PERCEPTUAL SPACES FOR SOUND EFFECTS OBTAINED WITH AN INTERACTIVE SIMILARITY RATING PROGRAM

Gary P. Scavone[1], Stephen Lakatos[2], Perry R. Cook[3], Colin Harbke[2]

[1]Center for Computer Research in Music and Acoustics (CCRMA), Stanford University
[2]Department of Psychology, Washington State University
[3]Departments of Computer Science and Music, Princeton University

Corresponding Author's e-mail: gary@ccrma.stanford.edu

Abstract

Parametric signal-processing models of acoustic signals have recently begun to encourage some in the psychoacoustics community to reframe the problem of complex auditory perception of non-speech sounds from a descriptive perspective to a more hypothesis-oriented one. In particular, several recent studies have taken an ecological approach to timbre by positing that listeners infer the physical properties of sound-generating sources when they hear natural sounds. We describe a series of experiments that attempts to merge developments in the study of sound source perception and physical modeling to yield a better understanding of listeners' criteria in rating auditory timbre. The starting point for our efforts has been the need to obtain similarity ratings from human listeners for several hundred sounds in order to train an automated computer audio classifier. Traditional multidimensional scaling algorithms do not permit the testing of large stimulus sets because they require that listeners make all pairwise comparisons between stimuli, resulting in an exponential increase in the number of comparisons as a function of stimulus set size. In addition, it is difficult, if not impossible, for listeners to maintain stable comparison criteria across large numbers of comparisons. In order to circumvent such limitations, we have created an innovative graphics-based program for collecting similarity data for such large sets. The program initially assumes that the optimal perceptual space is two-dimensional, and listeners rate timbral similarities within this space. Additional dimensions can be added based on lack-of-fit measures for the initial two-dimensional space. Task demands for listeners are reduced through redundant mnemonic aids, and experimenters have considerable flexibility in specifying several adjustable stimulus comparison parameters. We have begun to use our program to investigate the role of mental imagery in listeners' evaluations of complex real-world sounds, as well as the degree to which such imagery is auditory or multimodal.

INTRODUCTION

The recent emergence of heuristics-based parametric signal-processing models of acoustical signals [1] provides opportunity to use such models to better understand human perception of sound source characteristics. In particular, the availability of real-time listener control over the parameters of such models now permits direct testing of a broad range of hypotheses about sound source perception. For example, Lakatos, Cook, and Scavone [2] used a probe-signal paradigm to demonstrate that listeners could attend selectively to the parameters of physically informed models of percussive musical instruments. Scavone, Lakatos, and Cook [3] also used a learning paradigm to examine how listeners acquire knowledge about the physical parameters of such percussive instruments, and the identities of the instruments themselves. Psychoacoustical data can also be used to fine-tune the parameters of such models and to aid in the selection of an optimal analysis/synthesis model for a given sound source. This paper focuses on our attempts to develop innovative techniques for obtaining similarity data for large sound sets in order to train our classifier.

Although multidimensional scaling (MDS) techniques have been used to study timbre for the past four decades, they are non-optimal for obtaining perceptual spaces for large numbers of stimuli. Traditional MDS methods require N*N comparisons from the complete pairing of N stimuli (or (N*(N-1))/2 comparisons for a half-matrix without diagonal). With 100 stimuli, participants must make a minimum of 4950 judgments; aside from considerations of fatigue, it is doubtful that one can maintain stable criteria across so many comparisons. Existing methods also set limits on participants' strategies for making comparisons within large stimulus sets: (a) constraining participants to make pairwise comparisons prevents them from adopting more complex and intuitive comparison strategies,

(b) Randomized stimulus presentation makes it impossible to return to previous comparisons if the participant wants to change his or her criteria based on current comparisons, (c) Continuous rating scales may encourage participants to think unidimensionally about stimulus relations in cases where a dimensional model may not be appropriate. Interactive MDS algorithms [4] have been designed that reduce the number of comparisons by using incomplete designs for a subset of stimulus pairs, selected randomly or according to mathematical criteria (see [5]), but they have met with limited success.

Given limited options for obtaining similarity data for large stimulus sets, we developed an interactive graphical program for collecting similarity data that reduces task demands with mnemonic aids and flexible comparison parameters. Our Linux-based program is inspired in part by Bonebright's [6] psychophysical comparison of the results obtained from a two-dimensional sorting environment with those of traditional pairwise comparisons. Our program provides flexibility in positioning, grouping, and classifying stimuli in the two-dimensional plane of the screen, and offers options for determining whether a two-dimensional interpretation is valid or whether additional dimensions are warranted. Although this approach stands in contrast to established MDS approaches, we find that an interactive environment gives participants welcome control over their comparison strategies. We outline below the main features of the program, and then describe our preliminary use of the program with 150 sound effects to examine how listeners' ratings change depending on whether listeners focus on the timbral properties of the sounds or on the mental images that the sounds generate.

PROGRAM FEATURES

The program offers several features that contribute to a robust and flexible comparison environment:

• The program provides a two dimensional palette in which sound items may be compared, contrasted, and grouped. The interface provides drag-and-drop functionality for sound item movement and placement. Sound playback is randomized when multiple items are selected.

• Several mnemonic cues help participants track the nature and extent of past comparisons, including: (1) a message box indicating the number of times a sound has been played and the sounds to which it must be compared (if pairwise comparisons are specified), (2) a corresponding visual cue that flashes the icons of those sounds to which a specific sound must be compared, (3) a "Remaining Comparisons" button that flashes all the sounds for which pairwise comparisons are still required, (4) A feature that progressively desaturates the color of an icon the more frequently it is played.

• Participants can create categories with color labels.

• Resources are provided for entering verbal descriptors for individual stimuli.

• The program provides the option for obtaining confidence ratings regarding the final position of each stimulus once the participant finishes all required comparisons. Confidence ratings provide the information concerning sounds that do not "fit" well in the two-dimensional space, either because additional dimensions may be required to account for the variance associated with such stimuli or because certain stimuli have unique characteristics not shared by any other stimuli in the set.

Figure 1. *An illustration of the graphical interface with a participant's final stimulus groupings.*

• Following the completion of required comparisons, options are provided for collecting traditional pairwise similarity ratings for all stimuli within each of the participant-defined categories, as well as similarity ratings for a randomly selected subset of stimuli drawn across all such categories.

• Comprehensive data output is provided for statistical analysis and multidimensional scaling.

TIMBRE/MENTAL IMAGERY TESTS

In our first use of the classifier, we obtained similarity judgments for 150 complex sounds from each of 28 participants, with the goal of using the resulting dissimilarity matrices to train an automated computer audio classifier. Stimuli were sound effects recorded from a variety of effects libraries (e.g., BBC, O'Connor, Sound Effects Toolkit). A specific feature of the sound effects is that all of them are made and controlled by human gestures; further, they are generally single- or multiple-object systems within a contained space that the human interaction can control. Original loudness and duration values were left unaltered since equalization would have unduly altered/truncated signal content regarding source properties. Since most of these stimuli evoke strong mental images of the sources or objects generating them, we compared participants who were instructed to focus on the stimulus timbre with those who focused on the mental image generated by each stimulus. Participants were asked to generate 5-15 non-overlapping stimulus categories using the interface, and participants in the mental imagery condition were also asked to provide verbal descriptors for each stimulus. In addition to participant-directed organization of stimulus icons on the screen, participants were required to perform 150 randomly-selected pairwise comparisons from among all such possible comparisons, in order to encourage them to make at least one pairwise comparison for each stimulus. To test the validity of a two-dimensional assumption for this space, we subsequently obtained pairwise similarity ratings from each participant for all within-category comparisons, as well as a randomly selected subset of across-category comparisons. Each participant took approximately 10-14 hours to complete the experiment.

Preliminary results from the two conditions are shown in Figures 2 and 3 for the timbre and mental image conditions, respectively. The spaces in these figures were generated by submitting dissimilarity (distance) matrices for all subjects in each condition – computed from the distances between icons on the two-dimensional interface - to the multidimensional scaling program Clascal [7]. A clear overall difference in the organization of the two spaces is discernable, with most sounds neatly clustered according to source attributes in the mental imagery condition, while sounds in the timbre condition are grouped in much less interpretable ways. We are currently working on isolating acoustical correlates for the dimensions of the spaces, although the temporal complexity of the sound effects make the extraction of correlates like spectral centroid and rise time problematic. Most participants in the timbre condition noted the difficulty of applying the traditional operationalization of timbre as sound quality independent from pitch and loudness to sounds of such complexity as those tested here. Perhaps our most interesting general observation was the degree to which participants, when provided with the opportunity to choose their own strategies for organizing and grouping sounds according to timbre or mental imagery, became engaged in the task, and the degree to which they avoided making strictly pairwise comparisons whenever possible. In tape-recorded interviews, many participants remarked that the required pairwise comparisons seemed to be an unintuitive and artificial method for comparing sounds, and that they rarely engaged in it when given the option of other comparison strategies. In sum, our graphical interface may serve not only as an opportunity to develop novel data collection strategies for large stimulus sets, but also to test long-held assumptions about the appropriateness of more traditional comparison techniques for such contexts.

REFERENCES

[1] Cook, P. R. (1997). Physically inspired sonic modeling (PhISM): Synthesis of percussive sounds. *Computer Music Journal*, **21**, 38-49.

[2] Lakatos, S., Cook, P. R., & Scavone, G. P. (2000). Selective attention to the parameters of a physically informed sonic model. *Acoustical Research Letters Online*.

[3] Scavone, G.P., Lakatos, S., & Cook, P.R. (2000). Knowledge acquisition by listeners in a source learning task using physical models. Paper presented at the 139th Meeting of the Acoustical Society of America, Atlanta, GA.

[4] Young, F. W., & Cliff, N. (1972). Interactive scaling with individual subjects. *Psychometrika*, **37**, 385-415.

[5] Spence, I., & Domoney, D. W. (1974). Single subject incomplete designs for nonmetric multi-dimensional scaling. *Psychometrika*, **39**, 469-490.

[6] Bonebright, T.R. (1996). An investigation of data collection methods for auditory stimuli: Paired comparisons versus a computer sorting task. *Behavior Research Methods, Instruments, & Computers*, **28**, 275-278.

[7] Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, **58**, 315-330.
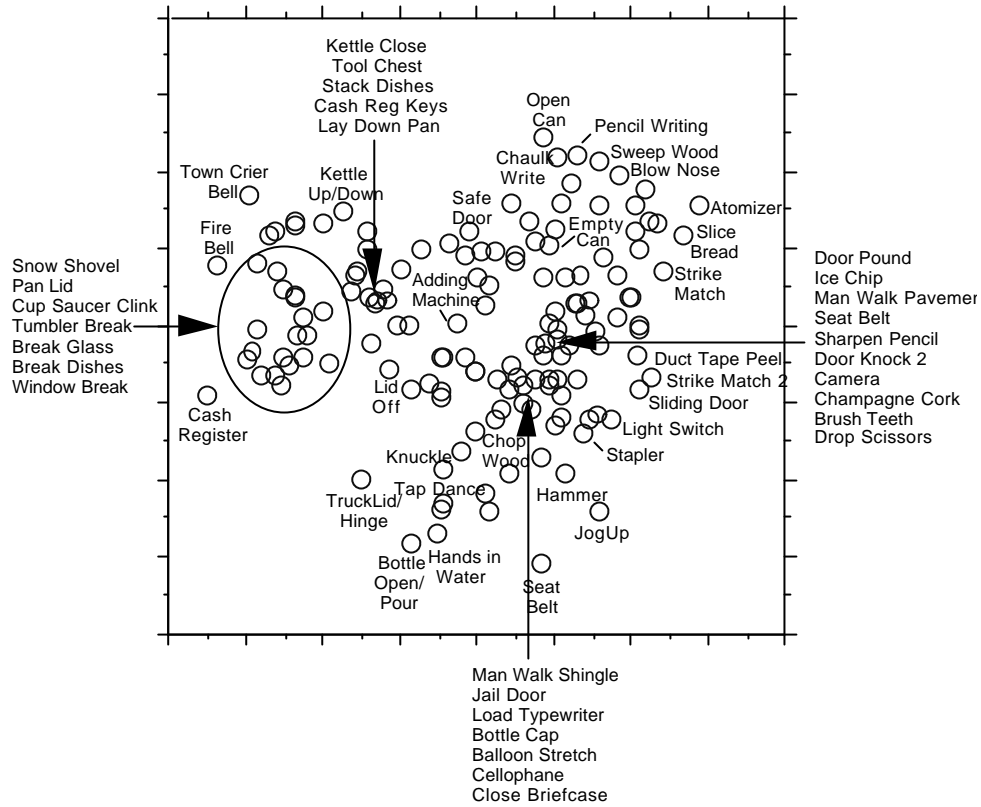
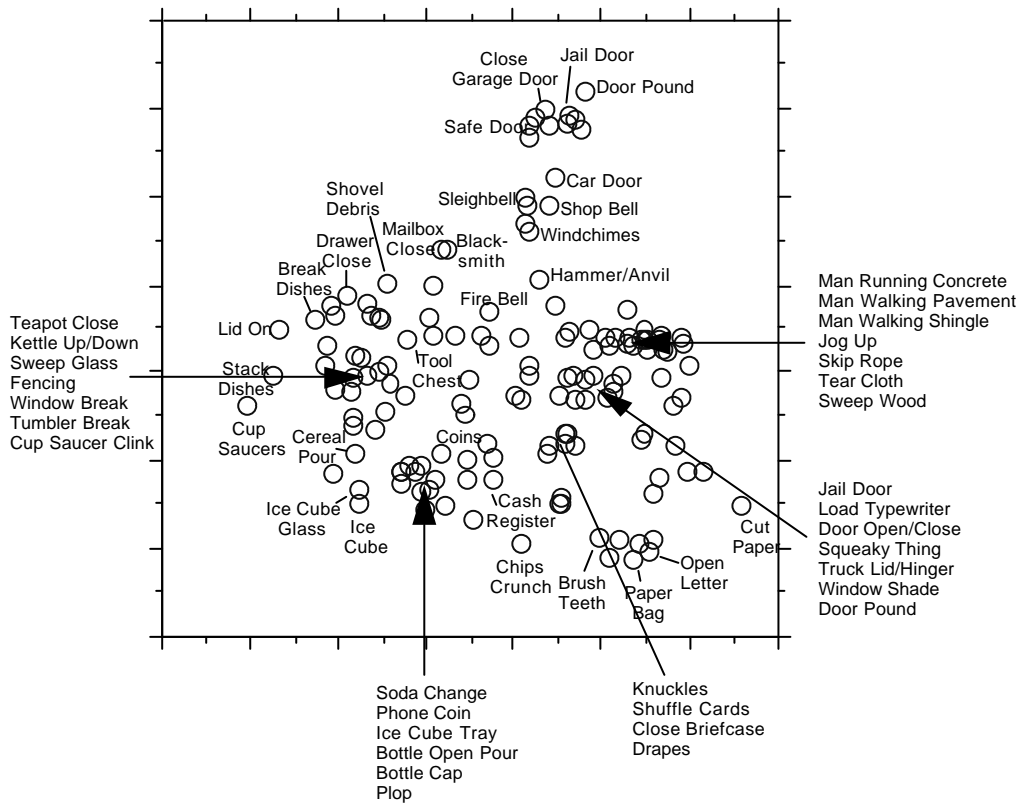Figure 2. *MDS solution for participants' spatial arrangement of stimuli in the timbre condition.*

Figure 3. *MDS solution for participants' spatial arrangement of stimuli in the mental-imagery condition.*