AUDITORY STIMULUS DESIGN: MUSICALLY INFORMED

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

[1]Rice University
[2]University of Houston-Clear Lake
Houston, TX

stallmann@rice.edu

ABSTRACT

This paper discusses an approach to auditory stimulus design that appropriates concepts and techniques commonly used in music composition. These ideas are used to create referential sound cues that orient a listener along a timeline. The functions these cues serve may include: emphasizing arrival at targeted goals; providing orientation information relative to beginning and end times; or creating a sense of imminent closure indicating a predictive end to an ongoing process.

The stimulus used in the pilot study represents this team's first attempt at integrating musical ideas with stimulus design. In an effort to make this stimulus comparable with previously used experimental stimuli, extreme restrictions were placed on the design. Although the resulting stimulus is not to be confused with 'music' in the proper sense of that term, it is interesting to note how an extremely restricted set of elements can be manipulated to create an aesthetically satisfactory experience that rivals responses to 'real' music in untrained listeners. The application of musical techniques towards the construction of effective auditory stimuli that are, at the same time, rated aesthetically satisfactory by users, is a long-term object of study by this team.

1. INTRODUCTION

The root of the work described here is found in the relationship between an auditory stimulus and a subject's ability to estimate time that has passed (retrospective time estimation). Initial experiments were conducted within the framework of a subject's ability to estimate time spent 'on-hold' while waiting on a telephone [1,2,3]. As the work has evolved, several questions have arisen that include: What effect does a pleasurable experience have on a subject's time estimate? Can an audio stimulus be encoded to provide cues to aid the subject in this task? If so, what are the most effective cues that can be included? What is the difference between retrospective estimation and prospective estimation? Does the change from retrospective to prospective time estimation require a change in stimulus design? What happens to the perceived duration of time when elements of sound are coordinated to produce an effect of 'closure' at the end of the stimulus? Although some preliminary work has been done in this area [4], much is yet to be studied.

Earlier work done by companies interested in providing the best possible experience for their customers were not interested in accuracy of time estimation. Rather, they believed that if a customer 'underestimated' their 'on-hold' time, this would indicate a more satisfying experience. To encourage a positive response from callers (customers), their goal was to make the 'on-hold' experience as pleasurable as possible. To this end, in their experiments, these companies provided menus that gave customers the ability to choose a genre of music to fill the time as they waited 'on-hold.' It turns out that there is some experimental evidence to suggest that this procedure fosters an environment conducive to fairly accurate time estimates of time 'on-hold' [5]. This raises the question of whether or not any further work need be done with musical stimulus design. Is this technique already adequate enough? By empowering subjects with predictable choices (thereby generating the sequence of heightened attention, expectation, and subsequent fulfillment), the optimal conditions for retrospective time estimation may already be present. If we accept that, then the primary function of the auditory content is to provide a predictable, pleasurable experience that satisfies the expectation generated by choice.

If, on the other hand, one believes that a subject's performance can be manipulated by using specific cues embedded in an auditory stimulus design, a serious problem with the previous method becomes apparent. How can the precise relationship of the auditory stimulus to the subject's performance be studied if the only attribute given to that auditory stimulus is its genre? All that could be measured would be the relative satisfaction of the overall experience in relation to accuracy in the performance of estimating time passed.

With the aim of more objectively defining the design of the auditory stimulus itself to discover links between performance and specific variables of stimulus design, Kortum, Peres, Knott, and Bushey [1] prepared auditory stimuli based on simple, measurable quantities. These stimuli were created using laboratory sinusoids as sound sources. The stimulus designs included the dimensions of pitch and duration across increasing and decreasing conditions. The quantities used were precisely measured. The simplicity and clarity of the auditory stimuli design allowed for precise study of relationships between changing conditions and a subject's performance. One interesting piece of information observed in this experiment was that the increasing condition in the dimension of duration resulted in generally better performances than other possible combinations, although these differences were not significant. A problem with the stimulus design, however, was that all subjects of the experiment rated the overall aesthetic experience of the stimulus as very unsatisfactory. The researchers attributed this to the 'laboratory-like', mechanical quality of the sounds.

With the goal of matching the best performance results found in the sinusoid study while also increasing overall satisfaction, Kortum and Peres conducted another experiment [2]. This time, complete popular songs (as identified using music industry data reported in Billboard Magazine) were chosen with specific durations similar to the 30, 60, 120, and 240 second stimuli used in the earlier sine tone experiment. This differed from earlier experiments that used music as a stimulus because in this case a complete song that filled the duration of the stimulus was used, whereas in earlier experiments the songs were simply cut off midstream at the end of the required duration. Another difference is that in this experiment, the songs were assigned rather than chosen. Results of the 'whole song' experiment showed that as auditory stimuli, the 'whole song' stimuli matched the best performances found in the sine tone experiment and also increased the overall satisfaction of the experience. It did not, however, attempt to answer what aspects of these songs could be attributed to the matched performance results and improved aesthetic ratings.

Noting the higher satisfaction ratings associated with musical stimuli while also understanding that using completed songs as stimuli did not provide the necessary precision to explain performance differences, Peres, Kortum and Stallmann turned to music composition to find ways to improve the aesthetic appeal of their stimulus designs while also attempting to maintain a strict degree of experimental control over the stimulus as noted in the earlier sine tone experiment. In an attempt to answer how the timbre and 'laboratory-like' quality of the sound affected aesthetic ratings by subjects, a small group of composers guided by Stallmann derived closely related stimuli from the template design of the original sine tone stimulus. A stimulus designed by composer A. Alon was chosen from the pool of new stimuli and used in a follow-up experiment. It was decided that a human performer would play the stimulus on an acoustic instrument, in this case a cello. Adjustments were made to adapt the stimulus to human performance including the use of common scales, a standard tuning system, duration changes that reflected common notation practice, and an embellished musical fragment employed in an attempt to make the original pitch stimulus slightly more music-like. Although these changes were necessary, they were considered as slight enough to allow for a comparison between the cello and original sinusoid stimuli. On the other hand, these changes brought the pitch dimension into alignment with cultural norms. In the original sinusoid experiment, increasing and decreasing pitch were tracked to 100 Hz intervals. Given the linear nature of the changes over the entire frequency range of the stimulus, the resulting musical intervals were not constant, but rather were frequency dependent. In this experiment with cello tones, changes made to the pitched stimuli resulted in comparatively better performance in the increasing pitch condition, however satisfaction overall was still low. These results confirmed that making a stimulus more attractive is not simply a matter of changing the timbre and tuning system. The actual design of the stimulus itself bears the brunt of its aesthetic appeal.

The next logical research step required a merging of desirable musical characteristics of 'whole song' stimuli with the strictness and simplicity of the original electronic stimuli. This required an overhaul in the overall stimulus design and brings us to our current study. In the current pilot study, the stimulus has been redesigned by Stallmann to replicate some aspects of the whole song, while also severely limiting the number of compositional elements to make the resulting stimulus comparable to earlier designs. Characteristics adopted from the 'whole song' stimuli include: 1) elimination of the extremely long silences (inter-stimulus intervals) that occurred between restatements of the auditory stimulus; 2) creating a sense of completion that coincided with the end of the allotted time by including a clear goal that creates a sense of closure at the end of the stimulus duration; and 3) using metrically organized pulses to create a reference tempo against which patterns can be discerned. These 'whole song' characteristics were merged with features from the original 'sinusoid' stimulus including: 1) an extensible design—one basic design template used for all time lengths with adjustments made to accommodate different durations; and 2) a limited number of discrete parts which make up the design, each part having a specific function in the whole.

2. STIMULUS DESIGN

The four 'electronic' stimuli for this experiment (30s, 60s, 120s, and 240s) are all based on an extensible design structure presented in its optimal form at 30s. This 'Basic Form' is stretched in time to accommodate the longer durations. As the same materials get stretched over longer periods of time, the stimuli for the longer durations become more static. In an attempt to compensate for the perceived reduction in complexity as durations increase, stimuli at 60s, 120s, and 240s are slightly altered from the Basic Form presented at 30s. These alterations are minor (with the exception of Part 5) and do not affect the function of Parts 1-3 presented in the Basic Form. Below is a description of the design of the Basic Form. Alterations to this basic design in the 60s, 120s, and 240s stimuli are described afterwards. Examples of all of these can be heard online (see section 6 for more information).

It was determined, at the beginning of this study, that the design of the stimulus would: 1) take into account information gathered from previous experiments; 2) be extensible in time; 3) be conducive to 'multi-tasking' and require relatively low cognitive functioning; 4) be generally 'likable'; and 5) be useful for retrospective time estimation. As a final condition, the stimulus should be very simple to construct using commonly available tools.

Earlier experiments showed that when using pitch and duration, the increasing dimensions resulted in better retrospective time estimation performance than decreasing. Earlier stimuli also used inter-stimulus intervals. It was believed, after reviewing results from 'whole song' stimuli, that the long silences introduced by the ISI's were a contributing factor to the negative 'likeability' ratings of the previous stimuli. As a result, the design for the stimulus includes a sequence of rising pitches not separated by silences.

Other changes from earlier stimuli include: the use of a musical meter (4/4) to facilitate pattern recognition; the inclusion of a constantly reinforced goal tone so that there is a convergence in pitch signaling a clear end to the stimulus; and the use of other reference cues described below. The stimulus was created using a very common sound design tool, Logic Audio, Version 7. All of the timbres used in the stimuli use 'stock' patches in virtual instruments found in Logic 7 (descriptions of timbres used for these parts is at the end of the paper in Section 6).

2.1. 30 Second Stimulus (Basic Form)

The 'Basic Form' of the 'electronic' stimulus consists of three parts (see Figure 1).

Part 1 of the stimulus steps through a one octave scale using nine notes: C, D, E flat, F, G, A, B flat, B natural, C. The scale operates in a 'C' tonality, meaning that the tones are not weighted equivalently, but rather are coordinated to reinforce the presence of the tone ‘C’. Part 1 both begins and ends on C. The timing of the last three notes (B flat, B, and C) is compressed with their succession occurring twice as fast as the previous notes. The time compression at the end of the line in itself serves as a strong cue to indicate to the user that the end of the stimulus is near while also emphasizing the leading tone tendency to create a heightened expectation of the return of C at the end. The amplitude, or volume, of Part 1 increases from medium loud to loud, also communicating to the user that he is getting closer to the end. Transitions between successive notes in the scale are overlapped to convey a feeling of connectedness while moving forward in time.

Part 2 articulates a repetitive, rhythmic pattern that stays constant throughout the entire stimulus. The repetitive nature of this part reinforces the basic metrical organization and offers a tempo, or ‘pulse’ rate as a reference (120 quarter notes per minute).

Part 3 serves as a pitched reference point in the stimulus. This reference point plays a repeating rhythmic pattern on C that is two octaves higher than the first note of Part 1. When Part 1 reaches its last note, C, it and the reference point are one octave apart, and the resulting perfect consonance signifies the end of the stimulus followed by one conclusive unison event that confirms arrival. The duration of each reference note is kept constant except for this last event, which is slightly longer as a cue for the end of the stimulus.

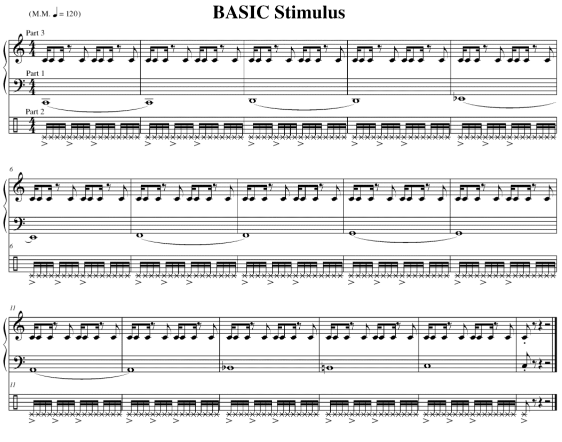

Figure 1: Musical Notation for ‘Basic Form’ of the ‘electronic’ stimulus

2.2. 60 Second Stimulus

In the 60 second stimulus, Parts 1-3 of the 30 second stimulus simply double their overall durations. The underlying tempo remains the same (120 quarter notes per minute). An addition to the 60-second stimulus is Part 4, a subtle marimba rhythm that enters at the halfway point (at 30 seconds) and again emphasizes ‘C’. This part elaborates the basic metrical subdivisions while also splitting the entire duration of the stimulus into two halves. Part 4 consistently enters in the second half of the 60, 120, and 240-second stimuli.

2.3. 120 Second Stimulus

In the 120-second stimulus, Parts 1-4 of the 60-second stimulus again double their overall duration while maintaining the underlying pulse (120 quarter notes per minute). The structure of the 120 second stimulus is very similar to the 30 and 60 second stimuli. Because of the increase in the overall time of the stimulus, three additional changes have been made to provide yet more location information for the listener and to make the stimuli aesthetically viable: a decrease in the amplitude of each individual note event of Part 1 to produce a dynamic, time varying envelope; subtle variations introduced in Part 3 to reinforce motion along the scale; the addition of Part 5, a time-compressed scale as an additional reference point.

Part 5, a time-compressed variation of Part 1, coincides with the entrance of each pitch of the rising scale in Part 1. As the scale ascends, the starting note of Part 5 ascends in direct relation. Part 5 serves three purposes. First, it provides an additional flourish at the beginning of each scale tone giving each step of the rising scale more emphasis. Secondly, because the first tone of each of the scales is rising along with Part 1, the time-compressed scale becomes shorter as the listener progresses through the stimulus. This feature provides the listener with auditory cues about the current position in the stimulus timeline. Thirdly, since the reference note is played at the end of each compressed scale, the final note, or “goal,” of the stimulus is constantly reinforced.

2.4. 240 Second Stimulus

In the 240 second stimulus, Parts 1-5 of the 120 second stimulus double their overall duration for the last time while the underlying tempo remains the same (120 quarter notes per minute). The structure of the 240-second stimulus is very similar to the 120-second stimulus. The only difference is the addition of Part 6. Because of the increase in time from 120 to 240 seconds, a conga sample was added to the second half of the duration of each scale tone in order. Since each scale tone of Part 1 is now 30 seconds long and they decay to silence (similar to the 120s stimulus), the conga serves to mask the decay while emphasizing the flow of time through each section.

3. A PILOT STUDY

In order to test these new ‘electronic’ stimuli with real users in a real task, a small pilot study was conducted using an ‘on-hold’ telephone experimental protocol. 14 people participated in the study. Each participant listened to each of the four electronic stimuli (30s, 60s, 120s, 240s) and, following the completion of the stimuli; they made a retrospective time estimate of the length of time they were ‘on-hold’. Results of this pilot study indicate that the performance of the electronic stimuli (the accuracy with which users estimate the amount of time they have been on hold) is as good as the other APB stimuli that we have tested to date. As seen in Figure 2, the best performance (indicated by the lowest proportion of the estimate error) occurred in the Cello tones group. However, this difference was not significant ($F(3,219) = 1.735$, $p = .160$) indicating that there are no performance differences for the different APBs. Further, the preference ratings given to the electronic stimuli equal those given to the ‘whole song’ stimuli and are significantly better than other periodic stimuli that we have constructed in the past. Figure 3 shows the results of the comparison of participants’ preferences across the four APB types. It is clear from this figure that the type of APB influences participants' preferences (F (3,215) = 34.01, p < .0001). Specifically, they preferred the Electronic and Song APBs to the Cello and Sine Tone APBs. Further, Figure 3 also indicates that participants liked the Sine Tone APBs the least. Tukey's post hoc analysis showed significant differences between all of the APBs but the Electronic and Whole Song.

This pilot study suggests that constructed stimuli can bridge the gap between satisfaction and accuracy while still maintaining restraints over the stimulus design. While 'whole song' stimuli demonstrate good accuracy and high preference ratings, they suffer from the fact that they are not extensible. For many applications that could potentially use auditory progress bars, the ability to have stimuli that can be constructed with specific time lengths is very important.

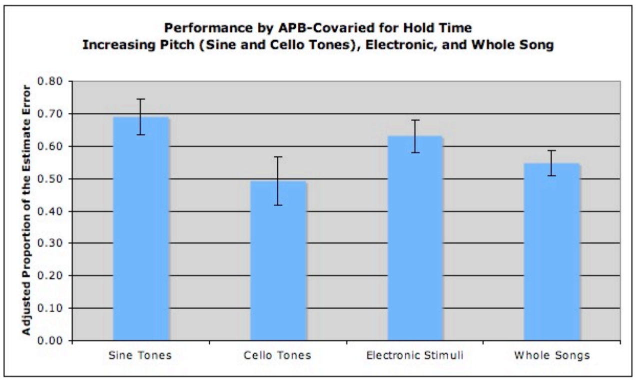

Figure 2. Mean adjusted proportion of the absolute value of the error participants made when estimating their time on hold by the type of APB. Error bars represent standard errors for each group mean. Means were co-varied for the hold time (or length of the APB).

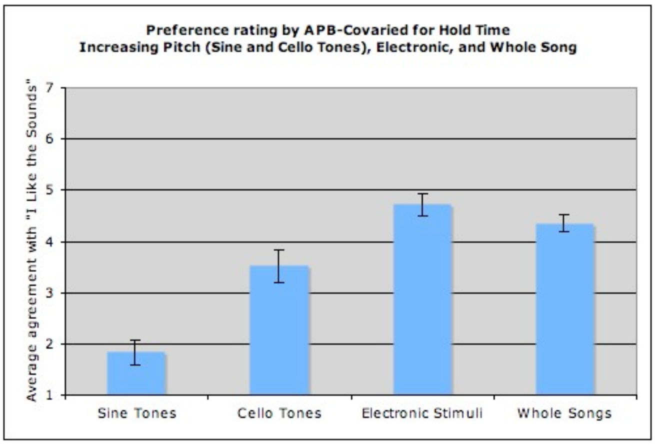

Figure 3. Mean preference ratings by the type of APB. Error bars represent standard errors. Means were co-varied for the hold time (or length of the APB). Scale used was 1-Strongly Disagree to 7-Strongly Agree.

4. DISCUSSION

One of the things we have been considering is the way in which auditory progress bars may actually be used in other environments. To date we have had participants estimate the amount of time they've been on hold after the auditory progress bar has completed. While this is certainly a valid measure of time estimation, we believe that one of the important ways people use progress bars is the estimation of the amount of time that remains in an ongoing process. This prospective time estimation may demand that different compositional elements be considered and used in the construction of these APBs.

As we conduct more studies, we hope to identify and categorize types of cues that are most effective for the time estimation tasks. For example, the current 'electronic' stimulus aims towards its end by reaching towards a goal. The use of an end goal at the conclusion of the stimulus may, in fact, be better suited for prospective rather than retrospective time estimation given that the design emphasizes forward anticipation of completion, not backward assessment of growth relative to a start position. The lack of superior performance results with the current stimulus may be due to a mismatch between the type of cue (anticipation of a final goal indicating closure) and the task of the subject (retrospective time estimation).

In contrast, it may be that by using cues that more explicitly emphasize cumulative growth from a starting position work better for retrospective time estimation. The notion of 'accumulation' in the mind of the subject lends itself well to retrospective time estimation. It suggests a rate of growth as measured against time thereby putting the function of passing time in relief.

We draw a distinction here between types of auditory stimuli that use 'musical' sounds from other types of stimuli that use sounds with higher symbolic value. For example, we would like to test a stimulus that uses a clock ticking with chime rings at periodic intervals throughout the stimulus and compare performance results with our 'musically informed' stimuli.

We also draw a distinction between a subject's performance based on a single hearing of the stimulus versus performances based on a learned stimulus. For example, Part 4 of the stimulus consistently enters the 60s, 120s, and 240s stimuli during the second half of the overall duration. If a user was taught to recognize that fact, it could serve as a reliable reference cue for orientation. We note that one of our group's interests is to design stimuli to be used in interactive interfaces that require orientation cues about where one is located in an ongoing process. The embedded cues may go unnoticed in a first hearing, but the subject could be taught to hear these cues, thereby improving orientation and perhaps performance.

One of the great challenges we have encountered in the design of the stimuli is how to develop a palette of techniques to create effective extensible designs. Stimuli optimized to be heard in 30 seconds cannot be easily expanded into 240 seconds and remain effective. In our first attempt here, we have compensated for the longer durations by increasing the amount of surface activity, thereby complicating the surface texture. While this technique can be useful, especially for embedding location cues, we also plan to study other techniques, perhaps using traditional contrapuntal methods of augmentation and diminution, or repetitive figuration.

5. CONCLUSION

We are interested in designing APBs that help users be accurate in their time estimates. To date, extensible APBs have been judged harshly in terms of their satisfaction. 'Whole Song' APB's have demonstrated good satisfaction and excellent time estimation accuracy, but they are not extensible. This Pilot Study suggests that these constructed stimuli can bridge the gap between satisfaction and accuracy.

In constructing these new electronic stimulus APB's, we combined three elements identified in the whole song stimuli (elimination of ISI's, creating sense of completion, creating a reference tempo). While this combination seems to have been effective, the relative contributions of each of the elements cannot be discerned from these data. Future research should examine each element in isolation in order to allow a more accurate picture of the relative contributions that each of these elements may or may not make to the combined stimulus. A more granular understanding of these elements would contribute to a development of design patterns using highly quantifiable data.

The results of pilot testing suggest that, while there has yet to be a measurable difference in performance of our subjects, there was a marked increase in the 'likeability' of the basic stimulus. The stimulus is now quite tightly controlled and exhibits the same degree of likeability as the whole songs. The lack of superior performance may be due to the fact that since the effort was to create a final goal of closure, we may have created a stimulus better suited for prospective time estimation than retrospective time estimation. We hope, in future work, to test this stimulus and others in prospective time estimation experiments.

Finally, we believe it is important to note that these stimulus designs should not be confused with music. We are simply using elementary techniques of musical control over the basic design of the stimulus so that effects between stimulus and performance can be studied.

6. ADDITIONAL FILES AND TIMBRE LIST

Additional sound files accompanying this paper may be downloaded online at: http://www.trigonmusic.com/icad.html

All Sound files are monophonic in MP3 format.
'Basic Form' Part 1
'Basic Form' Part 2
'Basic Form' Part 3
Electronic Stimuli 30 seconds
Electronic Stimuli 60 seconds
Electronic Stimuli 120 seconds
Electronic Stimuli 240 seconds

Timbre List:
All timbres used in the stimuli were selected from readily available common patches found in LogicPro 7.2.3.
Part 1: Dyno Rhodes Pad patch (ES2 virtual instrument)
Part 2: audio loop (Synthetic Chime Loop)
Part 3: default patch (EVD6 virtual instrument)
Part 4: marimba sample (EXS 24 software sampler)
Part 5: default patch (ES2 virtual instrument)
Part 6: audio loop (Conga Groove 04)

7. REFERENCES

[1] Kortum, P., Peres, S.C., Knott, B. & Bushey, B. (2005). The Effects of Auditory Progress Bars on Consumer's Estimation of Telephone Wait Time. Proceedings of the 49th Annual Human Factors and Ergonomics Society. Pp 628-632, Santa Monica, CA, Human Factors and Ergonomics Society.

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

[3] Peres, S.C., Kortum, P., and Stallmann, K. (2007). Auditory Progress Bars: Preference, Performance and Aesthetics. Proceedings of the International Community for Auditory Display, ICAD 2007.

[4] North, A. and Hargreaves, D. (1999). Can Music Move People? The Effects of Musical Complexity and Silence on Waiting Time. Environment and Behavior. Pp. 136-149, Sage Publications, Inc.

[5] Knott, B.A., Kortum, P., Bushey, R. and Bias, R. (2004) The Effect of Music Choice and Announcement Duration on Subjective Wait Time for Call Center Hold Queues. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, pp 740-744, Santa Monica, CA, Human Factors and Ergonomics Society.

8. ACKNOWLEDGEMENTS

We would like to thank Alicia Ling for her assistance in preparation of the stimuli, the stimuli descriptions and data collection.