THE EFFECT OF PITCH SHIFTS ON THE IDENTIFICATION OF ENVIRONMENTAL SOUNDS: DESIGN CONSIDERATIONS FOR THE MODIFICATION OF SOUNDS IN AUDITORY DISPLAYS

Brian McClimens, Justin Nevitt, Cheng Zhao, Derek Brock and James A. Ballas

Naval Research Laboratory,
Washington, DC, 20375

mcclimen@aic.nrl.navy.mil, jrn5a@virginia.edu,
brock@itd.nrl.navy.mil, ballas@itd.nrl.navy.mil

ABSTRACT

To examine the plausibility of dynamically adjusting the sounds presented by an auditory display, a study addressing the effects of pitch shifting on the identifiability of a set of forty-one environmental sounds was carried out. The sounds were shifted both up and down in pitch and presented to listeners who were asked to identify them. Results show that pitch shifting is detrimental to the identification of environmental sounds, suggesting that benefits gained from dynamically manipulating sounds in an auditory display must be carefully weighed against perceptual effects on their identifiability. Results also indicated that the sounds in our study better retained their identity when shifted down in pitch than when shifted up. This result however is believed to arise from confounds in the study.

1. INTRODUCTION

In an effort to reduce the manpower required to operate vessels, the U.S. Navy is funding research with the goal of improving the efficiency of operator interactions with onboard systems. Naval operators are expected to conduct several tasks simultaneously, and the multimodal watchstation, as described in Osga [1] has been designed in order to help effectively manage critical responsibilities. Although considerable effort has gone into making efficient use of visual information, workstations are showing a trend towards utilizing more visual space; the multimodal watchstation, for instance, is designed to be used with up to four separate displays. In the anticipation of a visual saturation point where more screen real estate no longer translates into performance gains for the user, we have been studying the effective use of auditory displays.

2. BACKGROUND

Currently, auditory signals are a greatly underutilized resource in user interface design. However, studies at the Naval Research Laboratory (NRL) have recently shown the utility of auditory displays for purposes of managing operator attention in multitask environments (e.g., Brock et al. [2]). It is reasonable to assume that as the benefits of informational uses of sound become more widely appreciated, interfaces will grow dense with auditory information in much the same way that they are now becoming saturated with visual information. Therefore, it is important that we make concerted efforts towards utilizing auditory space as efficiently as possible.

Sounds engineered for alerting purposes often make use of a variety of techniques to maximize their perceptual salience. These include solutions like modulating fundamental frequencies, covering a broad spectrum of frequencies, and increasing the amplitude of the signal. Many of these techniques cause alerts to spread out across auditory space (i.e., the alerts have a high power level in many frequency ranges). In situations where one alert should be clearly dominant (e.g., fire alarms, air raid sirens), this is acceptable. In a complex auditory display, there will be a large set of alerts and sonifications, some of which may vary in urgency over time. The alerts will have to be able to grab the user's attention without drowning out other auditory information that may be sounding at the same time. Well thought out auditory displays can avoid many presentation conflicts through the careful design of sound materials. Unfortunately, in dynamic systems it may not be possible to resolve all potential conflicts during the design phase. This is the one of the main motivations for the concept of self organizing auditory displays.

Self-organizing auditory displays (SOADs), as envisioned in Brock et. al [3], are systems which autonomously adjust the audio signals they present in order to maximize their effectiveness. One of the integral features of SOADs is the ability to make adjustments to auditory materials in order for the signal to transmit an appropriate level of information in a wide variety of situations. The SOAD must also be able to adjust auditory signals to minimize the interference between sounds that are being presented simultaneously. In many situations, it will be crucial for the user to recognize and respond to an alert as quickly as possible. In order to provide SOADs with the ability to manipulate sounds without overly increasing operator reaction times, we must first gain a better understanding of the ways in which we can reliably modify auditory signals without compromising their identifiability. In this paper, we review an experiment that explores the effects of pitch shifts on the identification of everyday sounds.

3. EXPERIMENT SETUP

3.1. Environmental Sounds

In the formative stages of this study, we were faced with the choice of using either engineered sounds or environmental sounds to perform pitch shifts on. Gygi [4] describes the latter of these as complex, naturally occurring, non-speech sounds, and notes that other researchers have used terms such as common, familiar, everyday, or naturalistic to refer to the same class of sounds. While auditory displays typically make use of

sounds designed for specific purposes, several advantages of using a base set of environmental sounds led us to reject the use of engineered sounds in this initial experiment. We believe that the identities of certain sounds are inherently more resilient to the effects of pitch shifting than those of other sounds. Presently, it is unclear whether or not specific properties are responsible for this effect. It was decided that a set of engineered sounds would be more likely to share common traits that might skew the results of our study. Additionally, it was important for listeners to have some pre-existing familiarity with the stimuli since we did not have the time required to train subjects on a set of novel sounds. Due to these considerations, it was determined that environmental sounds would be most appropriate for the present study.

Ballas [5] describes a series of experiments concerning the identification of brief everyday sounds. The first of these was concerned with the relationship between response times and causal uncertainty. This experiment made use of an uncertainty statistic that assigned a numeric value to the variation among answers provided by listeners attempting to identify the stimuli. Correlations between the resulting statistic, referred to as the 'measure of causal uncertainty', and mean identification times were significant. The current study makes use of Ballas' sound materials and the uncertainty statistic he used to measure the variability among his listeners' responses.

The forty-one sounds used in [5] are edited recordings of environmental sounds that are each approximately 650 milliseconds in length. These sounds represent a wide variety of sources, including a bugle call, bacon frying in a pan, and a telephone ring. To keep these sounds as distinct from speech as possible, no vocalizations of any sort were used. Ballas observed a wide range of identification accuracy, with the most easily recognized sound being correctly identified by one hundred percent of all listeners, and the least recognizable sound being correctly identified by only four percent.

3.2. Sound Manipulation

Several manipulations can be used to achieve slight variations of sounds that may be able to enhance their perceptual salience in a dynamic context. In preparation for this study, several methods of modifying sounds were explored. Among these were several types of highpass, lowpass, band, and comb filters, amplitude modulation, vibrato or frequency modulation, and pitch shifting. We chose to focus on pitch shifting in this experiment because it is a relatively straightforward modification that only slightly distorts the original signal. Additionally, pitch shifting is a modification that is likely to help a sound fit into an auditory niche, similar to the ecological notion described in Krause [6]. The key frequencies for a given alert could be shifted away from a frequency band containing high levels of noise in a given environment to frequencies with relatively little background noise. Although it is expected that any pitch shift will negatively impact the identifiability of a sound, the ability to adjust a sound in such a way that it fills an unoccupied auditory niche is expected to outweigh this drawback.

3.3. Procedure

Five versions of each of the forty-one sounds from Ballas' study were used in the current study. For each sound, we used the original sound and versions of the sounds that had been shifted up in pitch by six and twelve semitones, and down in pitch by six and twelve semitones. These modifications were made using a pitch shifting effect in Sound Forge [7] that preserved the duration of the sounds. In combination with the original sounds, these modifications gave us a set of 205 sounds.

The sounds were divided into five sets of forty-one stimuli. Each set contained exactly one version of each sound with a mixture of pitch shift types across the sounds. Five classes of high school students were each presented with one such set of stimuli. Sounds were presented simultaneously to the entire class in random order, and students were given thirty-five seconds to write down responses for each sound they heard. Students were asked to identify each sound, indicate how difficult they thought it was to identify, as well as whether or not they believed the sound had been shifted up or down in pitch and by what amount. The response form had a blank line allowing for an open-ended identification of the sound, the difficulty rating was a scale from one to five, and the pitch shift scale offered answers of -12, -6, 0, 6 and 12.

4. ANALYSIS

4.1. Identification

The primary statistic used in our analysis was a measure of how accurately sounds from each pitch shift category were identified. Each listener's responses were grouped by the amount of pitch shift that had been applied to the corresponding stimulus. For each group of responses, the average percentage of correct responses was calculated. The mean of the listeners' scores in each pitch shift category are displayed in Figure 1. These means show that, as expected, sounds which have undergone no pitch shifting are easiest to identify. Additionally, correct identification of the sounds appears to be inversely related to the magnitude of the pitch shift.

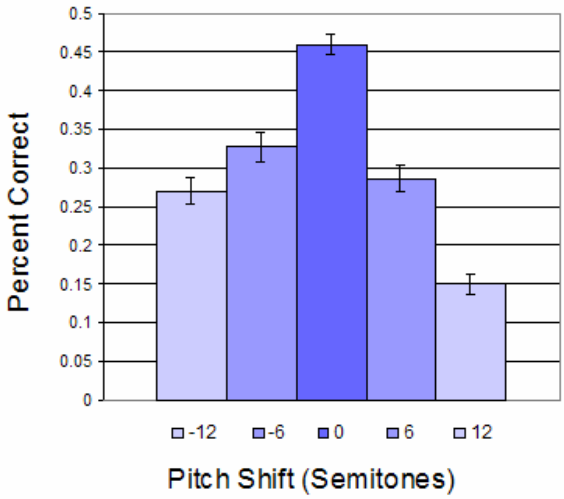

Figure 1. *Mean percent of sounds identified correctly in each of the conditions. Error bars show the standard error of the mean.*

A two-way ANOVA comparing magnitude (absolute value) of pitch shift and direction found significant main effects for magnitude and direction, as well as a significant interaction between magnitude and direction ($F(1) = 56.451$, $p < .001$, $F(1) = 12.196$, $p<.005$, $F(1) = 6.669$, $p < .05$ respectively). The significant effects of direction and the interaction are believed to be caused by an echo effect which is introduced by the pitch shifting algorithm. The reasons for and the implications of this are explored in the discussion section.

4.2. Variation in Responses

Scores of variability for each pitch shift and sound combination were also analyzed. To get this measure, all answers for each sound were sorted into category bins. In [5], Ballas uses an uncertainty statistic which is calculated by the following formula:

$$Hcu_s = \sum_{b=1}^{n_s} (a_{bs}/a_{ts}) \log_2 (a_{bs}/a_{ts}) \qquad (1)$$

where Hcu_s is a measure of the variability in the answers provided for sound s, a_{bs} is the number of answers sorted into bin b for sound s, a_{ts} is the total number of answers given for sound s, and n_s is the total number of separate bins for sound s.

When all answers for a sound fall within the same bin, the equation for Hcu_s yields zero. In all other cases, the equation yields a negative number, with the magnitude of that number increasing until it reaches its maximum when all answers fall into separate bins. The maximum possible magnitude is dependent on the maximum value of n, which is equal to the number of subjects giving an answer for sound s. In Ballas (1993), this number remained equal across all sounds. In the current experiment however, there were five separate groups of participants, each of which differed in size, so a normalization of this score was required. This equation:

$$V_s = \frac{Hcu_s}{\log_2(1/a_{ts})} \qquad (2)$$

transforms values of Hcu_s, such that $0 \leq V_s \leq 1$, and the magnitude of V_s is not dependent on the number of subjects as Hcu_s was.

Equation 2 was used to measure the certainty with which subjects identified sounds. Results from this measure are very similar to the results measuring the correctness of answers. The main difference between these two methods of measuring responses is that sounds which are consistently identified as one of a small number of things, or sounds which are consistently attributed to an incorrect source while rarely being classified correctly do not affect the variability as much as sounds which are very difficult to identify. Separate means of variance for each of the pitch shift conditions were calculated, and are shown in Figure 2. The results from this analysis closely, but inversely, mirror the results of the correctness analysis, with least variation occurring in sounds that had not been shifted, and variability increasing with the magnitude of the pitch shift.

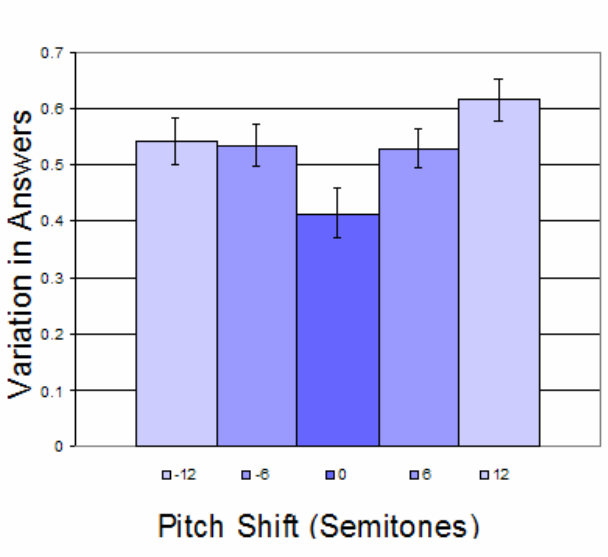

Figure 2. *Variation among answers in each of the conditions. Error bars show the standard error of the mean.*

4.3. Ranking and variation consistency

In order to determine if certain sounds were affected by pitch shifts in significantly different ways from the general trends discussed above, we analyzed the rankings of individual scores across conditions. Scores were ranked by the percentage of people who correctly identified them in their original form. The sounds in each category of pitch shift were also ranked in this way. Then the ranks of the sounds in their original form were compared to their ranks after pitch shifts had been applied. The results of this analysis are shown in Figure 3.

If the pitch shifts had a consistent effect on all of the sounds in this study, the resulting graph would show a straight diagonal line. If a data point for a sound falls above this line, it indicates that the sound was affected more adversely than the average sound by the pitch shifts. Conversely, the data points which fall below this line indicate sounds that may be more resilient to pitch shifts. A similar analysis was applied to the variability among the answers given for each sound, and results are shown in Fgure 4.

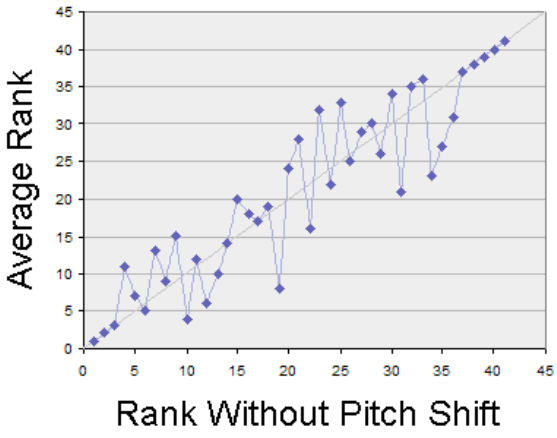

Figure 3. *Percent correct rank consistency plot.*

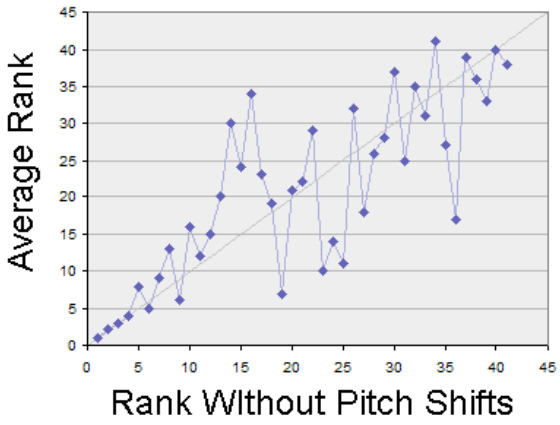

Figure 4. *Variation in answer rank consistency plot*

5. DISCUSSION

The three main results of this study on the effect of pitch shifting on the identifiability of environmental sounds were the significant effects of magnitude, direction, and the interaction between these factors. The effect of pitch shift magnitude on sound identification was expected, and supports the notion that any benefits gained from modifying a sound within an audio user interface must be carefully weighed against detrimental effects to its identifiability. Conservative use of pitch shifting is still likely to be a useful technique for self-organizing auditory displays, however more research is needed to establish practical parameters for this type of modification.

The results suggesting that shifting sounds up in pitch was more detrimental to their identification than a comparable shift down in pitch came as a surprise. However, we have a plausible explanation for this effect. When an audio signal is subjected to a pitch shift which raises its pitch, an echo effect is typically introduced into the sound. As the amount of pitch shift is increased, the delay between the original signal and the echo is increased correspondingly. Although the echo did not seem to be a prominent feature in most of the sounds shifted up by twelve semitones, it was perceptible. In addition to this confound there is another effect that could contribute to the difference between the positive and negative shifts. Many of the sounds in this study contain a fair amount of background noise that is relatively low in frequency. When sounds were shifted down in pitch, this background noise became less perceptible. Sounds that were shifted up often had the effect of making the background noise much more salient. We believe that the significant effect found in relation to the direction of pitch shift as well as the interaction effect between direction and magnitude are due mostly to the echo and effects on noise which were introduced during positive pitch shifts. In a more extensive study, we would have liked to explore issues relating to the effects of different pitch shifting algorithms.

The two sounds which were most easily recognized across all conditions were a bugle call and an automatic rifle. The automatic rifle had a very distinctive temporal pattern of sound bursts which carried very little tonal information. Because the duration of these sounds were preserved during the pitch shift, the temporal pattern of the automatic rifle was not affected by the pitch shifts. This suggests that a strong rhythmic or temporal component to auditory alerts may increase their chances retaining their identity through manipulations related to pitch. The bugle call is believed to have retained its identity because the relative pitch differences between notes in the melody were preserved when the pitch of the entire sound was adjusted. We believe that alerts containing a distinct series of tones will be able to be recognized by the relationships between tones, allowing the whole alert to be shifted up or down in pitch without adversely affecting the alert's identity. In the future, we would like to test these notions about which characteristics preserve identifiability when pitch shifts have been applied. Further study would allow us to more effectively design sounds for use in dynamic settings by self-organizing auditory displays.

6. ACKNOWLEDGEMENTS

This work was funded by the Office of Naval Research. The authors wish to thank Greg Trafton for his analytical advice. We would also like to thank Thomas Jefferson High School for Science and Technology and F.J. Berenty for facilitating Cheng Zhao's student internship and research.

7. REFERENCES

[1] G. Osga, "21st century workstations: active partners in accomplishing task goals," *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meething,* San Diego, CA, 2000.

[2] D. Brock, J.L. Stroup, J.A. Ballas, "Effects of 3D auditory display on dual task performance in a simulated multiscreen watchstation environment," *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, Baltimore, MD 2002.

[3] D. Brock, J. A. Ballas, B. McClimens, "Perceptual issues for the use of 3D auditory displays in operational environments," *Proceedings of the International Symposium on Information and Communications Technologies*. Trinity College, Dublin Ireland, 2003.

[4] B. Gygi *Factors in the Identification of Environmental Sounds,* Doctoral Dissertation, Indiana University, Bloomington, IN, 2001.

[5] J. A. Ballas, "Common Factors in the identification of an assortment of brief everyday sounds," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 19, pp. 250-267.

[6] B.L. Krause, "Bioacoustics, habitat ambience in ecological balance," *Whole Earth Review.* Vol. 57, Winter, 1987.

[7] Sound Forge Sonic Foundry Version 5.0, 2001.