USING AUDEMES AS A LEARNING MEDIUM FOR THE VISUALLY IMPAIRED

Steven Mannheimer, Mexhid Ferati, Donald Huckleberry and Mathew Palakal
Indiana University School of Informatics
535 West Michigan Street, Indianapolis, Indiana 46202

Keywords: Audeme, sound, Blind and visually impaired, Children, Cognitive long-term memory, Education.

Abstract: In this paper we demonstrate the utility of short, nonverbal sound symbols—called "audemes"—in the encoding and recalling of text-based educational materials. In weekly meetings over a school year with blind and visually impaired pre-college students, we explored their capacity for long-term memory of individual audemes, audeme sequences, and textual content presented in conjunction with these. Through interviews and group discussions, we also explored the ability of these students to create intuitive narratives enabling memory of complex audemes and series of audemes. Further, we explored the mnemonic power of positive affect in audemes, and the impact of thematic association of information-to-audeme. Our results showed that the use of audemes can improve encoding and recall of educational content in the visually impaired population. The ultimate goal of our work is implementation of an "acoustic interface" allowing users to access a database of audemes and associated text-to-speech content.

1 INTRODUCTION

In recent years, the proliferation of primarily visual, screen-based information technologies has accentuated the difficulties of the blind and visually impaired in education. Although "screen readers" (text-to-speech) applications are widely available, their use is tedious and affect-less. Very little research or development work has exploited the capacious human memory for everyday noises, music and other cognitively rich, non-verbal sounds, and almost none has addressed complex sound combinations. Advances in digital technologies have made sound recording and reproduction very easy, but little work has been done before now to explore the mnemonic or semantic power of nonverbal sound and its potential use in informational systems. (Turnbull et al, 2006).

The preliminary hypothesis of our research project is that short nonverbal acoustic symbols, called *audemes* (to suggest an auditory lexeme and phoneme) can serve as substitutes for visual labels/icons to improve computer access to educational material for visually impaired users.

An audeme is a crafted audio track, generally in the 3-6 second range, used to signify and cue memory about a theme and/or text. Audemes comprise combinations of 1) the iconic sounds made by natural and/or manufactured things (e.g. rain on the pavement, car ignition); 2) abstract synthesized sounds (e.g. buzzes, blips); 3) music; and 4) occasional snippets of language gleaned from songs (e.g. *I've Been Working on the Railroad*). In our work with the students from the Indiana School for the Blind and Visually Impaired (ISBVI) we found that audemes work best when combining at least 2 but no more than 5 individual sounds. We also found that nearly all iconic sounds could be recognized quickly, in 2 seconds or less. This let us construct relatively complex audemes containing 2-5 sounds in sequence and/or simultaneous layers, to convey relatively complex significations. For example, an audeme for the American Civil War contained snippets of the *Battle Hymn of the Republic* and *Dixie,* staggered but overlapping and conflicting, followed by the sound of rifle and cannon fire, all combined in a 5-second audeme. Audeme design was best approached through a dialogue between designers and users to determine the most effective combination of sound elements to best represent a target theme. As our work evolved, we came to understand the components of an effective audeme to include acoustic uniqueness (one audeme is not easily confused with another), thematic relevance (the audeme contains music or sounds easily associated with the theme) and emotional quality.

Mannheimer S., Ferati M., Huckleberry D. and Palakal M. (2009).
USING AUDEMES AS A LEARNING MEDIUM FOR THE VISUALLY IMPAIRED .
In *Proceedings of the International Conference on Health Informatics*, pages 175-180
DOI: 10.5220/0001532801750180

We also noted the semantic flexibility of many audemes and sequences, i.e. that they could suggest a range of possible interpretations depending on surrounding audemes or explicit context. This flexibility produced much of the “fun factor” for our student subjects. They enjoyed generating internal narratives that could explain audemes or audeme sequence and competed to offer ingenious explanations.

We collaborated with the staff and students of the ISBVI, enrolling a variable cohort of approximately 20 students. In our initial studies, they were divided into three groups: two groups heard informative essays with a thematically related audeme, while the control group heard the same essay without the audeme. In tests conducted two weeks later one group was tested on this essay while hearing the audeme; two other groups were tested without the audeme. This same test structure was repeated three times with three distinct audemes and three separate essays. In each of the three tests the group that heard the audeme with the essay and also during the test showed the greatest improvement in recalling the information. The group that heard the audeme during encoding but not during testing also showed superior results over the control group, which never heard the audeme. We conclude that the use of audemes improves the participants’ abilities for the functions of both encoding and cuing memory of information.

2 PREVIOUS WORK

Foundational work in psychoacoustics (Back, 1996) raised questions about how speech and non-speech stimuli earned long-term memory. With the advent of the personal computer, the Graphical User Interface (GUI) and mouse navigation in the 1980s, important work in the development of acoustic or auditory interfaces was performed in the late 1980s by researchers such as W. Gaver (Gaver, 1989), S. Brewster (Brewster, 1994), M. Blattner (Blattner et al, 1989), A. Edwards (Edwards, 1989) and others. Most of this concerned auditory enhancements to GUIs. Researchers predicted that sound-based interfaces could be useful for the blind or in “eyes free” contexts such as driving (Edwards, 1989)(Stevens and Brewster, 1994).

Smither (Smither, 1993) and Brewster (Brewster, 1994) agreed that natural speech is more readily understood and remembered than synthesized. Some explored the relative value of abstract sound (buzzes, beeps, et al.) or earcons (Blattner at al, 1989), vs. natural sounds or acoustic icons (e.g., the sound of rain). Gaver (Gaver, 1989) suggested that natural/iconic sounds are both more long-lasting in memory and better able to conjure a range or depth of content associations. Conversy (Conversy, 1998) suggested that it is possible to synthesize abstract sounds for natural phenomena such as speed, waves or wind, and these will fully convey meaning. Back and Des (Back and Des, 1996) report that popular media have had a strong influence on how we expect the natural world to sound. Mynatt reports that a recorded sound must fit the mental model we have for that sound: “Thus, thunder must crack, boom, or roll…listeners will reject any of the myriad of other sounds made by thunder or seagulls as not authentic.” (Mynatt, 1994) Some researchers believe that a judicious mix of all types of sound cues may be the best way to proceed (Frohlich and Pucher, 2005).

Studies have demonstrated that sound can be a powerful catalyst to memory (Sanchez and Flores, 2004). Some studies involved visually impaired students (Doucet et al, 2004) and generally report higher mnemonic performance than sighted students (Sanchez and Jorquera, 2001). A few of these studies have highlighted learning and short-term memory (Sanchez and Flores, 2004). We are not aware of studies of acoustic enhancement of long-term memory. This study helps to fill this gap.

3 METHODS

3.1 Experimental Environment

3.1.1 Audemes

Audemes are short sound combinations, approximately 3-6 seconds, of sound icons or effects, music and abstract sounds, with occasional song lyrics or, rarely, recorded speech. We constructed our audemes with commercial sound effects libraries and Soundtrack Pro software. We established meanings for our audemes initially through discussion with our subjects and then through researchers using their own best judgment. For the three initial tests we created audemes for “Radio,” “Slavery” and “US Constitution.” The “Radio” audeme was the sound of a radio dial being twisted through stations. The “Slavery” audeme combined a short passage of a choir singing “Swing Low, Sweet Chariot” followed by the sound of a whip crack. The “US Constitution” audeme combined the sound of a gavel (symbolizing courts), the sound of quill pen writing and the *Star Spangled Banner*. We also created three thematically appropriate essays, each approximately 500 words, from Web-based source.

3.1.2 Participants

We conducted weekly, one-hour sessions with approximately 20 students from the Indiana School for the Blind and Visually Impaired (ISBVI). Students were of different ages, ranging from 9 to 17 years old; 11 of them were completely blind, and 9 were partially blind. External commitments meant the number of weekly participants fluctuated between 15-20. For their recruitment, consent of the school and their parents was granted. They were randomly recruited based on their willingness to participate and subject to the ISBVI staff's approval.

3.2 Experiment 1

Students were divided into three groups, carefully selected by ISBVI staff to ensure a balance of age, learning abilities, and level of visual impairment. Group I was the control group, while Group II and Group III were the experimental groups. A multiple-choice pretest of the yet-to-be-heard thematic essay was conducted with all groups to establish a baseline of their existing knowledge of the themes. The pretest contained 10 questions derived from the lecture and these were printed in Braille or large-print sheets. All three groups took the same test in the same classroom with no talking during testing. Afterwards, Group I was moved to a separate classroom; Groups II and III remained together. Group I (the control group) listened to a reading or text-to-speech rendition of the associated essay without any audemes being played. Group II and III listened to the same lecture with the relevant audeme played between each paragraph of the text, resulting in the same audeme being played 8-10 times for each essay. Two weeks after each initial session, a posttest was given. This test contained the questions from the pretest in randomized order and with three new questions serving as statistical noise. All questions were read aloud by researchers. Group I and II took the posttest without hearing the audeme, while Group III heard the audeme played before and after each of the test questions. This allowed researchers to track how well students remember the lecture by itself after two week (Group I); if audemes enhanced encoding (Group II); and how well the audemes enhanced both encoding and recall (Group III).

Table 1 demonstrates that exposure to audemes in conjunction with text increased encoding and recollection of associated content. For Radio content, Group III showed a 52% increase in tested knowledge (from 4.2 correct answers to 6.4), factored against the pre-knowledge. For US Constitution, Group III showed a 65% increase (from 3.3 correct answers in pre-test to 5.50 correct in post-test). For Slavery, Group III showed an 80% increase (from 3.75 correct in pre-test to 6.75 correct in post-test). Group II showed a 38% increase in knowledge for Radio (from 4.2 to 5.80 correct); and a 16% increase for US Constitution (from 5.16 to 6.00 correct) and a 12% increase for Slavery (6.25 to 7.00). Group I, the control group, demonstrated a 47% increase in knowledge for Radio (3.40 to 5.00), then a 3.6% *decrease* in knowledge for US Constitution (4.67 to 4.50); and a 20% increase for Slavery (5.00 to 6.00).

Table 1: Results for all groups.

Group	I	II	III	I	II	III	I	II	III
Count(#students)	5	5	5	6	6	6	4	4	4
Mean(difference)	1.6	1.6	2.2	-0.17	0.83	2.17	1	0.75	3
% Improvement	47	38	52	-3.6	16	65	20	12	80

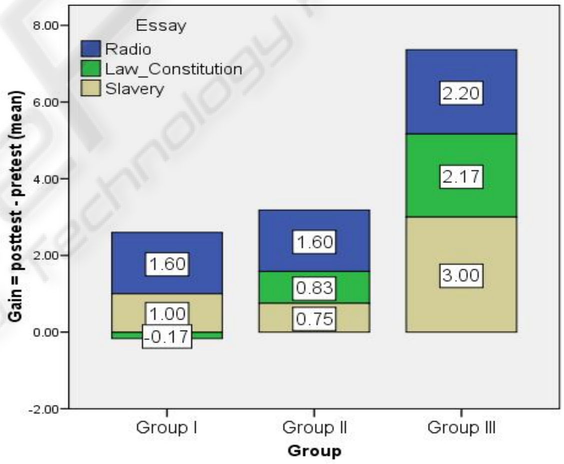

Figure 1: Cumulative improvement for each group.

3.2.1 Data Analysis

The statistical analysis of the data began by computing the difference between the pretest and posttest scores for each participant. Afterwards, we analyzed those differences in a One-Way ANOVA. This difference was called Gain.

$$\text{Gain} = \text{posttest} - \text{pretest}$$

Table 2: Anova.

Gain

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	60.023	2	30.011	7.641	.001
Within Groups	329.931	84	3.928		
Total	389.954	86			

The p-value is .001 ($p<.05$), which means that there is significant difference in the level of improvement among the three groups.

3.3 Experiment 2

We then tested audeme sequences. Participants heard three distinct sequences named C (Charlie), D (Daniel), and E (Edward). Each sequence contained 6 audemes. The participants were asked to memorize the individual audemes, their order and their sequence identity. Through earlier interviews with the students we learned they can devise sequential narratives to encode 3, 4 or 5 previously unheard audemes. Students reported that six audeme sequences were difficult to narrate this way. As such, we decided to use six-audeme sequences to explore the possibility that students' memory might work in smaller "chunks" within longer, less easily encoded sequences. We first named and played the three audeme sequences separately for the students. One week later we played various shuffled mixtures of 3-6 audemes drawn from all three original sequences. We asked students to decide which of the original sequences (C, D or E) each shuffled sequence most strongly resembled.

We begin this experiment with four hypotheses in mind:

H1: Majority should win. For example, in the test sequence E5-D2-E4-E3, students should identify this shuffled sequence as most resembling original audeme sequence E.

H2: Audemes from the first (C1, D1, E1) and last (C6, D6, E6) positions in original sequences should disproportionately impact resemblance of a shuffled sequence to an original.

H3: Audemes given first (e.g., E4 in E4-D4-C4) or last (e.g. C4 in E4-D4-C4) position in a shuffled sequence should disproportionately impact resemblance of a shuffled sequence to an original.

H4: Consecutive audeme "chunks" (e.g., C3-4-5 in shuffled D5-C3-4-5-D4-2), should disproportionately impact resemblance to an original audeme sequence.

3.3.1 Testing for Majority

In the Majority test we tracked students' identification of resemblance between a shuffled test sequence and one of the three original sequences when a majority of the test audemes had been taken from that original sequence. We used a total of nine new audeme sequences devised from the original audeme sequences (C, D, and E) to test for Majority.

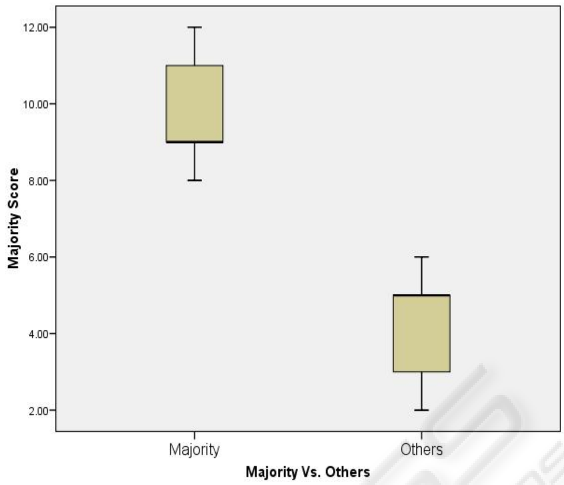

Figure 2: Majority test results.

Figure 2 shows the mean of 9.78 (variance = 1.694; stdev = 1.302) against 4.22 (variance = 1.694; stdev = 1.302) in favor of Majority. This means that 69.85% of the time participants identified resemblance of a test sequence to a given original sequence when a majority of audemes in the test sequence were taken from that original audeme sequence, against 30.14% for other cases.

3.3.2 Test for Original Position

The Original Position test tracks the relative influence of the First and/or Last audemes in an original sequence on students' ability to identify resemblance between original and test sequences. We used a total of nine new audeme sequences devised from the original audeme sequences to test for Original Position.

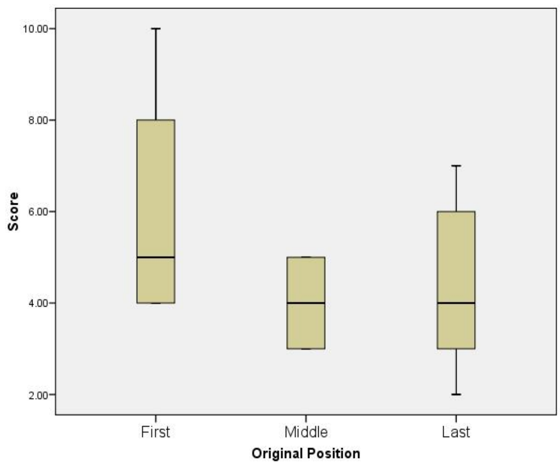

Figure 3: Original Position results.

Figure 3 shows the mean of 6.2 (variance = 7.2; stdev = 2.683) or 44.28% in favor of the First position against Middle position with mean 4.00 (variance = 3.5; stdev = 1.871) or 28.57% and Last position with mean 4.2 (variance = 3.511; stdev = 1.874) or 30%. This shows that the First position had the greatest impact on students' sense of resemblence between original and test sequences, and suggests a string impact on encoding sequence identity.

3.3.3 Test for Test Position

The Test Position test tracks the relative influence First and/or Last audemes in a new, shuffled test audeme sequence, regardless of their position in the original audeme. We used a total of nine new audeme sequences devised from the original audemes to test for Test Position.

Figure 4 shows the mean of 6.44 (variance = 6.528; stdev = 2.555) or 46% in favor of the Last position against the First position with mean 4.00 (variance = 2.5; stdev = 1.581) or 28.57% and Middle position with mean 3.33 (variance = 5.25; stdev = 2.291) or 23.78%. This proves that, all other factors being equal, audemes that occupied the last position in any original sequence had the greatest influence on students' interpretation of resemblance to a new test sequence. In short, the last audeme heard in a test sequence had the strongest influence on the students' perception of similarity to earlier, original sequences.

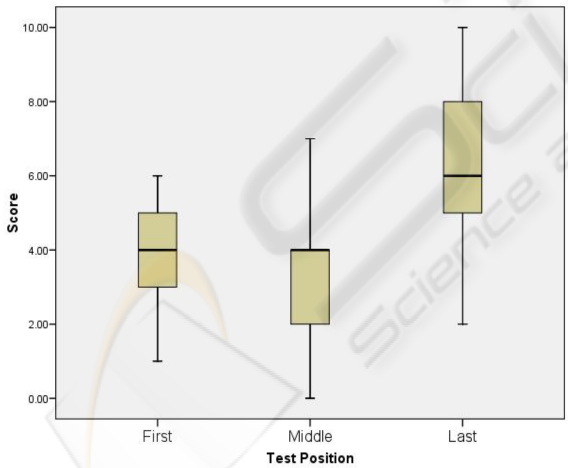

Figure 4: Test Position results.

3.3.4 Test for Consecutiveness

The Consecutiveness test tracks the impact of consecutive audemes taken from an original sequence and included, in the same order, in a new shuffled test audeme sequence. We used a total of 10 new audeme sequences devised from the original audemes to test for Consecutiveness.

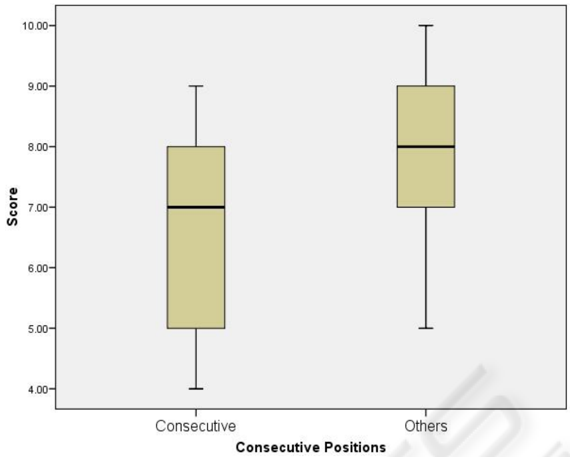

Figure 5: Consecutive Position results.

Figure 5 shows that the impact of the consecutiveness with a mean 6.6 (variance = 3.156; stdev = 1.776) or 47.14% is weaker than all other tested factors (Majority, Original Position, Test Position) which had a mean of 7.889 (variance = 2.861; stdev = 1.691) or 56.35%.

3.4 Experiment 3

We speculated that besides position of the audemes in a sequence, their affect (or perceived emotional quality as either negative or positive) should have great impact in their memorization and their ability to form lasting associations with other content. First, we presented all 18 audemes from C, D and E to the participants and they rated them as Positive or Negative. Preparing students for the test, we broadly suggested that they use any intuitive definition for "negative" (bad, unhappy, don't like, sad, unpleasant, etc.) and "positive" (good, happy, like it, pleasant, etc.). From their replies we devised a five-point emotional scale for these audemes.

Using this scale, we tracked the influence of emotional affect in memorization of 28 audeme sequences. The Positive audemes triggered better memorization in 67.86% of the cases against 32.14% of Negative audemes.

3.5 Experiment 4

We presented 20 short topics and texts along with a different audeme for each. Ten of the audemes were thematically or metaphorically related to the text (e.g., the crunch of footsteps in snow + gunfire = The Cold War) and 10 had no thematic relationship to their assigned texts, (e.g., a computer-made buzz + a brief passage of classical music = The National Grange). The audemes-text pairs were presented randomly. A week later we tested for recall of the associated text using 5-answer multiple-choice questions, with 10 questions for the themed audeme texts and 10 for the arbitrary audemes. Test results showed that the participants' recall of information associated with themed audemes was 67.82% (mean = 7.25; variance = 10.867; stdev = 3.296) better than their recall of information with arbitrary audemes 32.17% (mean = 3.437; variance = 8.396; stdev = 2.898).

4 DISCUSSION AND CONCLUSIONS

From our experiments and interviews with participants we have reached the following conclusions: (1) Audemes increase memory for associated text; (2) Participants display a strong ability to identify an individual audeme as a member of a six-audeme sequence and carry that sense of sequence identity forward; (3) Several factors impact the perception of resemblance or similarity between different sequences, including majority, position (when encoding), last position (when recalling), and positive affect of component audemes; (4) Themed and/or emotionally positive audemes are most memorable; and 5) intuitive narratives enhance affect and memory of sequences.

We believe our outcomes can be applied to a broader range of contexts for both disabled and mainstream populations. The increasing informational capacity of all technologies has elevated expectations that they will communicate, at least as to navigational cues and process status. These developments open the door for a broad range of new uses and new understandings of the power of acoustic information. Although current industry focus is on speech as the primary input/output system for acoustic interfaces, our studies strongly suggest that nonverbal acoustic information may actually prove more powerful and user-friendly in the conveyance of information.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Nina Mason Pulliam Charitable Trust.

Researchers want to thank the students and the staff of Indiana School for the Blind and Visually Impaired.

REFERENCES

Back, M. and Des, D. (1996) "Micro-Narratives in Sound Design: Context, Character, and Caricature in Waveform Manipulation," ICAD.

Blattner M.M., Sumikawa D.A. and M. Greenberg R.M. (1989) "Earcons and Icons: Their structure and common design principles." *Human-Computer Interaction*, Vol. 4, pp. 11-44.

Brewster, Stephen A, (1994) "Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces," PhD thesis, University of York.

Conversy, S. (1998) "Ad-hoc synthesis of auditory icons," ICAD.

Doucet, M.-E., Guillemot, J.-P., Lassonde, M., Gagne, J.-P., Leclerc, C. & Lepore, F. (2004) *Blind subjects process auditory spectral cues more efficiently than sighted individuals*. Springer-Verlag.

Edwards, A.N.D. (1989) Modelling Blind Users' Interactions with an Auditory Computer Interface, *International Journal of Man-Machine Studies*.

Frohlich, P. and Pucher, M. (2005) "Combining Speech and Sound in the User Interface," ICAD '05.

Gaver, W.W. (1989) The SonicFinder: An Interface That Uses Auditory Icons, W.W. Gaver, *Human-Computer Interaction*, Vol. 4, No. 1, pp. 67-94.

Mynatt, E. D. (1994) Designing with auditory icons: how well do we identify auditory cues? *Proceedings of the 2nd International Conference on Auditory Display.*

Sanchez, J & Flores H (2004) AudioMath: blind children learning mathematics through audio. *Proceedings of Fifth Conf. Disability*, Virtual Reality & Assoc. Tech., Oxford, UK.

Sanchez, J. & Jorquera, L. (2001) Interactive virtual environments for blind children: usability and cognition. Department of Computer Science, University of Chile.

Sanchez, J. & Flores, H. (2004). Memory enhancement through Audio. Department of Computer Science, Chile.

Smither, J.A. (1993) Short term memory demands in processing synthetic speech by old and young adults. *Behaviour & Information Technology* Vol. 12, No 6: 330-335.

Stevens, RD and Brewster, SA. (1994) Providing an audio glance at algebra for blind readers, *Proceedings of ICAD.*

Turnbull, D., Barrington, L., Torres, D. & Lanckriet, G. (2006) Modeling the Semantics of Sound. Department of Computer Science and Engineering, UCSD.