Abbrevicons: Efficient Feedback for Audio Interfaces

Matthew Hockenberry, Sharon Cohen, Zachary Ozer, Tiffany Chen,
and Ted Selker

The Media Laboratory, 20 Ames St. E15-322, Cambridge MA 02139
{hock, selker}@media.mit.edu
{sbcohen, zozer, chent}@mit.edu

Abstract. Abbrevicons are a technique for encoding audio information efficiently while still providing the user with useful and appropriate feedback. Ideally, abbrevicons are targeted towards audio interfaces that are infrequently used and where the user doesn't have the opportunity to develop a strong model of the system or make sense of rich encoding schemes. Abbrevicons allow us to present natural language feedback that is identifiable and comprehensible, while taking significantly less time than traditional verbal feedback. Abbrevicons can decrease the speed of short descriptive feedback by up to 200% while still ensuring user comprehension. These techniques seek to provide richer user feedback in audio interfaces while still keeping in mind the need for efficient user interactions.

1 Introduction

In audio interfaces such as telephone trees, tones are often used to provide confirmation of user actions. These systems are inefficient because these tones provide little feedback or information to the user's specified action – they only indicate that an action has occurred. [1]

Abbrevicons are intended to replace these generic tones in audio interfaces. Abbrevicons are created by compressing a base word or phrase so that it plays significantly faster than normal speed. Abbrevicons can provide significantly more information to the user than a generic tone and has the potential to significantly improve the efficiency and quality of feedback in audio interfaces. A generic tone provides only awareness that the user has made a selection from the menu; it does not confirm their choice. Our goal in creating abbrevicons is to provide the user with as much information as possible within a very compressed space.

1.1 Existing Approaches

There is a large body of work in the literature on providing feedback mechanisms for audio interfaces. Audio icons and earcons both represent established methods for providing feedback based on user interactions, but neither offers feedback that is *both* innately meaningful and can be mechanically synthesized for occasional-use audio interfaces.

Audio icons are rich naturalistic sounds that encode underlying system mechanics in sound. The classic example of this is the ARKOLA simulation where Gaver used the sounds of a bottling factory to offer feedback on the state of the system. [2]

Generally, audio icons are well received by users and can communicate a lot of information. The difficulty comes in their construction; there has been little success in creating effective audio icons that can be generated through purely mechanical methods. [3]

Earcons are generated synthetic tones that can vary in rhythm, pitch, timbre, register, and dynamics. Earcons can use these five factors to encode different kinds of information such as the type of application, family, or number of a file on a desktop. Although they can be very flexible and computer generated, they require that the user learn how each of the factors encodes or alters the encoding of information. [4] This is quite unlike audio icons that have an obvious meaning that relates directly to the encoded information. Still once the system has been learned they offer a lot of opportunity for improved efficiency in presentation speed. [5]

1.2   Creating Abbrevicons

Abbrevicons are created by starting with a 'base phrase' recorded either directly from a human speaker or (ideally) through a speech generation tool. This base phrase is then processed to drop frames of audio content. The result is an abbrevicon which shares the same tonal properties of the base phrase, but which sounds as though it was spoken very quickly. We chose to drop frames so that we could keep the natural sound of the base phrase without distortion (such as changes to pitch, etc…). This process is relatively simple and can be performed with a large number of audio editing programs (such as soundforge, audacity, and so on).

2   Evaluation

We have run a series of studies in our exploration of the effectiveness of abbrevicons. We have had over 100 subjects complete 100 trials for each of three conditions studying both recognition and recall of abbrevicons. Our recognition conditions presented the subject with an abbrevicon and a set of either five visual or five auditory cues. We also did a flat recall that required the subjects record what they believed they heard. The abbrevicons were constructed from a fixed, known lexicon with which the subjects were familiar. We found that the comprehensibility of abbrevicons depends on four main factors: the amount of compression of the base phrase, the number of syllables in the base words, the presence of distinct spaces between words, and the consonant sounds within the base words. For our purposes a "comprehendible" abbrevicon is one for which the subjects were able to identify its meaning in more than 95% of trials.

One of the largest influences on the level of comprehension of abbrevicons is the amount that the base phrase is compressed. The difficulty of comprehension increased as the base phrases were increasingly compressed. Speeding the base phrase up by 75-150% tended to create reasonably understandable abbrevicons (the mean error rate is .866% for a 75% increase and 1.088% for a 150% increase with 99.5% confidence), while increases of speed by 200-250% rendered abbrevicons that were significantly less comprehensible (though still acceptable, the error rate for 225% was 3.053% with 99.5% confidence). Tripling the speed of the abbrevicons was the maximum limit of

any comprehension and we see the error rate rise exponentially at this point. Subjects also reported feeling uncomfortable listening to audio at this rate.

Increasing the number of syllables within a word increases the difficulty of comprehension. Abbrevicons that consisted of single-syllable base words were generally easier to comprehend, even at higher speeds, than abbrevicons of multiple syllable base words. Abbrevicons that contained two or three-syllable base words were not only harder to understand, but also had less potential to be compressed further and retain comprehensibility. For example, of the abbrevicons "confirm selection" and "confirm choice," the latter abbrevicon demonstrated higher clarity at the same compression rate of 100%. We attributed this easier understanding to "choice" having fewer syllables than "selection."

To make Abbrevicons more comprehensible at higher speeds, we found that intentionally adding spaces between words allowed for greater clarity, even at higher speeds. This was especially effective when using computer-synthesized voices, rather than human recordings of abbrevicons. For example, the abbrevicon "please hold" could be heard with much more clarity if a deliberate pause was added in between the two words. With the added pause, the Abbrevicon could be easily comprehensible even at 150% or 200% compression (with an error rate of 1.088% as reported earlier), whereas without the space, the Abbrevicon was only readily understandable at 75-100% increase in speed (with an error rate of 2.097%).
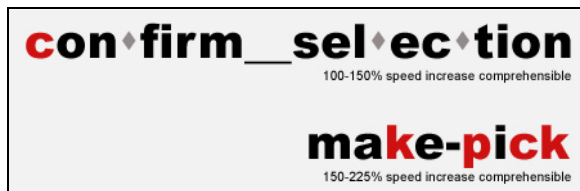

Fig. 1. 'Confirm selection', because of its lack of harsh consonants and multiple syllables can only be increased by 100-150%, while a base phrase like 'make pick' can achieve an increase of 150-225% and remain comprehensible

Another interesting conclusion about abbrevicons is that words with hard consonant sounds were easier to comprehend. In the previously discussed abbrevicon "confirm choice," the beginning "k-" and middle "-ch-" sounds are both weak consonant sounds. Abbrevicons with these weak consonant sounds tend to be readily understandable at 100-150% increase in speed. However, words with hard consonant sounds tended to be greater in clarity. "Make pick" consists of the "m" and "k" harsh consonants for the first word as well as the "p" and "-ick" harsh consonant sounds for the second word. The two words combined with a deliberate space in between, this abbrevicon had high clarity even after being speed up by 150-200%. Once again, computer-synthesized voices, which emphasized the harsh consonant sounds, were easier to comprehend at higher speeds than human recordings.

3   Conclusions and Future Work

From our studies, we concluded that the most easily comprehensible abbrevicons consisted of words with harsh consonant sounds and few syllables. In addition, a clear

space between words increases clarity even at high compression rates. Using these methods, we can compress information to up to 360 words per minute, twice the speed of normal conversational, without significantly hindering comprehension. Unlike other methods for providing feedback in audio interfaces, abbrevicons have both intrinsic meaning (assuming the listener can understand the language) and the capacity for computer generation (by scripting sound applications with text-to-speech engines). This could allow us to use abbrevicons as a method for providing feedback in general purpose audio interfaces with little effort on the part of the interface or system developer. The goal is that we can use these techniques to generate feedback that is useful to the user while still presenting information as efficiently as possible.

References

1. Resnick, P., and Virzi, RA. Relief from the Audio Interface Blues: Expanding the Spectrum of Menu, List, and Form Styles. ACM Transactions on Computer-Human Interaction, 2, 2 (1995), 145-176.
2. Gaver, W. W., & Smith, R. B., O'Shea, T. Effective sounds in complex systems: the ARKOLA simulation. In Proc. SIGCHI conference on Human factors in computing systems: Reaching through technology. (1991), p. 85-90.
3. Gaver, W. W. Synthesizing Auditory Icons. In Proc. INTERCHI. (1993).
4. Brewster, S. A., Wright, P.C. & Edwards, A.D.N. A detailed investigation into the effectiveness of earcons. In G. Karmer (ed.), Auditory display, sonification, audification and auditory interfaces. In Proc. The First International Conference on Auditory Display. (1992) pp. 471-498.
5. Brewster, S. A., Wright, P.C. & Edwards, A.D.N. Parallel Earcons: Reducing the length of Audio Messages. International Journal of Human-Computer Studies, (1995).