# Design and Evaluation of an Auditory Glance at Algebra for Blind Readers

R. D. Stevens, S. A. Brewster, P. C. Wright, and A. D. N. Edwards
Department of Computer Science
The University of York
York, UK YO1 5DD
Email: robert@minster.york.ac.uk

**Abstract**

An overview or glance is important in planning the reading process. Such a high-level view is not possible for a blind person when reading by listening. This paper describes the sonification of algebra notation to provide an *audio glance* to facilitate planning prior to reading. Prosodic cues used in spoken algebra were combined with pre-existing guidelines for earcon design to produce a high-level view of expression structure called *algebra earcons*. The initial evaluation indicated listeners could recover syntactic information from an algebra earcon and pick an appropriate expression. Some defficiencies were revealed, that were amended and then retested in a second experiment showing an improvement in recognition of syntactic features. Analysis of verbal protocols revealed the level of syntactic information the listeners could retain and recall, and also the strategies used.

## 1 Introduction

We normally associate reading with vision, but for blind people reading often has to be done through listening. The listening reader plays a passive role in the reading process, and cannot use the external memory provided by the printed page to achieve the speed and control inherent in the visual reading process. This paper reports on part of the Mathtalk program, which uses synthetic speech (including prosody), nonspeech audio and browsing techniques so blind people can read algebra in an active way [6]. One of the major difficulties a listener has is the inability to plan the reading process. At present, an algebra expression has to be read in full detail or not at all. Current speech based technology, such as tape recorders and screenreaders for computers, have a poor representation of the material and lack adequate control over the information flow. This means a syntactically complex expression may overwhelm the reader's resources and lead to frustrating rereading to gain the structure and detail of the expression.

The spatial arrangement of printed algebra notation means a sighted reader can quickly judge the syntactic nature of an expression, even at the basic level of length and presence of complex items. This *glance* or preview allows a sighted reader to *control* the pace of the reading process. Ernest [4] proposes planning and decision making as the first stage in the process of reading an expression. Such a glance or overview is not really possible for the listener, because of the transience of the speech signal and lack of control over the information flow. Such a glance should present a high-level view of the expression to be read. From this view the reader should be able to judge the syntactic complexity of the expression and so be able to plan a reading strategy, using the browsing functions provided in Mathtalk. If possible this glance should also give the reader a framework into which the details of the expression may be fitted as it is read.

Speech was rejected as a means of providing this glance because a description would be longer than the expression itself. One of the principles of the Mathtalk system is that the interface should perform no mathematical interpretation, as is the case with a printed expression. This rules out a high-level mathematical description as a glance. Instead nonspeech sound was explored as an option for an audio glance. Nonspeech sounds have the potential for communicating complex messages to a listener in a form that does not interfere with speech. This paper describes a novel method for using earcons to sonify algebra notation. An abstract presentation of an algebra expression could give noninterpretive information about that expression and hide its content, both necessary components of a glance.

Earcons were used to provide this glance. These are abstract nonspeech sounds whose structure provides information to the listener [1, 2]. The parameters defining earcons are rhythm, pitch, timbre and dynamics, the same parameters used to describe prosody [5]. Mathtalk uses prosody in its synthetic speech output to indicate the syntactic structure of an expression to the listener [7]. Algebra syntax, algebraic prosody and guidelines for earcon construction [2] have been combined to make earcons whose structure represents the syntax of an algebra expression. These earcons are called *algebra earcons.*

Algebra earcons work by representing only the syntactic type and not the instance of an item in an expression. Different musical timbres were used to represent the basic syntactic types within an expression. The rhythm, pitch, duration and timing characteristics of the spoken form were used to guide the creation of the algebra earcon.

Two experiments were conducted to test the ability of listeners to recover syntactic structure from an audio glance. The first investigated whether the basic principle of algebra earcons would work and tested the rules for their construction. The second experiment tested the amended rules and investigated the mental representation derived from the audio glance.

## 2 Experiment One

A multiple choice paradigm was used with two conditions. The first used only syntactically simple expressions, the second syntactically complex ones. Construction of suitable distractor expressions in the multiple choice responses would enable testing of the effectiveness of each aspect of the earcons. The recognition task inherent in this paradigm was not realistic of the ultimate planning task, but would provide an initial indication of the effectiveness of algebra earcons. If a listener could recover sufficient information to select an expression from a series of similar alternatives, then the audio glance would be said to be effective in presenting syntactic information.

A total of 12 subjects were used. All had mathematics qualifications, but varied in daily use of mathematics from postgraduate research work to virtually none. All were familiar with standard algebra notation.

A total of 30 expressions were made, equally divided between syntactically simple and complex. The simple expressions had no complex items, but could have many simple ones. The complex expressions always had at least one complex item, but could also include simple items. Two stimuli are shown in Figure 3.

The multiple choice format comprised the correct answer plus three distractors. Each distracter differed from the stimulus in one respect that would probe the effectiveness of the earcon to present the syntax of the expression. Examples of stimuli and distractors may be seen below in Section 3.3.

| Item | Timbre |
|---|---|
| Base-level operands | Acoustic piano |
| Binary operators | Silence |
| Relational operators | Marimba |
| Superscripts | Violin |
| Fractions | Pan pipes |
| Subexpressions | Cello |

Table 1: Table of musical timbres used in algebra earcons.

## 2.1 Algebra Earcon Design

Algebra earcons are constructed by blending the visual representation of algebra syntax with the prosodic cues used when it is spoken. Different items within an algebra expression are replaced with sounds with different musical timbres, enabling a listener to discriminate elements within the expression without knowing the instance. The sounds used are shown in Table 1. The timing, pitch and amplitude of these sounds are then manipulated according to the rules below.

To establish a rhythm that facilitates recovery of grouping and retention of information, a bar length must be defined for the earcon. This is based on the length of the longest term in the expression. For simple terms, each item contributes one beat to the bar length. The last operand in a term contributes two beats. This lengthening mimics the final syllable lengthening in speech. An extra silent beat is added for a printed binary operator. A silent beat is used because for a glance the division into terms is the important feature, not the nature of the operator. In contrast, relational operators are important cues to the structure of an expression. In length calculations, a relational operator is included in the following term, being counted as one beat, plus a separator of a silent beat.

All complex items are represented by a continuous tone with a constant pitch, similar to parenthesised expressions in speech. This simply indicates that such an item is present, but revealed nothing of its contents, except its length. This is consistent with the idea that an algebra earcon is a glance. The lengths of complex items were calculated as above, but binary operators did not make a contribution. This reflected the faster, pausless uttering of these items in speech.

After the maximum term length had been calculated, each term in the expression is fitted into this bar length for the algebra earcon. Shorter terms are padded at the right with silent beats to preserve the rhythm of the algebra earcon.

Algebra earcons are played in the C major scale. The pitch of each new term starts at middle C ($C_3$). Subsequent items are played at one note below the previous. The last term's pitch starts at $A_4$. This mimics the sharp pitch fall at the end of an algebraic utterance, that indicates the impending end of the expression to the listener. If the relational operator precedes the final term, the note representing the first operand is played at $F_4$, as the relational operators are also played at $A_4$. Superscripts are played at a pitch two notes higher than their base, in the octave above. Subexpressions are played two octaves and two notes below the preceeding item.

Simple and complex fractions are both represented by pan pipes, but with a different pitch profile. Simple fractions have the same pitch fall throughout as simple terms, but there is a one octave drop at the start of the denominator. The last note in a simple fraction is lengthened as if it was the last note in a term. Complex fractions are represented by two long notes of constant pitch, separated by two silent beats. The second note, the denominator, was played two notes lower than the first. This difference reflects that seen in speech, where a complex fraction is spoken as two bracketed subexpressions, separated by an 'over' operator.

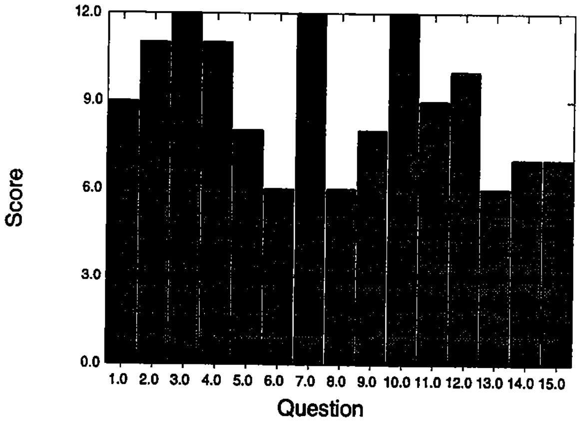

Figure 1: Scores across questions for the simple stimuli expressions.

For example the expression $3x + 4 = 7$ has three terms making a three bar algebra earcon. The first term '$3x$' has a length of four beats: A note of one beat for the '3', two beats for the '$x$' and one silent beat for the '+' which separates it from the following term. The second term '4' has a length of three beats and the final term '= 7' has a length of four beats. Therefore, the bar length of this earcon is four beats. The first and third terms already fit into this bar length. The second has an extra silent beat added to make it fit this length.

The next stage is to assign pitches and timbres. A piano note at $C_3$ is used for the '3' and one at $B_4$ for the '$x$'. For the start of the new term, the note representing '4' is again played at '$C_3$'. The marimba timbre used for '=' is played at $A_4$. To emphasise the pitch fall at the end of the expression, the piano note for '7' is played two notes below this at $F_4$.

This is a simplified set of rules for constructing algebra earcons. Care was taken to give the earcons a strong rhythmic component, to aid retention and discrimination of syntactic structure [3]. Distinctive timbres must be used to aid this discrimination [2].

The algebra earcons were constructed with a sequencer. The sounds were all played through a Yamaha DMP11 mixer controlled by an Apple Macintosh and presented via external loudspeakers.

## 2.2 Procedure

The experiment was in two parts: In the first part earcons for the simple expressions were presented and in the second the earcons for the complex expressions. For each part there were seven training earcons that reflected all the material in the condition. Each training earcon was played once. An appropriate expression was spoken by the experimenter and the earcon was explained in terms of the prosodic cues used. Each stimulus earcon was played once. When the earcon was finished the subject chose one of four possible response expressions.

## 2.3 Results and Discussion

Subjects performed much better than chance in both simple and complex conditions. In both conditions the means were approximately 11 correct in 15 responses (see Figures 1 and 2 for scores across subjects for each condition). A binomial test for 11 correct in 15 responses, with a probability of success being 0.25, gave a probability of this result happening by chance of 0.0001. Listeners were able to recover enough syntactic information from the algebra earcon to choose an appropriate expression from a list of similar alternatives.

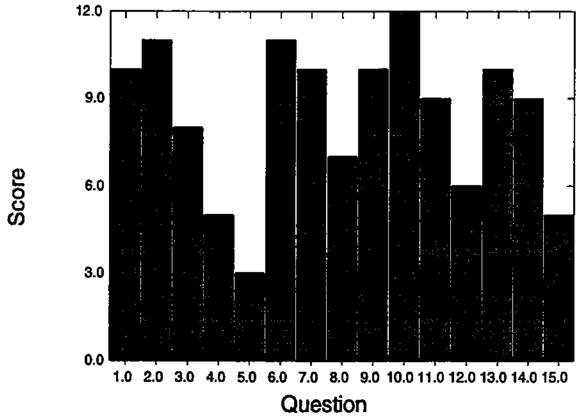

Figure 2: Scores across questions for the complex stimuli expressions.

Those subjects who had musical training performed significantly better in this task ($T(22) = 3.94, p = 0.0007$). This training may have enabled them to extract more information, or recover the information more easily from the audio glance. Earlier work on earcons [2] did not find this difference in performance. The difference found here may be due to the more complex and variable stimuli and the very different task. Whether this finding has important implications for the use of algebra earcons will only be shown by the influence of this training on the type of mental representations derive from the glance. This suggested that algebra earcons could convey the syntactic complexity of an expression. However, it told us nothing about what kind of mental representation a listener had of the expression after having heard the earcon. The subject could use the printed expression in the recognition task, something a blind user could not do. To be useful, a listener needed to retain and recall an expression's structure from the glance.

An examination of the results across questions revealed which presented the most problems. The choice of incorrect answer should high-light problems with the glance. Figures 1 and 2 show the scores for each question. Incorrect answers were usually concentrated on one or two of the distractors, making the determination of faults in earcon design easier.

For question 15 (see Figure 3 top), only two subjects gave answer **B**, which differed from **A** by omitting the coefficient from the first term. This is a trivial mistake and may not affect how the user would plan a reading. If either **A** or **B** were held as mental representations by the user, they would be a good framework by which to guide the reading. Three subjects did not recover a superscript from the earcon and chose answer **C**, a more severe error. Such a mental representation, if it were held, would provide a good guide to syntactic complexity, but not an exact guide to the expression. These are examples of the common error of missing an item or group of items, but to pick an expression of the right form. Alternative **D** is never chosen, probably because of the equals symbol. Subjects commented that the relational operator was a very useful discriminant in the glance. Question 5 (Figure 3 bottom) is the only question in which the correct answer was not the most frequent response. In choosing answer **B**, the error was not to perceive $\frac{a}{b}$, but $ab$. Subjects may have remembered the two sounds, but not their form. It may have been that the fraction timbre was not distinctive enough. Many subjects complained that these sounds were too fast, and even if recognised as a simple fraction, the internal structure was not noted. Mistakes with the representation of the simple fraction were common in other questions. The choice of **C** is a timing error. A gap is perceived between the fraction and the subexpression. Mistaking products as sums, and vice versa, was also a common timing fault in other earcons. Not choosing **D** is an example of how strong a cue the relational operator is for choosing an expression. This was reflected throughout the experiment, but subjects noted that

15 **A** $\boxed{ax^3 + bx^2 + cx + d}$
**B** $x^3 + ax^2 + bx + c$
**C** $ax^2 + bx + cx + d$
**D** $ax^3 + bx^2 + cx + d = e$

5 **A** $\boxed{\frac{a}{b}(cd + e) = f + g}$
**B** $ab(cd + e) = f + g$
**C** $\frac{a}{b} + (cd + e) = f + g$
**D** $\frac{a}{b}(cd + e) + fg$

Figure 3: Questions 15 from the simple condition and 5 from the complex condition. Correct responses appear in boxes.

the timbre used could be overwhelmed by other sounds.

The errors described in the previous two examples, and those found in the other questions, have been put into the following categories:

- Simple fractions were the cause of many errors. The type was often not perceived. Discrimination of any detail was rare.
- Distinguishing the temporal relationship between items caused many problem.
- Both the pan pipes used for fractions and marimba used for relational operators were not prominent enough. Subjects forgot mappings between timbres and syntactic types.
- Items were sometimes missed or forgotten. This frequently happened with items from the start or end of the expression. Superscripts were particularly prone to this problem.

The rules devised for constructing algebra earcons gave a glance at an expression's syntax, but improvements could be made. A new timbre for relational operators had to be found. The representation of simple fractions had to be improved. These two changes would give the greatest enhancement to the earcons. Some changes were also needed for terminal superscripts to make them easier to recognise.

The two other major sources of error were the recovery of grouping information and loss of algebraic items. Perception of timing structure may improve with practice. Investigation of strategy and appropriate retraining may reduce the errors due to missed or forgotten information and misperception of timing in algebra earcons.

## 3 Experiment Two

A second experiment was performed to assess the changes to the rules for earcon construction and to investigate the mental representation a listener could derive from the audio glance.

The second goal was important because the user will ultimately have to recall information about the expression implied by the glance to plan the reading of the expression. Algebra earcons allow a listener to derive a variety of internal representations. A full and correct representation would give complexity and a cognitive framework into which a listener could fit lexical details as he or she heard them. The lowest level at which the audio glance could work is as a crude guide to length and complexity of an expression.

The same multiple choice design and stimuli were used to ensure a valid test of the new earcon rules. Three months had elapsed between the two experiments, minimizing any effects of remembering stimuli. The design for this experiment differed in that subjects were asked to 'think aloud' after hearing an earcon. This verbal protocol was divided into two parts. The first, occurring just after hearing the stimuli, was a description of what the subject recalled of the earcon or what representations they had of the expression. The second part was a description of how the subject was choosing his or her answer from the multiple choice question. This approach should have revealed what types of representation were derived and retained from the glance and what strategies were used to select an answer.

## 3.1 Sounds

The same basic rules for earcon construction were used, but were altered as follows:

- The more distinctive 'rim shot' percussion timbre was substituted for the marimba for representing relational operators.
- The simple fractions had a one beat pause introduced between numerator and denominator. The first item in the denominator was no longer emphasised, but the pan pipe timbre was increased in volume relative to the other timbres. These changes attempted to make the simple fractions more prominent and easier for subjects to ascertain the internal structure.
- Superscripts were all played at the pitch $C_2$. This should avoid any confusion with terminal superscripts.
- Subexpressions were all played at pitch $B_4$. This change, and the superscript pitch change, made the task of constructing earcons easier, without any decrease in information content.

## 3.2 Procedure

Six of the original twelve subjects were retested in this experiment. Three subjects from the top of the range and three from the bottom were chosen. Two of the top scoring subjects were not available, so the next best were substituted.

Each subject was trained as before. After the earcon was played, the subject was asked to describe what he or she had heard. Once this description was given, the subject was asked to verbalise the process he or she used to select the answer.

## 3.3 Results and Discussion for Experiment Two

Figures 4 and 5 show the scores for each question in the second trial. The results showed subjects still achieved a good score, despite the interference of 'thinking aloud'. Five of the six subjects showed an improvement, the overall score of the sixth decreasing by three, causing the improvement to become nonsignificant. Subjects at the bottom of the range showed the greatest improvement.

Figures 4 and 5 show the scores for individual questions. Many of the same stimuli still caused problems, but the proportion of errors decreased. On others less severe errors were made. For example, choosing an alternative that differed from the correct answer by omitting one item.

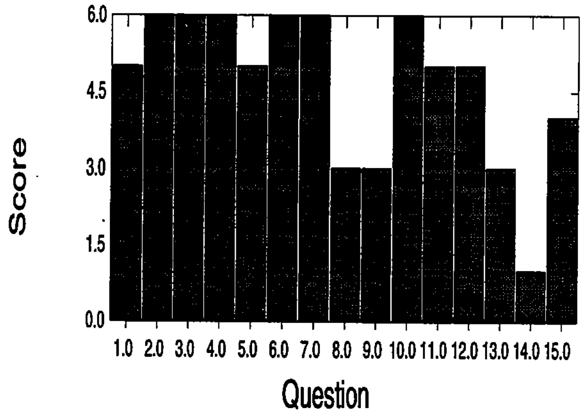

Figure 4: Scores across questions for simple expressions.

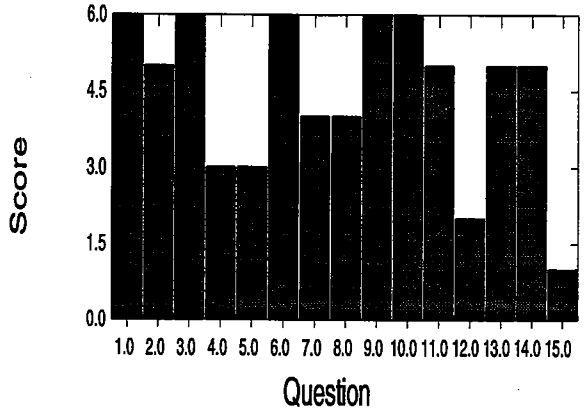

Figure 5: Scores across questions for complex expressions.

Performance for the two example questions above showed a marked improvement. For question 5 half answered correctly, as opposed to only one quarter. Only one subject did not distinguish the simple fraction from a product, indicating that the representation of simple fractions worked much better. Two subjects made timing errors, making the left hand side a sum, rather than a product. Therefore, perception of timing groups may still be a problem.

For question 15, a majority (4/6) answered correctly. The error of missing an initial term did not occur, perhaps indicating use of a better strategy. One error is a trivial missing of a coefficient from the first term. The inclusion of the equals symbol in one answer is unusual and has no apparent explanation as relational operators were sometimes missed, but hardly ever erroneously included.

Examination of other answers supported the improvement of recognition of simple fractions. Errors due to a missing relational operator did not occur in either experiment, but complaints about difficulty in discrimination of its timbre were not made in the second experiment. A small number of errors were made by transposing an equals sign with another operator. This could have been due to poor recall, but the rim shot timbre on the synthesiser was sustained and some subjects complained of its 'overlapping' with other sounds.

Two major sources of error remained. Many errors were due to superscripts not being perceived or recalled. The changes made did reduce some of the errors, but superscript representation may have to be redesigned. The other major error was still perception of timing groups. It was hoped that listeners may have learned this aspect of the glance, but this appears not to have happened. The next round of algebra earcon design will make the timing structure more prominent. One facet of timing problems may be the selection of a response with a missing item. In many cases an answer with the right structure was selected, but may have had a single operand missing from a term, too short a subexpression or one of a repeated term missing (see question 15). In the context of an algebra earcon as a glance this type of error may not be too important.

### 3.3.1 Protocol Analysis

The recall part of the 'thinking aloud' experiment was used to elicit the type of mental representation gained from the glance. The descriptions of expressions given during the 'thinking aloud' part of the experiment have been put into the following categories:

- idea of complexity or length,
- low-level knowledge of complexity: Equation or expression, balance of left and right and sides, some knowledge of syntactic items,
- knowledge of major syntactic features, some detail and knowledge of their order, and
- detailed representation of structure, a framework into which detail could accurately be placed during reading.

All of these representations could be useful as a glance because they would indicate the syntactic complexity of an equation. However, a strong, but inaccurate framework has the potential to mislead a reader. As algebra earcons were only designed to provide a glance, such inaccuracies would not be too great a problem because any glance is not supposed to be entirely accurate. A good representation of the equation would be a bonus for the reader. The task forced subjects to recover as much information as possible from the earcon and meant that subjects were probably not using the earcons simply as a glance. Recovering information from the glance may be a difficult task, as described by many subjects, but this may be exacerbated by the novelty of the audio glance and the artificial nature of the experiment. In addition, the difficulty of using the audio glance has to be balanced against having to use a full utterance to guide the reading process.

Those subjects retaining a good framework for an expression were able to directly match this to one of the alternatives presented and choose the correct answer. In many cases, for both good and poor representations, the written expressions acted as a prompt, reminding subjects of features that may have been lost or not discriminated. Such external help would not be a strong feature of reading using Mathtalk, but indicated that a slightly inaccurate representation from the glance would not cause too many problems.

As the quality of the representation gained from the glance decreased, subjects used more elimination strategies to choose their answer. Prominent amongst these was use of the relational operator. Often the scope of the choice could be reduced simply by remembering whether a relational operator was present. Location of such an operator within the expression was also useful, allowing a choice to be made between relative sizes of two sides of an expression. A similar strategy was to use the presence or absence of significant syntactic subunits derived from the glance. These strategies suggest the algebra earcon is working as a glance.

## Conclusions

The first experiment showed that listeners were able to derive syntactic information from an algebra earcon and use this information to select an appropriate expression from a list of similar alternatives. Experiment one suggested several changes that should be made to the earcons. Amendments were made to the representation of fractions, relational operators and superscripts. Changes to the first two proved successful. Both experiments revealed the importance of being able to use the timing information to apprehend the structure of the expression. One of the major differences between high and low performers seemed to be the ability to perceive and use this structure. This ability may depend on musical training. The rules for algebra earcons will be changed to make the timing more distinctive.

These experiments have shown that algebra earcons can work as an audio glance at an algebra expression. Listeners should be able to use the high-level information on an expression's complexity to plan a reading strategy using the other functions of the Mathtalk program. Further changes will be made to the earcon rules to enhance their use. This sonification of algebraic structure has the potential to facilitate enhanced reading of such information by blind people using either speech or braille. The use of such audio glances could also be extended to other planning task such as orientation and navigation while reading.

## References

[1] Blattner, M., D. Sumikawa, and R. Greenberg. "Earcons and Icons: Their Structure and Common Design Principles." *Human-Comp. Int.* **4(1)** (1989): 11–44.

[2] Brewster, S. A., P. C. Wright, and A. D. N. Edwards. "A Detailed Investigation into the Effectiveness of Earcons." In *Auditory Display: The Proceedings of the First International Conference on Auditory Display*, edited by G. Kramer, 471–498. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[3] Deutsch, D. *Psychology of Music.* London: Academic Press, 1982.

[4] P. Ernest. "A Model of the Cognitive Meaning of Mathematical Expressions." *British J. Ed. Psychol.* **57** (1987).

[5] Lehiste, E. *Suprasegmentals.* Cambridge: MIT Press, 1970.

[6] Stevens, R. D., and A. D. N. Edwards. "Mathtalk: The Design of an Interface for Reading Algebra Using Speech." In *Computers for Handicapped Persons: Proceedings of ICCHP '94, Lecture Notes in Computer Science 860*, edited by W. L. Zagler, G. Busby, E. R. xuxxia, and R. Wagner, 313–320. Berlin: Springer-Verlag, 1994.

[7] Stevens, R. D., P. C. Wright, and A. D. N. Edwards. "Prosody Improves a Speech Based Interface." In *Ancillary Proceedings of HCI'94*, edited by D. England. London: British Computer Society, 1994.