Use of Sound for the Interpretation of Impact-Echo Signals

M. L. Valenzuela, M. J. Sansalone and W. B. Streett
School of Civil and Environmental Engineering,
220 Hollister Hall, Cornell University,
Ithaca, NY 14853,
607-255-3438 (voice), 607-255-9004 (fax),
e-mail: MLV4@cornell.edu,
MJS39@cornell.edu, WBS3@cornell.edu

C. L. Krumhansl
Department of Psychology,
214 Uris Hall,
Cornell University,
Ithaca, NY 14853,
607-255-6351 (voice), 607-255-8433 (fax),
e-mail: CLK4@cornell.edu

ABSTRACT
This paper describes the development and evaluation of techniques for using sound to aid in the interpretation of signals obtained from the nondestructive testing of concrete using the impact-echo method. The impact-echo method and the significance of using sound for the field engineer are introduced. The auditory representation scheme developed and the software used are described. Psychological experiments that evaluate the effectiveness of the representation scheme are discussed. Results indicate the success of using sound to enhance signal interpretation in real-time and also suggest ways of using sound to train field engineers in the proper use of the impact-echo method.

Keywords
nondestructive, evaluation, concrete, sound, impact-echo

INTRODUCTION

This paper describes the development and evaluation of techniques and tools for using sound to aid in the analysis and interpretation of signals obtained from the nondestructive testing of concrete using the impact-echo method. A brief introduction to the impact-echo method is presented along with a discussion of the significance and implications of using sound for the field engineer. The auditory representation scheme, which maps data into a simple sound pattern, is described and the software and hardware tools are briefly discussed. The paper then focuses on the psychological experiments which are used to evaluate the learnability, patterns of confusion and the transferability of the representation scheme. Results indicate the success of using sound to enhance signal interpretation in real-time. Observations from the experiments also suggest ways of using sound in training to decrease the amount of time required to gain the expertise necessary for the field engineer to properly use the impact-echo method.

The Impact-Echo Method

Impact-echo is a method for the nondestructive testing of concrete and masonry structures that is based on the use of impact-generated stress (sound) waves that propagate through concrete and masonry and are reflected by internal flaws and external boundaries [1]. Impact-echo can be used to determine the location and extent of flaws such as cracks, delaminations, and voids in plain and reinforced concrete structures. When properly used, the impact-echo method has achieved unparalleled success in locating flaws and defects in highway pavements, bridges, buildings, tunnels, dams, piers, sea walls, and many other types of structures.

Impact-echo is based on the use of transient stress waves generated by elastic impact (see Figure 1). A short-duration mechanical impact, produced by tapping a small steel sphere against a concrete or masonry surface, is used to generate low-frequency (1 kHz to 80 kHz) stress waves that propagate into the structure and are reflected by flaws and/or external boundaries. Surface displacements caused by reflections of these waves are recorded by a transducer located adjacent to the impact. The resulting displacement versus time signals are transformed into the frequency domain to yield spectra which plot amplitude versus frequency. Patterns present in the waveforms and spectra (especially the latter) provide information about the existence and locations of flaws.

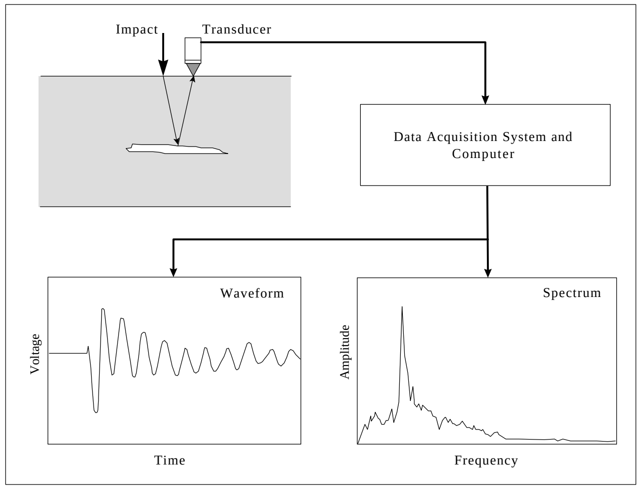

Impact
Transducer
Data Acquisition System and Computer
Waveform
Voltage
Time
Spectrum
Amplitude
Frequency

Figure 1

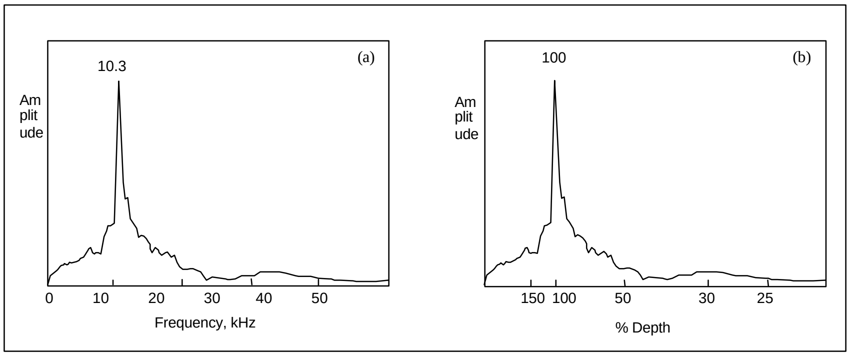


Figure 2. Spectra for a test of a solid plate: (a) amplitude spectrum, (b) normalized spectrum. The maximum peak is located at 10.3 kHz and represents the solid thickness frequency, f_T, that is, the frequency of waves reflected from the boundary of a solid plate.

The present study focuses on four major categories of flaw cases that are represented by a typical set of spectra shown in Figure 3. They are large flaw, shallow delamination, and small flaw (I) and small flaw (II), where the latter two are distinguished by the presence of a peak in the spectrum that indicates the depth of the flaw in the plate. Note the presence in each case of a large peak that is shifted from the 100% depth scale and is one of the major indicators of the presence of a flaw in plate structures.

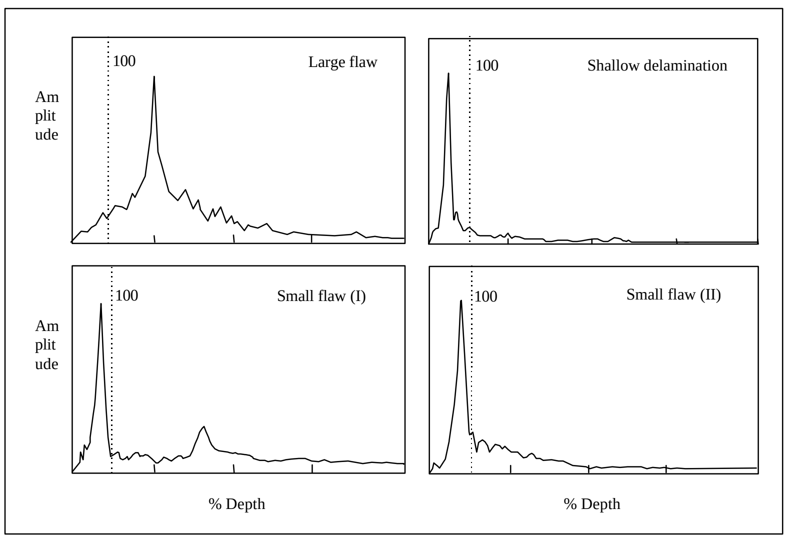


Figure 3. Four categories of flaws commonly found in plate structures. Note that the maximum peaks in the four normalized spectra are shifted from the 100% depth scale, represented by the dotted line.

AUDITORY REPRESENTATION SCHEME

In conceptualizing an auditory representation scheme for impact-echo signals, it is useful to discuss Kramer's distinction between *sonification* and *audification* [3]. The use of sound in general to represent data can be referred to as sonification whereas the direct playback of streams of data can be referred to as audification. Two examples in the literature that help illustrate the distinction are Lunney's work on auditory representation of infrared spectra for chemists [4] and Hayward's research in the audification of seismic waveforms [5]. Both examples are germane to the impact-echo method in their focus on the representation of spectra and waveforms, respectively.

Lunney developed a simple sound system for representing infrared spectra for the identification of organic compounds by visually impaired students and scientists. Continuous spectra are simplified into "stick" models that label the significant peaks. These peaks are then translated into a sequence of musical notes whose durations are determined by the height of the peak. Lunney's work is as an example of sonification.

Hayward explored the use of simple processing techniques to render seismic data into audible sound data. These techniques include frequency doubling and time compression. Hayward's work is an example of audification where there is a more direct and natural relationship between seismic waves (which obey the wave equation) and audible sound waves.

The use of the two labels helps distinguish between the application of auditory display to data exploration versus signal classification. A direct representation of data waveforms via audification lends itself to data exploration where more fundamental information, relationships, and trends in the time domain are being sought. Sonification, which uses a more symbolic approach than audification, seems more appropriate for signal classification by introducing one level of interpretation between the data and the listener, ideally the interpretation of an expert as embodied in the auditory mapping scheme.

Ultimately, the task of the engineer performing impact-echo testing is the task of signal classification and so the present work relies on sonification of the impact-echo spectrum. Our main objective in the development of an auditory representation scheme is to provide a means of distinguishing among the four different flaw categories and the solid category in real-time. The scheme highlights the significant peaks of the spectrum by representing each peak as a note in a musical sequence. The note value of the significant peaks is a step function of the frequency of the peak where the 100% depth is represented by a note value of middle C (MIDI Note = 60, frequency = 262.6 Hz) and the 1% depth is mapped to the C above middle C (MIDI Note = 72, frequency = 523.2 Hz). A graph of the relationship between percent depth and MIDI note value is shown in Figure 4. This tone mapping centers the values of percent depths within the musical range of audible tones. The duration of each note value is a function of the relative heights of each significant peak. Details about the determination of significant peaks can be found in the first author's dissertation [6]. The sound sequence begins with the reference tone (which represents the solid case for any plate regardless of depth) and is followed by the significant peaks of the spectra as they are read from left to right.

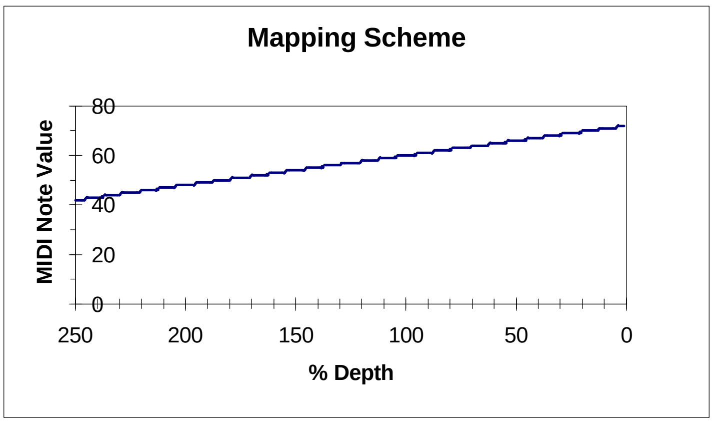


Figure 4. This graph represents the mapping of peak location (indicated by percent depth) to MIDI note value. Perhaps one of the most difficult concepts in impact-echo is the inverse relationship between percent depth and frequency. Note that the scale of the horizontal axis ranges from 250 % to 1 %.

The auditory mapping scheme developed in this work bears a strong resemblance to Lunney's work although devised without knowledge of the earlier efforts. In addition to the different characteristics of the spectra being sonified, the main difference between the present scheme and Lunney's work is the exploitation of the reference tone representing the 100% depth value. The use of the normalized spectrum and the 100% depth reference value helps reduce the number of auditory patterns needed for training engineers and reduces the burden on memory placed by variations in tone that might otherwise be due to different thicknesses and wave speeds.

Implementation

This highlighting scheme lends itself to the use of MIDI (musical instrument digital interface) for digital sound synthesis. MIDI provides the necessary standards for specifying the note value, duration and sequence of notes for easily producing a system that can represent sounds in real-time based on the data obtained from an impact-echo test. Because the existing data acquisition software for the impact-echo portable field unit has been written in VisualBasic, the sound generation software was also developed in VisualBasic on a Windows95 compatible machine with General MIDI capabilities. VisualBasic has proven flexible and easy to use in providing an interface to the MIDI standard.

CONCEPT ATTAINMENT EXPERIMENTS

The primary purpose of performing human subjects testing was to establish the learnability, to determine the pattern of confusion error, and to explore the transferability of the auditory representation scheme to a typical context encountered by field engineers. Experiment I used a classic paradigm for the study of concept attainment [7] and comparison was made between the use of auditory representations versus visual display of token spectra of the different signal categories. Experiment II required the subjects to apply their newly acquired concepts to a simulated field investigation.

Other researchers have investigated the effect of using auditory, visual or both auditory and visual displays in other applications. A relatively recent example of a careful study is the work of Fitch and Kramer [7] that examined the use of auditory display for the complex task of interpreting eight streams of physiological data encountered in the field of anesthesiology. One group of subjects was presented with auditory representations of the data while a second group was presented with standard visual displays. Because the subjects were not anesthesiologists, all subjects in both groups completed a three-stage training process that incorporated auditory, graphical and text information in preparation for a final testing stage that consisted of identification of "critical" conditions and selection of proper response. The published results showed significant improvement in response time for the group presented with auditory representations. However, no emphasis was placed on the results of the three-stage learning process leading up to the final testing stage. Focusing on a less complex auditory representation (in conjunction with a more simplified task definition) and examining the learning process using sounds might prove useful in gaining insight into the strengths and weaknesses of the particular auditory representation under study.

Experiment I

Experiment I has elements in common with classic concept attainment studies by both Bruner *et al.* [7], and Posner and Keele [9]. The classic study of concept attainment by Bruner *et al.* examined the strategies used by subjects in the attainment of arbitrary concepts. Typical stimulus materials were cards with geometric forms varying in number, shape, color and presence versus absence of borders. The rules defining the concepts were either conjunctive (*e.g.*, one black border *and* small red square) or disjunctive (*e.g.*, black border *and/or* red square). The latter concepts were more difficult to learn, which was attributed to a greater burden on memory and a general, cultural dislike of disjunctive concepts. The studies in [7] focused on the attainment of a single concept by subjects who are presented with tokens or examples of that concept.

In their research on generalization from past experience, Posner and Keele studied the ability of subjects to classify a set of patterns that were distortions of a prototype. Nine dots in a 30x30 grid were arranged to form prototype patterns such as a triangle or the letter F. The prototypes were then distorted by shifting each dot randomly a certain distance and direction according to a statistical rule. Tokens of these prototypes were presented to subjects until correct identification of a series of these tokens. The research showed that variability in the patterns had a positive effect on the learnability of concepts and transferability to new tokens (*i.e.*, correctly classifying tokens that were not part of the original training series). Patterns with low variability were easier to learn but those with high variability led to better transfer.

As in the studies by Fitch and Kramer and Posner and Keele, two groups of subjects participated in Experiment I. The results can be used to compare the effects of learning signal classification using different modes of presentation. One group (Group A) was presented with auditory representations of the token spectra of the above signal categories while a second (Group V) was presented with visual representations.

Subjects

Group A consisted of 2 staff, 6 graduate and 6 undergraduate students, all in civil engineering except for one undergraduate in computer science, at Cornell University. The 14 subjects of Group V consisted of 13 third-year civil engineering students and one graduate student in civil engineering all from Cornell. All subjects were given $5 for their participation except for one undergraduate who earned points towards credit in a psychology course.

Materials

The experiment software was developed in VisualBasic for both the control of the tone generation using the MIDI interface and the control of the graphical generation of simulated spectra. The tokens of the categories were all generated by the computer during the time of testing rather than relying on a limited set of actual field data. In the auditory version of Experiment I (Exp. I_A), parameters such as the number, location and height of significant peaks in a spectrum were chosen at random from a range of values appropriate to a given signal category. Subjects in Exp. I_A used headphones to listen to the sounds that were produced by the SoundBlaster 16 Value PnP sound card with OPL3 FM synthesis. Stimuli for the visual version of Experiment I (Exp. I_V) were similar except for additional parameters such as the sharpness of the significant peaks and the noise in the graphical spectra. Subjects used the Micron Millennia Plus computer with the Windows95 operating system.

Procedure

Experiment I required the subjects to correctly categorize 20 consecutive tokens of the five different signal categories. The testing software displayed on screen a typical spectrum for each of the five categories. In addition, the software for Exp. I_A

provided a description of the scheme for creating auditory representations as described above. Figure 6 shows the computer interface used in Exp. I_A. Subjects controlled when the tokens were presented, and for Exp. I_A, they could choose to play the representations as many times as desired before making a classification. Once the subject initiated the trial, the software selected one of the five signal categories at random along with a random sampling of the appropriate parameters and then presented the token (either aurally or visually). If the subject made a correct categorization, appropriate feedback was given and a visual counter was updated for the user. If the subject made an incorrect categorization, the software presented the correct answer and reset the counter to 0. The software kept a record of all actions by the subject, including the representation presented by the test software, the subject's categorization, the time interval between initiation of the trial and categorization by the subject, and the number of times a representation was played in the case of Exp. I_A.

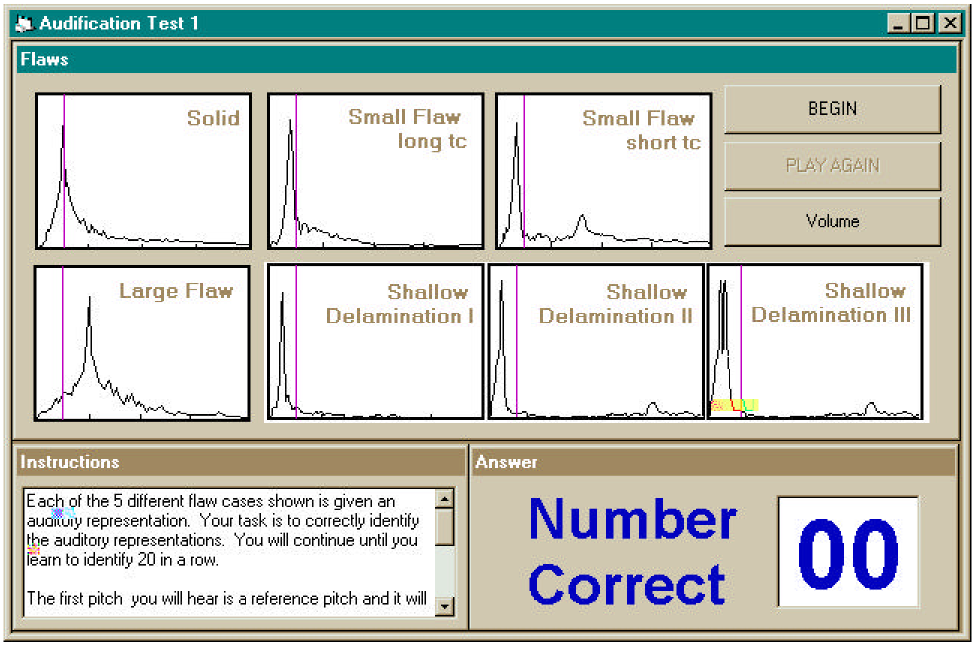


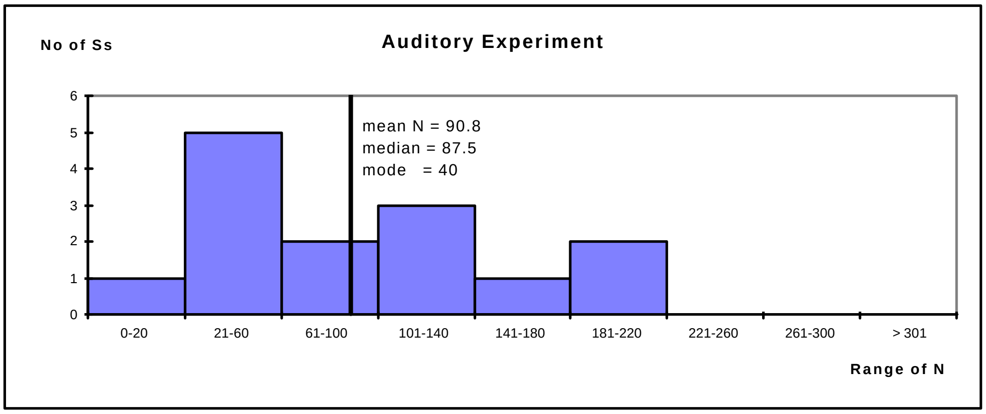
Figure 6. Computer screen interface used for Exp. I_A. Subjects used the mouse to initiate the trial by clicking on the button labeled BEGIN. The window labeled ANSWER also provided the correct answer when the subject answered incorrectly.

Results

Figure 7 shows a histogram of the number of trials, N, to criterion (correct categorization of 20 tokens in a row) for both Exp. I_A and Exp. I_V. The mean and variance for Exp. I_A are (90.76, 3539), respectively, and the mean and variance for Exp. I_V are (223.07, 31813), respectively. The histogram seems to indicate a decrease in the mean N when using the auditory representation. Graduate and undergraduate students who participated in Exp. I_A were evenly distributed within the histogram and no trend was easily discerned with respect to level of education. The one graduate student who participated in Exp. I_V completed the experiment in 184 trials, very close to the mean of 223.

To test the significance of the effect, a t-test is performed to check on the hypothesis that no effect has been detected (the null hypothesis). A group t-test assuming unequal variances yielded a result of $t(16) = 2.63$ and $p = 0.0090$ (one-tail) and $p = 0.018$ (two-tail), indicating that the null hypothesis could be rejected (*i.e.*, the difference in the means is statistically significant). Observe that the variance in the visual experiment is 10 times larger than in the auditory experiment. For measures with a boundary at zero (such as N trials to criterion), variances tend to increase with means. In this case, the increase in variance in itself is notable because large individual variations in the visual experiment imply the difficulty for some subjects in learning the concepts. Hence, training for the field application using visual display might prove difficult for some engineers.

Table 1 shows the results of the confusion matrix for Exp. I_A and I_V. The confusion matrix tabulates the classifications of the subjects in the two groups in response to the presentation of the different tokens. Individual subject responses were first normalized by column to mask out the effect of subjects who took longer to learn the concepts. Next, the entries from each individual confusion matrix cell are added together and divided by the total number of subjects. This gives the proportion of errors for the entire group. Observe that the values along the diagonals can be taken as a measure of the ease with which subjects learned to classify a particular category. Subjects in Exp. I_A were able to correctly identify the Solid, Small flaw (II) and Shallow delamination cases with somewhat greater accuracy than subjects in Exp. I_V. During informal postexperiment interviews for both groups, subjects expressed particular difficulty with distinguishing among the two small flaw and the shallow delamination categories. The subjects' observations were borne out in the confusion matrices. Of particular interest in impact-echo testing is the error of classifying a flawed area as solid, that is, the top row. A general comparison of the first row gives an indication of the occurrence of this critical type of error. Subjects in Exp. I_A seemed to avoid this type of error more often, as shown by the lower values in columns two through five of the top row.



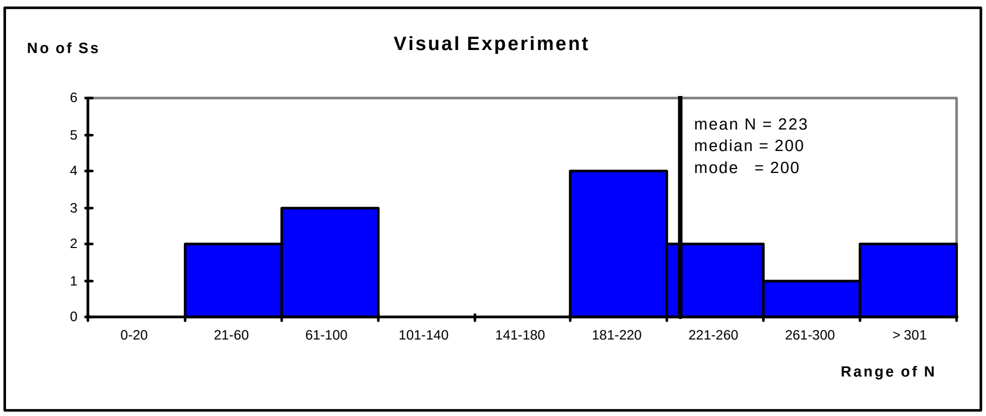


Figure 7. Histograms of the number of subjects (Ss) completing the task to criterion in N trials. The skew and kurtosis of the histogram for Exp. I_A are (0.132, 1.62), respectively. The skew and kurtosis for Exp. I_V are (1.59, 3.42) respectively.

Table 1. Confusion matrix for Exps. I_A and I_V.

Categorized as…		Represented as… Solid	Large flaw	Small Flaw (II)	Small flaw (I)	Shallow delam.
Solid	Exp. I_A.	0.981	0.012	0.019	0.002	0.006
	Exp. I_V.	0.887	0.019	0.074	0.007	0.027
Large flaw	Exp. I_A	0.002	0.910	0.041	0.004	0.032
	Exp. I_V	0.038	0.923	0.011	0.008	0.003
Small flaw (II)	Exp. I_A	0.014	0.039	0.776	0.025	0.128
	Exp. I_V	0.063	0.005	0.689	0.042	0.109
Small flaw (I)	Exp. I_A	0.000	0.028	0.017	0.751	0.125
	Exp. I_V	0.005	0.019	0.057	0.811	0.209
Shallow delam.	Exp. I_A	0.003	0.011	0.146	0.218	0.710
	Exp. I_V	0.008	0.033	0.169	0.132	0.651

Experiment II

After completing Exp. I, the subjects from Exp. I_A participated in a study of the transferability of the concepts learned. Rather than test the subjects' classification of tokens that were not previously encountered in the training set [*c.f.* 9], Exp. II examined the transferability of concepts to a typical context encountered in the field by engineers performing impact-echo evaluations. Namely, subjects from Exp. I_A were asked to evaluate a portion of an actual bridge deck for which impact-echo data was available. Because engineers work in an actual physical space, information about adjacent testing points may add to the conceptualization of the context of the structure under evaluation.

Expert engineers will also rely upon experience gained from the evaluation of structures under various situations to form an expectation of the type of flaw condition to be encountered. Consequently, in this experiment, subjects are told that an experienced engineer would expect to find shallow delaminations.

Ideally, subjects from Exp. I_V also would have participated in a similar experiment to explore the effect of different modes of presentation on the transferability of concepts. Practical time considerations precluded such an experiment so the current investigation focused only on the effect of context cues on the application of learned concepts for auditory representations.

Subjects

The same group of subjects who participated in Exp. I_A also participated in Exp. II. All subjects were given an additional $5 for their participation except for one undergraduate who earned points towards credit in a psychology course.

Materials

Exp. II relied on information obtained through impact-echo evaluation of an actual bridge deck [1, pp 103-110]. The information included the condition of the bridge deck at discrete points on the bridge (*i.e.*, solid or shallow delamination) and, in the case of shallow delaminations, the number, location and heights of the significant peaks in the spectrum.

Software for Exp. II was developed in VisualBasic and was similar to the one used in Exp. I_A. The software included an instruction/scenario sheet that explicitly stated: "Using good engineering judgment you suspect that the corrosion of the top layer of reinforcing bars due to the heavy use of road salt will cause shallow delaminations throughout major areas of the bridge deck." Subjects were also provided with plan and cross section drawings of the bridge. Subjects used headphones to listen to the auditory representation and used the mouse to select which grid points to examine and characterize based on the auditory representation. The same computer and sound synthesis hardware that were used for Exp. I were used for Exp. II.

Procedure

After subjects read the scenario and instruction sheet, they were presented with a 7 by 10 grid of points that represented a portion of the bridge deck. Subjects could play the auditory representations of any of the grid points and then evaluate that point using the five categories learned in Experiment I. If the subject evaluated the point incorrectly, an icon that indicated the choice made encased in the international "NO" symbol is presented at the grid point. If the subject was correct, an icon representing the categorization was displayed at the grid point. Subjects could revisit any grid point at any time and could change their answers as many times as desired. The software kept a record of the sequence of grid points visited, the time spent at each grid point and the actions performed (whether a token was played or whether a classification was made). Criterion is reached when the subject had correctly classified all the grid points of the bridge deck. The incorrect classifications were not recorded although the number of classifications at a particular grid point prior to correct classification was a good measure of error.

Results

Figure 8 shows the layout of the 7 by 10 grid points with a value indicating the average number of trials to correct categorization per subject. The 15 grid points that comprised the flawed areas of the bridge deck are shaded; the rest of the grid points are solid. As to be expected from Exp. I and confirmed by the average trials to correct classification, N, flawed areas caused more difficulty in categorization than the solid areas. Indeed, solid areas that seemed to present some difficulty lay close to the flawed areas. The subject's interpretation was perhaps influenced by the surrounding locations. However, identifying as flawed a point which is actually solid and not in need of repair is a conservative error. The more critical error is the reverse, classifying as solid a point that is actually flawed. On average, each subject made 2.67 errors in classifying the 15 flawed grid point during an average of 20 trials until correct classification of all the flawed grid points. The low error rate compared with Exp. I_A (which would predict an average of 5.8 errors) demonstrates that subjects successfully transferred their flaw categorization to the bridge scenario.

Number of trials to correct classification per subject

	1	2	3	4	5	6	7	8	9	10
A	1.13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.07
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.13
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.73	1.07	1.00
F	1.07	1.73	1.07	1.00	1.40	1.00	1.00	1.53	1.20	1.40
G	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.07	1.00

Figure 8. The 7 x 10 grid layout of the bridge deck used in Exp. II. Values can be taken as indications of the errors made in the classification of the flaws in the bridge deck. The shaded areas represent the flawed grid points and non-shaded areas represent solid points.

SUMMARY AND CONCLUSION

A system for the sonification of impact-echo signals has been designed, implemented and tested. The main findings of the psychological experiments were that (1) learning concepts using the auditory condition was significantly faster than using the current visual display and (2) the transfer of the auditory learning to a real-life situation — in this case a bridge — was successful.

Just as important as the improved learning rate in the auditory condition was the reduction in the critical error of judging flawed conditions as solid. A possible explanation for both improvements can be found in the work of Posner and Keele, which indicated that tokens of low variability were significantly easier to learn than tokens with high variability. The present auditory mapping scheme may work by eliminating the secondary characteristics of a spectrum and highlighting the essential information, namely the number, location and height of significant peaks.

Posner and Keele's work also showed that subjects trained on tokens with high variability had a significantly easier time transferring their newly attained concepts to tokens of the prototypes not previously encountered. Based on Posner and Keele's findings, it cannot be concluded easily that those subjects trained with visual displays would have had significantly better transfer to the bridge scenario. First, the error rates shown in Figure 8 are already quite low. Second, routine tests of a large structure would not exhibit great variability in the signals being presented (either as sounds or as visual displays). Third, it is uncertain what role context cues play in aiding the transfer rate for either visual or auditory displays.

Indeed, as Fitch and Kramer pointed out, it is quite difficult to devise "exactly equivalent" displays "across different sensory domains" [8, p 310] and thus test for effect of mode of display on rates of learning and transferability. In fact, an additional experiment is being carried out to determine if the advantage seen in learning rates and confusion errors for auditory displays is due to the mode of display or to the pre-processing of information prior to sonification.

Nonetheless, the results of these studies indicated that the proposed sonification of impact-echo signals was learnable and transferable to the case of the routine, rapid evaluation of large plate-like structures such as bridge decks. Regardless of the ultimate cause of the improved learning rate, the proposed sonification scheme provides the practical benefit of reducing the number of engineers and/or technicians required to operate the impact-echo instrumentation from two to one.

The experiments to study the effectiveness of sound in the interpretation of impact-echo signals is one step in a larger research effort to make sound available in real-time to the field engineer. One such part of this effort involves investigating various approaches to pick out the significant peaks of spectra for use in the auditory representation scheme. The full scope of this work forms the basis of the first author's Ph.D. dissertation in structural engineering [6].

REFERENCES

1. Sansalone, M. J., and Streett, W. B. *Impact-Echo: Nondestructive Evaluation of Concrete and Masonry*, Bullbrier Press, Ithaca, NY, 1997.

2. Pratt, D. G., and Sansalone, M. J. "The Use of Neural Networks for Automating Impact-Echo Signal Analysis." *Materials Journal of the American Concrete Institute*