an aerosol paint	
Lion		Removing lid of plastic container		Stirring coffee in mug	
Machine gun		Replacing a screw lid on bottle		Strumming harp	
Male speaking		Replacing the lid of an aerosol can		Sub dive horn	
Marker		Rice Krispies poured		Switching a lamp	

Match strike	Rice Krispies with water	Swords	
Taking a bowl from stack	Typing on a typewriter	Wax paper	
Tank drive by	Vacuum	Whip	
Tea kettle	Van pass-by	Whistle blowing	
Tearing cloth	Velcro	Whistling	
Tearing paper	Venetian blinds lowering down	Wind	
Telephone hung up	Video case	Wind chimes	
Thermos bottle	Violin	Windshield wipers	
Thunder rolling	Wade through water	Wolf	
Timer	Walking on gravel	Wood file	
Toaster	Walking with rubber soles	Woodpecker	
Toaster release	Water bubbling	Writing with pencil	
Trumpet	Water cooler bubbles	Yawning	
Tupperware	Water draining	Zipper	
Turning pages	Water drip		
Typing on a keyboard	Waves crashing		

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM D T DAVIS BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRS EA DR V J RICE BLDG 4011 RM 217 1750 GREELEY RD FORT SAM HOUSTON TX 78234-5002
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM DG J RUBINSTEIN BLDG 333 PICATINNY ARSENAL NJ 07806-5000
1 (PDF)	ARMY RSCH LABORATORY – HRED ARMC FIELD ELEMENT RDRL HRM CH C BURNS THIRD AVE BLDG 1467B RM 336 FORT KNOX KY 40121
1 (PDF)	ARMY RSCH LABORATORY – HRED AWC FIELD ELEMENT RDRL HRM DJ D DURBIN BLDG 4506 (DCD) RM 107 FORT RUCKER AL 36362-5000
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM CK J REINHART 10125 KINGMAN RD BLDG 317 FORT BELVOIR VA 22060-5828
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM AY M BARNES 2520 HEALY AVE STE 1172 BLDG 51005 FORT HUACHUCA AZ 85613-7069
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM AP D UNGVARSKY POPE HALL BLDG 470 BCBL 806 HARRISON DR FORT LEAVENWORTH KS 66027-2302
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM AR J CHEN 12423 RESEARCH PKWY ORLANDO FL 32826-3276
1 (PDF)	ARMY RSCH LAB – HRED HUMAN SYSTEMS INTEGRATION ENGR TACOM FIELD ELEMENT RDRL HRM CU P MUNYA 6501 E 11 MILE RD MS 284 BLDG 200A WARREN MI 48397-5000
1 (PDF)	ARMY RSCH LABORATORY – HRED FIRES CTR OF EXCELLENCE FIELD ELEMENT RDRL HRM AF C HERNANDEZ 3040 NW AUSTIN RD RM 221 FORT SILL OK 73503-9043
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM AV W CULBERTSON 91012 STATION AVE FORT HOOD TX 76544-5073
1 (PDF)	ARMY RSCH LABORATORY – HRED RDRL HRM DE A MARES 1733 PLEASONTON RD BOX 3 FORT BLISS TX 79916-6816
8 (PDF)	ARMY RSCH LABORATORY – HRED SIMULATION & TRAINING TECHNOLOGY CENTER RDRL HRT COL G LAASE RDRL HRT I MARTINEZ RDRL HRT T R SOTTILARE RDRL HRT B N FINKELSTEIN RDRL HRT G A RODRIGUEZ RDRL HRT I J HART RDRL HRT M C METEVIER RDRL HRT S B PETTIT 12423 RESEARCH PARKWAY ORLANDO FL 32826

```
 1      ARMY RSCH LABORATORY – HRED
(PDF)   HQ USASOC
        RDRL HRM CN   R SPENCER
        BLDG E2929 DESERT STORM DRIVE
        FORT BRAGG NC 28310

 1      ARMY G1
(PDF)   DAPE MR   B KNAPP
        300 ARMY PENTAGON  RM 2C489
        WASHINGTON DC 20310-0300

        ABERDEEN PROVING GROUND

 12     DIR USARL
(PDF)   RDRL HR
          L ALLENDER
          P FRANASZCZUK
          K MCDOWELL
        RDRL HRM
          P SAVAGE-KNEPSHIELD
        RDRL HRM AL
          C PAULILLO
        RDRL HRM B
          J GRYNOVICKI
        RDRL HRM C
          L GARRETT
        RDRL HRS
          J LOCKETT
        RDRL HRS B
          M LAFIANDRA
        RDRL HRS D
          K DICKERSON
          A SCHARINE
        RDRL HRS E
          D HEADLEY
```