AUDITORY PROGRESS BARS PREFERENCE, PERFORMANCE, AND AESTHETICS

Peres, S.C.[1], Kortum, P.[2] and Stallmann, K.[2]

[1]University of Houston Clear Lake
[2]Rice University
Houston, TX
peressc@uhcl.edu

ABSTRACT

This paper presents the comparison of Auditory Progress Bars using segmented cello tones to those using sine tones in an on-hold telephony setting. Previous research suggests that for segmented sine tone APBs, there is an interaction between APB type (pitch or duration) and APB polarity (increasing or decreasing). However, for the cello tones, there was a main effect of direction, with increasing APBs resulting in better performance than decreasing APBs, and there was no effect of APB type (pitch or duration). As anticipated, overall performances were very similar for both types of segmented APBs. Contrary to expectations, users gave the cello tone APBs equally low subjective ratings to those they gave the sine tone APBs. Whole-song APBs produced very positive subjective ratings and performances similar to the sine and cello tone APBs.

[Keywords: Aesthetics, Auditory progress bar, Time perception, Telephony]

1. INTRODUCTION

Many businesses provide customer support phone numbers that allow customers to call when they wish to seek technical help, lodge a complaint, or conduct general business with the company. The call to customer support is one contributor to the consumer's overall user experience and can play a significant role in how positively the customer views that company.

Along with the successful resolution of the consumer's problem or completion of the consumer's task, one key variable influencing the consumer's judgment of the call is the time they spend on hold waiting for a customer service representative to assist them. To create a situation in which a customer would never need to be put on hold, companies would have to always have more representatives than calls. However, this practice is cost prohibitive and thus companies attempt to balance their labor costs with the cost (in terms of customer satisfaction) of putting customers on hold for a brief time.

Human's perception of time varies, depending on the stimulus (or lack thereof) that fills the time a person is put on telephone hold [1, 2]. Callers tend to overestimate the time they have been on hold when they are subjected to silence during the wait, and tend to underestimate the time when they are given some kind of cognitive activity during the on-hold interval (e.g. music, advertisements, interactive speech browsers). Creating the right stimuli to fill the on-hold interval has the potential to increase (or maintain) customer satisfaction while simultaneously meeting the business need to place customers on hold.

Regardless of whether callers over or under-estimate the actual hold time, the shorter they perceive the hold time to be, the more satisfied callers are with the experience [1]. Thus, communications research has typically used retrospective time estimation protocols to evaluate on-hold paradigms as this measure is directly related to customer satisfaction. Furthermore, on hold paradigms that result in highly accurate retrospective assessments of the time on hold may also be providing information to the listeners about their position and progression through the queue. Although other measures may more accurately identify if listeners are getting specific location and progression information from these stimuli, the studies described in this paper use the standard retrospective protocol in order to allow direct comparisons to on hold paradigms using different stimuli.

The stimulus used to represent the amount of time a person will be on hold is called an Auditory Progress Bar (APB) [3]. The APB was originally envisioned as a supplement to the standard visual progress bar that is ubiquitous on computer displays. The visual progress bar is wonderfully elegant and conveys a significant amount of information regarding the temporal properties of the process. For those visual progress bars that "fill" from left to right, observing the location of the fill allows the user to ascertain the approximate percentage of the temporal sequence that has occurred. Further, by examining the rate of change of the fill in relation to the visual endpoint of the fill box, the user can make an assessment of how much longer the process is going to take. The idea proposed by Crease and Brewster [3] was to provide a supplementary interface using sound so that a user who was distracted from the primary visual interface could still receive the information concerning the temporal properties of the process.

We have extended the work of Crease and Brewster [3] into situations where there is no visual representation and the sound must carry the entire informational load. Polkosky and Lewis [4] used the rate of ticking to indicate temporal information in a hold situation. Kortum, Peres, Knott and Bushey [5] used regularly spaced stimuli that changed in either pitch or duration for the same purpose. In both cases, the display proved effective, allowing listeners to make relatively accurate time estimations, but were generally perceived as fatiguing and annoying. Others [6, 7, 8] have studied the effect of using musical stimuli in order to convey temporal information. In these studies, the stimuli were generally continuous songs and played for the duration of the hold time.

In all of these in previous studies, performance was measured by determining how much callers misestimated their time on hold. This value was computed by taking the absolute value of the over or under estimation and dividing that by the actual hold time. This gave a proportion of misestimation, relative to the duration of the particular hold time. It was found that callers' time estimates became more accurate as the length of the hold time increased, for both tonal and whole song APBs [5, 8]. This was unsurprising since small errors in short hold times result in larger proportional errors than do small errors in large hold times.

Figure 1 shows the performance results from two of these previous studies [5, 8]. This graph suggests an interaction between APB type (pitch [sine-pitch-increasing.wav] or duration [sine-duration-increasing.wav]) and APB polarity (increasing [sine-pitch-increasing.wav] or decreasing [sine-duration-increasing.wav]). Better performance was observed in the increasing condition for the duration APB and in the decreasing condition for the pitch APB. However, this interaction was not significant (p=0.067).

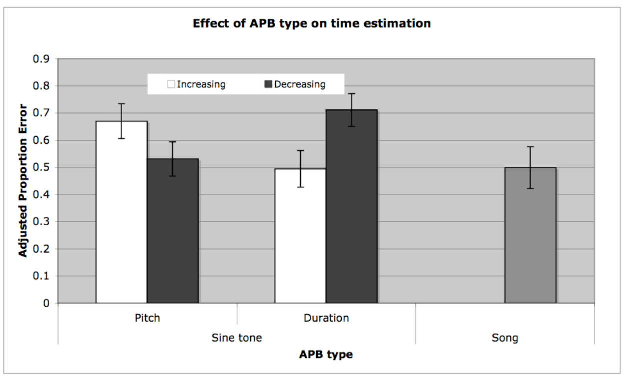

Figure 1. *Mean adjusted proportion of the error estimate for five APB types, 4 sine tone APBs and Whole song APB.*

The second metric collected was the user's satisfaction with the stimulus. As can be seen in Table 1, ratings of the segmented sine tone stimuli were quite low with users generally reporting that they strongly disliked these stimuli. The whole-song APBs [Whole-Song], however, produced very positive satisfaction ratings. The preference ratings of the APBs in these previous experiments did differ significantly (F (4,297)=51.14, p<. 001). A Tukey's pairwise comparison confirmed that the whole-song APB was rated significantly higher in mean satisfaction than all of the other four tonal APBs (all p's <.001).

APB type	Average Preference
Pitch Increasing	1.69
Pitch Decreasing	1.67
Duration Increasing	2.84
Duration Decreasing	3.08
Whole Song	4.75

Table 1. Mean rating for subjective likeability of the stimuli, on a 7-point scale, where higher scores indicate a more likable stimulus.

These findings illustrate that Whole-Song APBs resulted in similar performance but more positive subjective ratings than segmented sine tone APBs. This suggests that the design of the sine tone APBs (particularly the pitch decreasing and duration increasing APBs) provided as much information about time as the familiar songs used in the Whole-Song APBs. Participants' preference for the Whole-Song APB also suggests that the elements of a pleasing aesthetic that are typically used in popular music, may have improved participants' subjective rating of the APB.

The study presented in this paper was designed to extend our knowledge of APB structure by using segmented stimuli (following [5]) that are composed of recorded cello tones. The cello tones were produced by a live performer and were composed in a 12-tone tempered pitch system. This study allowed for an empirical investigation of the effects of aesthetic elements (in this case the use of a natural, versus synthetic, timbre and a tempered pitch system) on performance and preference with this type of stimulus.

2. METHOD

2.1. Participants

Twenty-five undergraduate students (13 females and 12 males) from Rice University participants participated in the study for course credit. They were all screened (self-report) for normal or corrected to normal hearing. Participants were initially told that they were participating in a voice recognition study to mask the true purpose of the experiment.

2.2. Stimuli

Sixteen different APBs were used in the study. Each consisted of a series of cello tone segments with a brief announcement at the beginning and end. Eight of the APBs varied in pitch over time. For four of the stimuli, the pitch rose from start to end, and had overall durations of 30s, 60s, 120s, and 240s. The other four stimuli had a falling pitch pattern with the same durations. The remaining stimuli had segments that varied in duration over time, with the duration increasing in four of the APBs and decreasing in the remaining four APBs. Replicating [7], this resulted in 4 different APB designs: Pitch Increasing (PI), Pitch Decreasing (PD), Duration Increasing (DI), and Duration Decreasing (DD).

Each APB had a 9-second introductory announcement that said "Thank you for calling. All operators are currently assisting other customers. Please hold for the next available operator." Three seconds following this announcement, the first segment played. Approximately 13.5 seconds of silence followed, and then the next segment in the sequence played. The exact inter-stimulus intervals varied slightly because the segment set (segment + ISI) is considered to be 15s from the start of $segment_n$ to the start of $segment_{n+1}$. In addition the individual segments varied in duration because they were recorded with a live cello. One second after the last segment in the sequence was played, there was a 9 second announcement that enabled the participants to make an inquiry concerning their current bill balance. This was the portion of the test that callers have been led to believe was testing a voice recognition application.

The tone segments used in the 'pitch' stimulus of this experiment here were designed to be aesthetically improved versions of stimuli used in an earlier experiment. In that experiment, sine tones were used to generate sounds rather than a live instrument. For the new stimuli, a live cellist was used to record the fragments. Although this choice has obvious benefits in terms of the 'naturalness' of the sound, it also allowed for less control over the exact details of the sound and introduced variations in duration, fluctuations in pitch, and changes in internal rhythmic declamations. Given the existence of these local variations, precautions were taken to keep the live performance within prescribed boundaries that were consistent with the earlier stimulus (including durations of ISIs, timing of overall stimulus, number of elements present, general patterns of rising/falling pitch and increasing/decreasing duration).

In the initial experiment, sine tones were used in stepwise progressions (rising and falling) at contiguous intervals of 100 hertz. Since the 100 Hz intervals remained constant throughout the rise and fall, the perceived pitch intervals changed with the rise and fall of the stimulus. Using a live instrumentalist, it was impractical to consider 100 Hz intervals. Thus it was decided to make the rise and fall using a common scalar pattern (Mixolydian mode). In its ascending form, the first note of the fragment rose in stepwise motion to the last note. This basic motion was elaborated with an octave leap and a passing tone [Pitch-Cello-Increasing.wav]. In the descending stimulus, the retrograde form of the fragment was used. The retrograde form was chosen for descent primarily because it allowed the corresponding content of ascending and descending versions to be identical and reversed (so that the stepwise pattern is now falling) [Pitch-Cello-Decreasing.wav].

For the 'duration' stimulus, again, it was impractical to notate with computer-like precision for a live performer. Instead, the performer used traditional musical notation to play an increasing duration. In this case, each stimulus increased by a sixteenth note value (where the quarter note rate is one per second). This gave an approximate increase of .25 seconds for each new stimulus. Again, the retrograde form of this progression was used for decreasing durations allowing the corresponding content to be identical [Duration-Cello-Increasing.wav, Duration-Cello-Decreasing.wav].

In the design phase of this project, many other stimulus designs submitted for this experiment were rejected because they veered too far away from the original constraints placed on the stimuli of the earlier experiment. The designs of the current stimuli were chosen because they were most closely aligned with the earlier stimuli, allowing for a direct comparison of performance results. Many of the rejected designs dealt more specifically with providing references that oriented the listener relative to the start and end times of the stimuli. This emphasis addresses more specifically our goal to provide location information concerning where a user is in an ongoing temporal process. We plan to use these other designs in future experiments.

2.3. Procedure

Participants in the study signed an IRB-approved consent form and were briefed on the general procedures used in the study. Participants were told that they would be making several calls to a telephone call center in order to test a customer service voice recognition system under development. They were also told that since the system is in development, and multiple locations are calling in to test the system, they might have to wait for the speech recognizer to become available.

Participants placed the calls on a standard Western Electric 2500 series analog phone that was hooked directly to a personal computer through the sound port. Stimuli were initiated via the PC, and no dial tone was present. Participants made one call at a time and heard one of the APBs designs (PI, PD, DI, or DD), counterbalanced across participants. Each participant heard all four durations (30s, 60s, 120s and 240s) of a single type of APB. At the end of each call, participants were asked to estimate the amount of time they were on hold. Participants were also asked a number of other questions concerning the aesthetic qualities of the stimulus and their overall satisfaction with the call, after which they initiated the next call.

2.4. Measures

The two constructs of interest for this study were performance and preference.

2.4.1. Performance

Identical to previous studies [5, 8] performance was measured using the proportion of mis-estimation or error made for each trial. As mentioned previously, this was calculated by dividing the absolute value of the over or under estimation by the actual hold time. The distribution of these scores was skewed, so a square root transformation was done to normalize the distribution, giving the adjusted proportion error scores.

2.4.2. Preference

To measure participants' preference ratings, the responses from the question, "I liked the music" was used. Participants responded to this question using a 7-point agreement scale with 1 being Strongly Disagree and 7 being Strongly Agree.

3. RESULTS

3.1. Effects of Practice and hold time

For both performance and preference, there were no effects of practice or hold time (all p's >.10). The fact that performance did not differ by hold time for the Cello Tones is not consistent with the results from the study using Sine Tones (p=0.008) or Whole Songs (p<0.001). It is important to note that the different findings for the cello tones may be due to the low power of the current study (Power=0.225). Thus, more data will be collected to increase the power and confirm this finding.

3.2. Comparing APB types

The primary goal of the current study was to compare the performance and preference associated with cello tone APBs to sine tone and Whole Song APBs. Thus the data from these three types of APBs were combined into one dataset. Additionally, given that there was not an effect of practice, the trials were treated as independent observations. Thus all analysis were 2 (Dimension: Increasing and Decreasing) X 2 (Dimension: Pitch and Duration) X 3 (APB type: Sine tone, cello tone, and Song) Factorial ANOVAs with either performance (adjusted proportion error) or preference (how much they liked the sounds) as dependent variables.

3.2.1. Performance

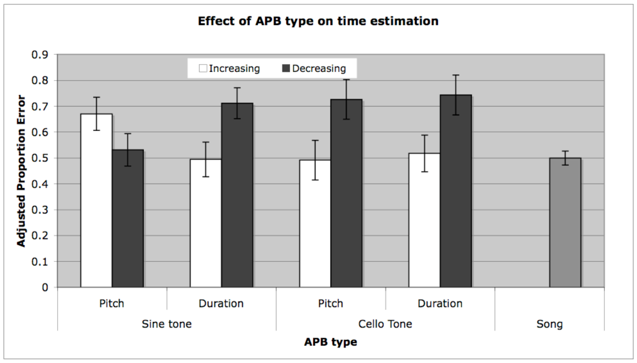

Figure 2. *Adjusted mean proportion of the error estimate by APB. Increasing APBs are in white, Decreasing APBs are in dark grey, and the whole song APBs are in light grey. Error bars represent the standard error of the mean.*

Figure 2 shows the means adjusted proportion error scores by APB. The figure shows that there is a main effect of direction p=0.007. Specifically, with the exception of Sine Pitch, increasing designs resulted in better performance than decreasing designs. Post hoc analysis comparing the 9 types of APBs represented in Figure 2 indicated that Whole Song APBs had better performance than Sine Duration Decreasing and Cello Duration Decreasing (all p's < 0.03), with the difference between Whole Song and Cello Pitch Decreasing approaching significance (p=0.059). No other comparisons were different.

3.2.2. Preference

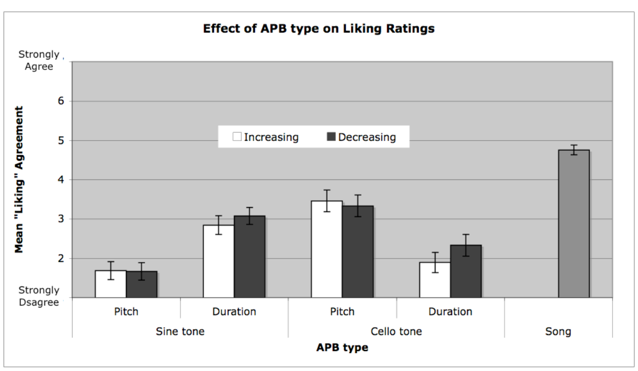

Figure 3. *Adjusted mean "Liking" agreement by APB. Increasing APBs are in white, Decreasing APBs are in dark grey, and the whole song APBs are in light grey. Error bars represent the standard error of the mean.*

Figure 3 shows the mean preference scores by APB. As seen in this figure, participants rated the whole song APBs more positively than the Sine or Cello tone APBs and this main effect was significant (p<0.001). Further, while there was no effect of direction on preference ratings, for the tone APBs (Sine and Cello) the interaction of dimension by APB type was significant (p<0.001). Finally, when examining Figure 3, three "groups" of preferences are apparent: Sine Pitch and Cello Duration; Sine Duration and Cello Pitch; and Whole Song. A Tukey's post hoc analysis indicates that these three groups are all significantly different from each other (all p's < 0.04) with the first group (Sine Pitch and Cello Duration) having the lowest ratings and the Whole Song APB having the highest ratings.

4. DISCUSSION

As compared to the Whole Song APB, the Tone stimuli (both Sine and Cello) used are different in many respects. Some of the most obvious features include:

- The whole song occupies all of the time of the required stimulus (in other words, there are no ISIs) leading us to question the role of ISI's in our results;
- The whole song represents a complete design structure with a beginning-middle-end leading us to question presuppositions in our design principles;
- The songs have repetitive rhythmic pulses that bind the fabric of the song together without pause;
- The songs are designed for the amount of time they fill - in other words, not necessarily extensible.
- The whole song shares the use of a tempered pitch system with the Pitch Cello stimuli as opposed to 100 hertz intervals used by the Pitch Sine stimuli. The tempered pitch system is not relevant to the Duration stimuli since in those only one pitch is used throughout.

The pitch stimuli used in the APBs are repeated transposed melodic fragments separated by ISI's. The length of the ISI's precludes any ongoing rhythmic patterns.

Despite these differences in design elements, given the results of the comparison of the Sine Tone APBs to the Whole Song APBs, we had anticipated that performances for the Cello Tone APBs would not differ from those of the Sine Tone APBs or the Whole Song APBs and this was the case. Indeed, with the inclusion of the Cello Tone APBs, the hint of an interaction between Direction and Dimension appears to not exist with the results strongly suggesting that APBs designed with an increasing element result in better performance than those with a decreasing element.

One of the main goals of this study was to determine the effect on subjective ratings of using elements of a more pleasing aesthetic when designing APBs. We had expected that using the timbre of a live instrument and a tempered pitch system would improve the subjective experience of listening to the APBs. However, the results suggest otherwise. Although, the Cello Duration APBs had ratings equally low as those for the Sine Pitch APBs, the Cello Pitch APBs had higher ratings than Sine Pitch APBs. It may be that a change from simple sine tones to natural timbre alone is not sufficient for a change in 'liking'.

The performance and preference results together are further proof that when designing stimuli like these, the designer needs to know that designs that result in better performance will not necessarily be more preferred. This is not a new finding, as the disconnect between preference and performance has been found in much of the human-computer interaction research as well as auditory and sonification studies [e.g., 9]. However, the results reported here go further to show that it requires more than "prettifying" the sounds to improve preference ratings. Indeed the finding that the Sine Duration tones were preferred significantly more than the Cello Duration tones refutes the assumption that all auditory displays should use instrument timbres [10].

With regard to the practical application of these results, this and previous studies [9] show that it is possible to create APBs that carry sufficient temporal information to allow listeners to be reasonably accurate in their estimation of time on hold. Furthermore, unlike the segmented tonal stimuli, the whole-song APBs were generally viewed as aesthetically pleasing. These finding indicated that it was indeed possible to construct APBs that have both pleasing aesthetics and good temporal information qualities.

Unfortunately, the use of whole-songs as APBs has significant difficulties if the solution is to be extended into field use. First, there must be different songs available for each anticipated hold duration. Second, finding songs for wait durations that are especially short or especially long is very difficult. Finally, the songs selected for the whole-song APB study were specifically picked by the demographic that would subsequently hear them during the hold experiment. While there are ways to maximize the probability of a listener obtaining the "right" music [2], it is not an optimal solution. The studies we have conducted have aimed to overcome these difficulties by utilizing segmented stimuli that are easily compressible and extensible, while adding a musical dimension to the elements to make the APB more song-like and, hopefully, more aesthetically pleasing.

In addition to practical applications, the results of this study have implications for more basic research on the effects of auditory display design on performance. The results from this experiment show that polarity has an effect on performance and thus suggests that performance associated with time estimation may be driven by polarity. Given the performance similarities between the best of the tone APBs and the Whole Song APBs, future studies in this line of research will focus on identifying what traits high performance tonal APBs and Whole Song APBs have in common.

In the most general sense, an APB is an auditory interface that provides temporal information to a user. This temporal information gives the user location information concerning where they are in that temporal process. The information may be as simple as an interface that signifies the completion of a temporal sequence (e.g. a cooking timer), or it may be as sophisticated as the interface we are describing here, one that is presented when a user is on telephone hold. Given an APB's potential for communicating more than retrospective time estimates, further studies in this line of research will measure different perceptual variables, e.g., time elapsed and projected time to completion.

5. ADDITIONAL FILES

Sound files referenced in this paper in brackets can be found at: http://coursesite.uhcl.edu/hsh/PeresSC/sonification .html

6. REFERENCES

[1] Knott, B. A., Pasquale, T., Joseph, K. M., Miller, J., and Mills, S., "Please hold for the next available agent": The effect of hold queue content on apparent hold duration. *Proceedings of the Human Factors and Ergonomics Society* Santa Monica, CA. Human Factors and Ergonomics Society. 2003, pp. 668-672.

[2] Knott, B.A., Kortum, P., Bushey, R., and Bias, R., The effect of music choice and announcement duration on subjective wait time of call center hold queues. *Proceedings of the Human Factors and Ergonomics Society* 48th Annual Meeting Santa Monica, CA. Human Factors and Ergonomics Society. 2004, pp. 740-744.

[3] Crease, M. & Brewster, S., Making Progress with Sounds. The Design Evaluation of an Audio Progress Bar, *Proceedings of the 5th Annual International Conference on Auditory Displays,* 1998.

[4] Polkosky, M. D. & Lewis, J. R., Effect of auditory waiting cues on time estimation in speech recognition telephony applications. *International Journal of Human-Computer Interaction*, 14, 2002, pp. 423-446.

[5] Kortum, P., Peres, S. C., Knott, B. & Bushey, B., The Effects of Auditory Progress Bars on Consumer's Estimation of Telephone Wait Time. *Proceedings of Human Factors and Ergonomics Society* Santa Monica, CA. Human Factors and Ergonomics Society. 2005, pp. 628-632.

[6] Kellaris, J.J. and Kent, R.J., The influence of music on consumers' temporal perceptions: Does time fly when you're having fun? *Journal of Consumer Psychology*, 1, 1992, pp. 365-376.

[7] Gueguen, N., & Jacob, C., The influence of music on temporal perceptions in an on-hold waiting situation. *Psychology of Music*, 30, 2002, pp. 210-214

[8] Kortum, P. and Peres, S.C., An exploration of the use of complete songs as auditory progress bars. *Proceedings of the 50th Annual Human Factors and Ergonomics Society*. Santa Monica, CA, Human Factors and Ergonomics Society, 2006, pp 2071-2075.

[9] Petrie, H. and Morley, S., "The use of non-speech sounds in non-visual interfaces to the MS-windows GUI for blind computer users," presented at International Conference on Auditory Displays, Glasgow, UK, 1998.

[10] Kramer, G., "Some organizing principles for representing data with sound," in Auditory display: Sonification, audification and auditory interfaces. Proceedings of the first international conference on auditory displays 1992, G. Kramer, Ed. Reading, MA: Addison-Wesley, 1994, pp. 185-222.

7. ACKNOWLEDGEMENTS

We would like to thank Aaron Alon for designing the stimuli and Jennifer Humphreys for her cello work. We also thank J. Barton, S. Conley, B. Nelson and A. Semmes for additional stimuli designs and Andy Su for data collection.