AUDITORY DISPLAY OF DIRECTIONS AND STATES FOR MOBILE SYSTEMS

Gunnar Johannsen

Systems Engineering and Human-Machine Systems (IMAT-MMS), University of Kassel,
Moenchebergstr. 7, D-34109 Kassel, Germany
Email joh@imat.maschinenbau.uni-kassel.de
http://www.imat.maschinenbau.uni-kassel.de ; http://www.EngineeringAndMusic.de

ABSTRACT

Auditory displays for mobile systems, such as service robots, have been developed. The design of directional sounds and of additional sounds for robot states (e.g., Heavy Load), as well as the design of more complicated robot sound tracks are explained. Basic musical elements and robot movement sounds have been combined. Two experimental studies, on the understandability of the directional sounds and the robot state sounds as well as on the auditory perception of intended robot trajectories in a simulated supermarket scenario, are described. Subjective evaluations of sound characteristics such as urgency, expressiveness, and annoyance have been performed by non-musicians and musicians. These experimental results are compared with the diagrams which have been computed with two wavelet techniques for time-frequency analyses.

1. INTRODUCTION

Mobile systems are part of our society in many different application domains. Examples are cars and other vehicles in transportation, conveyor belts and vehicles in production plants, and mobile robots – as well as mobile phones and computers carried by people everywhere. Further, movable tools for telemanipulation in environments such as undersea, space, hazards, medical operating theatres (with invasive surgery), and micro (electronic) production can also be regarded as mobile systems.

Often, human users of such systems want or need to know the directions of movement and characteristic states of these systems. This paper suggests that information about directions and states can be communicated to the human by non-speech auditory displays. Beyond its general importance of this approach in a world of visual overload, it may also be particularly useful for blind people.

2. SOUND DESIGN AND EXPERIMENTAL OUTLINE

In this research, auditory displays were developed for intelligibility tests and for autonomous mobile service robots (as an example application domain) [1]. The idea is to combine relevant original noise signals of robot movements with basic musical elements as intelligible auditory symbols (earcons) and, then, to create sound tracks from them. This goes far beyond the investigation of the very important auditory warning displays [2]. Robots communicate their actual positions, movements and intentions as well as failures and related warnings by means of non-speech audio symbolic expressions to the human. The rich body of knowledge from music theory as well as from auditory science and sound engineering is exploited.

Several auditory symbols were invented, as directional sounds and as robot state sounds. They were composed as pure musical sounds with different melody and rhythm patterns. Results in the literature "show that high levels of recognition can be achieved by careful use of pitch, rhythm and timbre" [3]. All the sounds were designed and created by the author with a powerful PC and Windows, the Logic Audio software, the Cool Edit audio editor, a MIDI synthesizer, and a keyboard. The pure musical sounds from the synthesizer were recorded with the Yamaha DSP Factory with its audio expansion unit under the Logic Audio software on the PC. Thus, wav-files were produced which needed some minor audio editing.

The equivalent directional sounds of robot noises were derived from DAT recordings of the movement noises of a real service robot in the laboratory. The variation of the directional sounds based on robot noises were generated through several steps of audio editing from the recordings of the movements of this service robot. The same melody and rhythm patterns were composed by time and frequency editing as in the pure musical sound cases. The Logic Audio editor was used for frequency transpositions of the pitch levels as requested in the different tones of the directional sounds. The Cool Edit audio editor was more appropriate for cutting and assembling the necessary time slices of each sound element to the desired directional sounds. Some amplifications with fading-in and fading-out effects were performed for achieving a clear separation between the different tones of each sound.

Several experiments were performed. In the first experiments, human subjects had to learn, understand and recall auditory symbols with related meanings about spatial orientation and directional movements. In succeeding experiments, intelligible auditory symbols and sound tracks were presented in a supermarket scenario with the simulated environment of mobile service robots. Intended trajectories with directional sounds and robot state sounds for moving-curved, Heavy Load, Waiting, Near Obstacle, and Low Battery were communicated by the robot to the human. A group of eight non-musicians and a control group of two professional musicians participated in all the experiments.

3. EXPERIMENTS WITH DIRECTIONAL SOUNDS

3.1. Sound Design

For the first experiments, a new set of auditory symbols was designed as directional sounds for eight possible directions of motion of the robot in space. These directions are the four main directions of Left, Up, Right, and Down as well as the intermediate directions of Down-Left, Up-Left, Up-Right, and Down-Right. Each directional sound consists of three tones. The musical basic elements rhythm and melody are used in the four main directions, independently of each other. The directional sound Up is represented by a melody upwards whereas a melody downwards denotes the sound Down. In both cases, each tone is of equal time duration. A rhythm of two short tones followed by one long one, all on the same pitch level, means Left. Consequently, a rhythm with one long tone followed by two short ones on the same pitch level expresses the direction Right. The musical elements melody and rhythm are combined in the intermediate directions with respective intermediate values of melody span and rhythm. Each of the eight directions was presented in four variations, i.e., with changed sound colour (timbre) or changed tempo. Music instruments and robot noises were used. The four sound presentations were realized by marimba P1, harpsichord P2, robot-movement sound P3, and fast tempo of marimba P4.

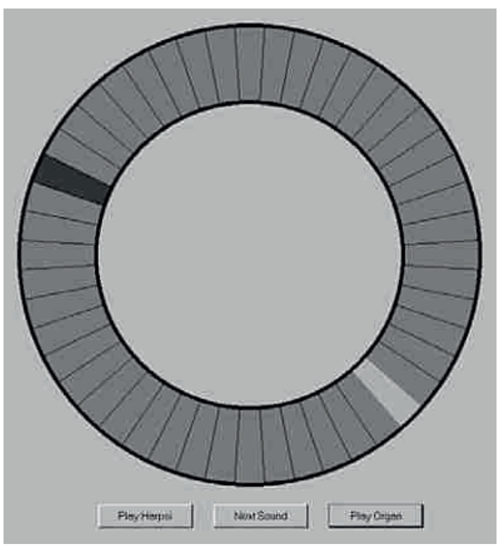

Figure 1. *Compass-card visualization for 48 directional sounds*.

The eight directional sounds are a subset of a more fine-grained resolution of 48 directions around the circuit of a compass card of 360°. This has recently been investigated in a student's project thesis. The experimental setup for the 48 directions is shown in Figure 1. For the eight directional sounds, only every sixth of all these directions, starting with the zero position for the direction Up, was included in the experimental set for the research reported here.

3.2. Experiments

The intelligibility and the recallability of auditory basic elements was checked in the first experiments. Human subjects had to learn and to recall the eight directional sounds with their related meanings. They had to process three test parts at the computer screen and heard the appropriate auditory symbols over loud-speakers. The experimental set-up was programmed in Delphi.

In the first part of the experiments, all eight directions in each of the four variations were presented, thus, altogether 32 auditory symbols, and that in random order. For each auditory symbol, its respective meaning was displayed immediately afterwards.

For refreshing the subjects' memory, they were able to select all eight directions again in the second part, each in the four variations, now in any order. After each auditory symbol, the subjects were asked for a subjective evaluation of its sound characteristics regarding Urgency, Expressiveness, and Annoyance. For this purpose, they had to make a selection in each case, and had to mark three Seven-Point scales with respect to their subjective impression

from - - - (very small) to + + + (very urgent),
from - - - (very small) to + + + (very expressive), and
from - - - (very small) to + + + (very annoying).

In the third part, the subjects should then show, how well they can recognize all eight directions independently of the presented four variations. They received a feedback after each of their decisions whether the respective auditory symbol had been recognized correctly or wrongly.

The experiments with the directional sounds were performed in all their three parts altogether three times, i.e., with two repetitions.

3.3. Results

The results of the experiments with the directional sounds show that the non-musicians were between 34% and 100% (one subject) correct whereas the two professional musicians were 97% and 100% correct. Subjective evaluations of sound characteristics such as Urgency, Expressiveness, and Annoyance were also performed by the non-musicians and the musicians. These characteristics were processed with respect to subject-related, indicated-direction-related, and presentation-related features.

During the second and last repetition of the experiments, the group of non-musicians was divided into two sub-groups of four subjects each. One sub-group was formed of the four subjects who showed higher performance during the first experiments whereas the other sub-group contained the four subjects with lower performance. Figure 2 shows the presentation-related features for the higher-performance sub-group of the non-musicians. The presentation of the robot-movement sounds P3 leads to clearly higher values of subjective evaluations for all three characteristics Urgency, Expressiveness, and Annoyance. All the presentations with the musical instruments are more or less on the same level. Only the faster presentation with the marimba P4 has been assessed slightly higher with respect to Urgency and Annoyance whereas the mean value for Expressive-ness is slightly lower.

Figure 2. *Presentation-related sound characteristics for non-musicians*.

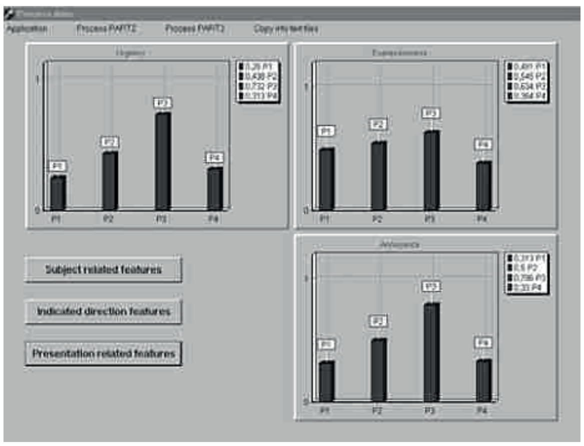

Figure 3. *Presentation-related sound characteristics for musicians*.

The same tendency with higher values for Urgency, Expressiveness and Annoyance for the robot-movement sounds can be observed for the musicians. Figure 3 with the mean values of the two musicians shows that this effect is much more pronounced for the characteristics of Urgency and Annoyance. These two characteristics seem to be highly correlated over all four presentations. For all three characteristics, the presentation with the harpsichord P2 leads to intermediate values between those for the marimba P1 and the robot-movement sounds P3.

The lowest evaluation of the Expressiveness within the indicated-direction-related features has been achieved for the direction Left (short-short-long) by the musicians. This effect can not be found with the non-musicians. The largest number of mistakes is made with the direction Down-Left by the non-musicians. One musician commented in the final interview that all the sounds towards the left with the rhythm short-short-long have been more difficult to be recalled. Nevertheless, he did not make more mistakes with such sounds.

4. EXPERIMENTS WITH SOUND TRACKS FOR MOBILE ROBOTS

4.1. Sound Design

In the experiments of the supermarket scenario, the auditory perception of intended robot trajectories and additional sounds was investigated. A simulated supermarket scenario was designed in such a way that a mobile service robot can make straight movements and turnings of 45 and 90 degrees. The sound tracks for the predictive display of the intended robot trajectories were composed of moving-straight segments and turnings. The moving-straight segments are represented by the above eight directional sounds. The directions Left, Right, Up, and Down are particularly used but also a few of the intermediate directions (Down-Left, Up-Left, Up-Right, and Down-Right) are sometimes possible. Down means downwards on the computer screen and towards the human subject (shown on the lower middle of the screen in Figure 6). Correspondingly, Up means away from the subject.

The turnings (moving-curved sounds) were derived from recordings of the original robot turning sound by transposition. They are always heard with any directional change between any kind of two complete moving-straight sections. If a complete moving-straight section consists of a number of straight segments of the same direction, the appropriate directional sound is repeated correspondingly without any turning sound in between. A segment is defined as the straight connection between two neighbouring active decision areas.

Additional sounds for robot states and situations were newly designed. These robot states and situations are Heavy Load, Waiting, Near Obstacle, and Low Battery. The latter sound was recorded from the original robot's indication of the low battery status. This is a continuous, quite annoying high tone. The other three sounds were played on the MIDI keyboard. The author tried to convey the subjective impression of the meanings of the three sounds. For example, the Heavy Load sound was played with three parallel Tubas as one accentuated short time-interval tone followed by one tone of a longer time duration. Figure 4 shows the score and the wave form of this Heavy Load sound.

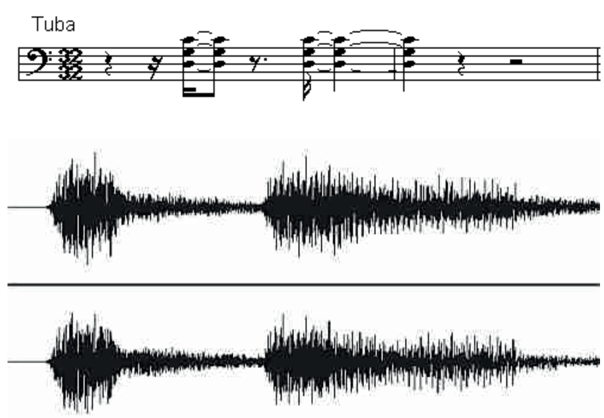

Figure 4. *Score and wave form of the Heavy Load sound.*

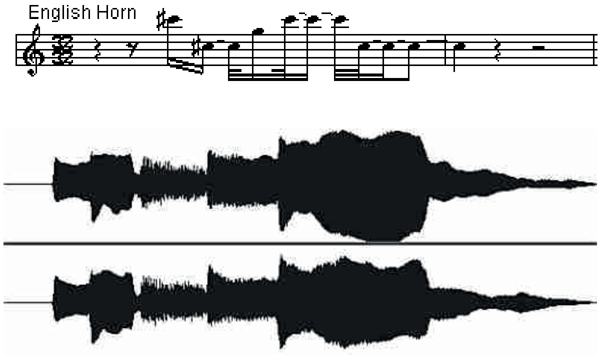

Figure 5. *Score and wave form of the Near Obstacle sound.*

The Near Obstacle sound was played with English Horn; its score and wave form is presented in Figure 5.

4.2. Experiments

In the second experiments, intelligible auditory symbols and sound tracks were presented in a supermarket scenario with a simulated (Windows and Delphi) environment of a mobile service robot. It is assumed that the supermarket is open during seven days a week for 24-hours. A mobile service robot for cleaning and for carrying goods will inform the human subject (the customer) with sound symbols of non-speech auditory predictive displays about the trajectory of its intended movements and about the additional robot states and situations Heavy Load, Waiting, Near Obstacle, and Low Battery. These additional sounds for the robot situations have to be learned by the human subjects in a training phase at the beginning of the second experiments. The subjects can listen to these sounds in any order as often as they wish.

A floor plan of the supermarket is visualized on the computer screen. The human subject is shown on the lower middle and the robot in different starting positions which are depending on the investigated trajectory; see Figure 6.

A matrix of decision areas was constructed. Any intersection between a horizontal and a vertical line together with the respective nearest surrounding of this crossing, in which alternative routes can be chosen (beyond returning the same way), is determined as an active decision area in the visual floor plan of the supermarket; see Figure 6.

The robot sound tracks actually used in the experiments of the supermarket scenario are the overlays of the sound tracks of the intended robot trajectories and, during some of their segments, of the additional sounds for the robot states and situations Heavy Load, Waiting, Near Obstacle, and Low Battery. The human subjects were asked to recognize and to understand the intended trajectory of the robot as well as the overlayed additional sounds of the robot situations, from listening to the robot sound track. They had to draw the auditorily perceived trajectory into the visual floor plan of the supermarket on the computer screen and had to mark the perceived additional sounds. The subjects were informed about the correctness of their auditory perception.

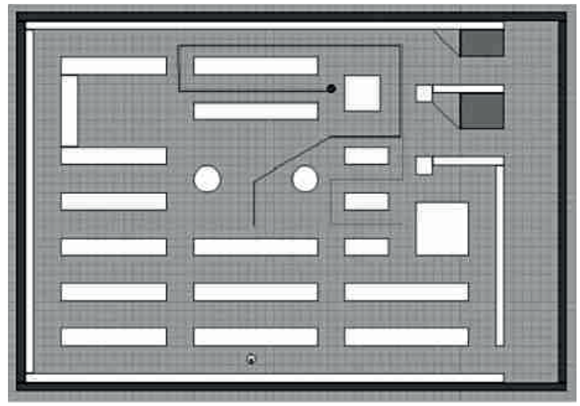

Figure 6. *Replay of intended and perceived robot trajectories in the supermarket scenario*.

Altogether, the subjects performed four sub-experiments, each with four different trajectories. The intended trajectory and the perceived trajectory as well as the intended and the perceived additional sounds were recorded. Also, the durations of the training phase and of the drawing of each perceived trajectory were measured. The intended and the perceived trajectories can be compared in the replay mode; see Figure 6.

In the last two of these sub-experiments, the sound tracks of the trajectories were composed of the same sound symbols of the intended trajectories and the overlayed additional sounds for the robot situations. However, the sounds of the real robot movements were now also overlayed. They were presented in real time whereas the intended trajectories are auditory predictor displays and, thus, faster than real time. This makes the scenario even more difficult for the subjects but it is also more realistic. In a real-world human-robot environment, the real robot movements are also always heard.

4.3. Results

The experimental results with the same eight non-musicians and the two professional musicians showed large differences in their auditory perception. The same three of these ten subjects who performed best in the first experimental study made only very few errors with the perception of the robot sound tracks of all trajectories as well as the additional sounds for the robot states and situations.

5. TIME-FREQUENCY ANALYSES AND SUBJECTIVE EVALUATIONS

5.1. Methods

A first attempt has been undertaken in this research to explain subjective evaluations of sound characteristics by means of image patterns in the time-frequency planes with the wavelet analysis techniques. The subjective evaluations of sound characteristics such as Urgency, Expressiveness, and Annoyance have been considered. Experimental results for selected directional sounds and robot state sounds have been compared with the diagrams which have been computed with two wavelet techniques for time-frequency analyses, namely the fast wavelet transform and the wavelet packet technique.

The directional sounds Down-Left and Left in their presentations with the marimba instrument and the robot-movement sound were analysed in more detail with the wavelet packet technique and the fast wavelet transform. In the same way, three of the additional sounds of robot states and situations were analysed with the wavelet packet technique, namely Heavy Load, Near Obstacle, and Low Battery.

The wavelet packet technique has been chosen for the time-frequency analysis of the selected sounds [4]. The wavelet packet analysis has been performed with the WaveLab toolbox as a software library of MATLAB routines [5]. Also, the fast wavelet transform has been accomplished with the same WaveLab toolbox.

The fast orthogonal wavelet transform with the Haar filter has been applied [4]. The transform computes the orthogonal wavelet coefficients and decomposes each sound signal into multiresolution signal approximations at different scales.

The wavelet packet technique segments the frequency axis of the time-frequency plane by means of a wavelet packet binary tree. The whole plane is segmented into Heisenberg boxes of the corresponding wavelet packet basis which are time-frequency atoms.

Both techniques, the fast wavelet transform and the wavelet packet technique, have been used for finding correlates to the human subjective evaluations of the sound characteristics Urgency, Expressiveness, and Annoyance. The main question is whether the diagrams of the time-frequency analyses can clearly express those differences in the subjective evaluations which have been found in the experimental investigation.

5.2. Results

The results show some surprising agreements between the experimental findings and the computational results.

Figure 7 shows the time-frequency plane of the wavelet packet analysis with the Heisenberg boxes for the directional sound 2-3 (Left – robot sound). Figures 8 and 9 show the same time-frequency planes for the robot state sounds Heavy Load and Near Obstacle.

Comparing the two time-frequency images of the directional sounds for Left, i.e., of the marimba instrument (not shown here) and the robot sound (sound 2-3; Figure 7), a much higher frequency proportion can be found for the robot sound. Also, it is much longer persistent over time. Both effects may lead to a stronger subjective impression of Annoyance.

This can be confirmed by comparing the additional sounds for robot states and situations. The sound Near Obstacle (Figure 9) is more annoying than the sound Heavy Load (Figure 8). Consequently, a much higher frequency proportion has been analysed for the Near Obstacle sound in the time-frequency image of Figure 9 as compared with the Heavy Load sound (Figure 8).

With respect to Expressiveness, the two marimba sounds for the directions Down-Left and Left can be compared with their time-frequency images (also not shown here). The Down-Left sound has a stronger expressive appearance over time in the time-frequency image. This correlates with the subjective evalutions of expressiveness which were found in the experiments with the musicians. Their assessment of expressiveness within the indicated-direction-related features was almost proportional to the melody span.

The same result is even more pronounced when comparing the diagrams of the fast wavelet transform for the two marimba sounds. The wavelet transform diagram of the melodic sound 1-1 (Down-Left) of Figure 10 has graphically a more expressive appearance than the wavelet transform diagram of the sound 2-1 (Left) with two shorter tones followed by a longer one on the same pitch level (Figure 11). The corresponding wavelet transform diagram of the robot sound for Left (sound 2-3) shows almost an "epileptic" pattern (Figure 12) which correlates with the quite annoying but less expressive assessment in the human subjective evaluations.

6. CONCLUSIONS

The design of directional sounds and of robot sound tracks has been accomplished with basic musical elements and recorded robot noise signals. The experimental studies show that the auditory symbols and the sound tracks are recallable and understandable, at least for more musical people. Positive training effects have been observed with all human subjects. A more detailed study comparing musicians and non-musicians is under way. The comparison between time-frequency analyses and subjective evaluations of sounds will be further pursued.

The investigated digital-audio sound tracks are feasible means of communication in human-machine interaction. Continuing research also considers other application domains, such as car driving and aircraft piloting.

In addition to the applications mentioned above, one can imagine that the directional auditory displays suggested here can also be advanced for applications in art and virtual reality scenarios, for example as a hearing education tool in music or as a notation and instruction tool in dance choreography (for musification of movements).

7. REFERENCES

[1] G. Johannsen, "Auditory displays in human-machine interfaces of mobile robots for non-speech communication with humans," *J. Intelligent and Robotic Systems*, vol. 32, pp. 161-169, 2001.

[2] N.A. Stanton and J. Edworthy (Eds.), *Human Factors in Auditory Warnings*. Ashgate Publishing, Aldershot, England, 1999.

[3] S.A. Brewster, P.C. Wright, and A.D.N. Edwards, "A detailed investigation into the effectiveness of earcons," in G. Kramer (Ed.), *Auditory Display – Sonification, Audification, and Auditory Interfaces*, pp. 471-498. Addison-Wesley Publishing, Reading, MA, USA, 1994.

[4] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, USA, 1998.

[5] J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. Scargle, *About WaveLab*. Version 0.8 http://www-stat.stanford.edu/~wavelab/ftp/AboutWaveLab.pdf, 1999.

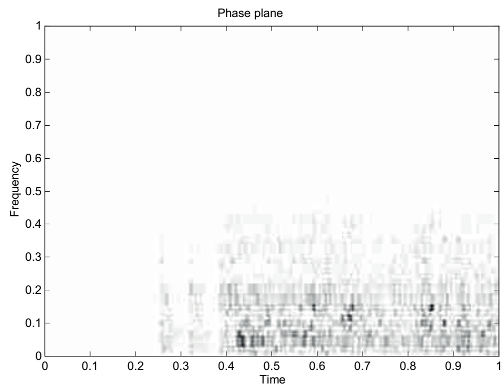

Figure 7. *Time-frequency image of sound 2-3 (Left – robot sound).*

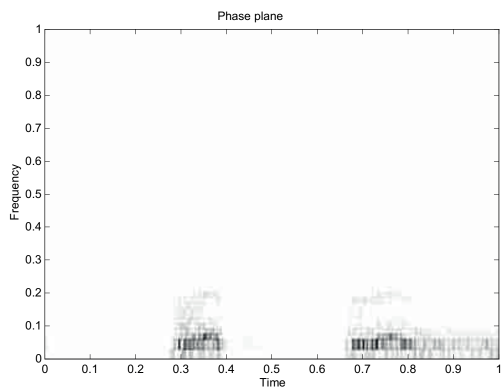

Figure 8. *Time-frequency image of sound Heavy Load.*

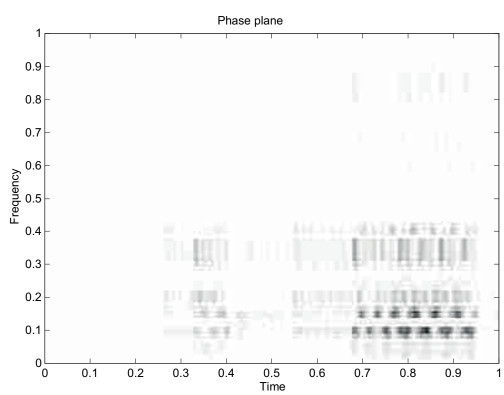

Figure 9. *Time-frequency image of sound Near Obstacle*.

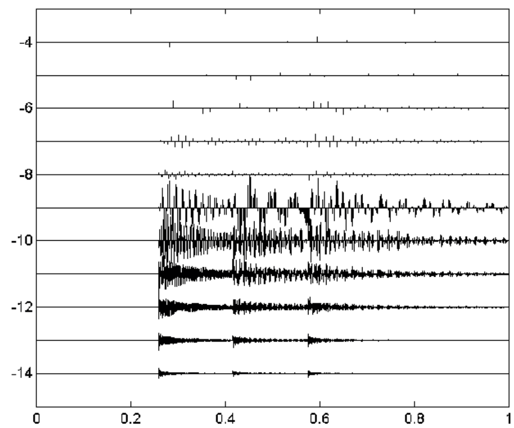

Figure 10. *Wavelet transform diagram of sound 1-1 (Down-Left – marimba).*

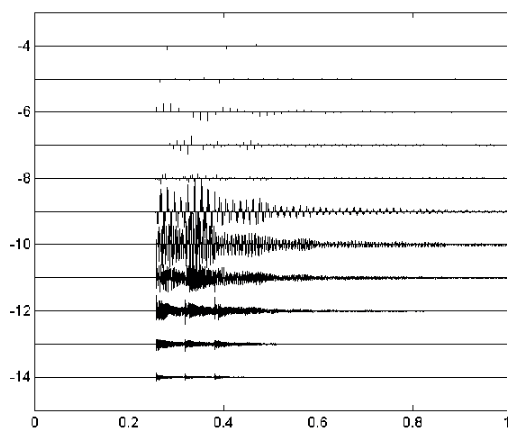

Figure 11. *Wavelet transform diagram of sound 2-1 (Left – marimba).*

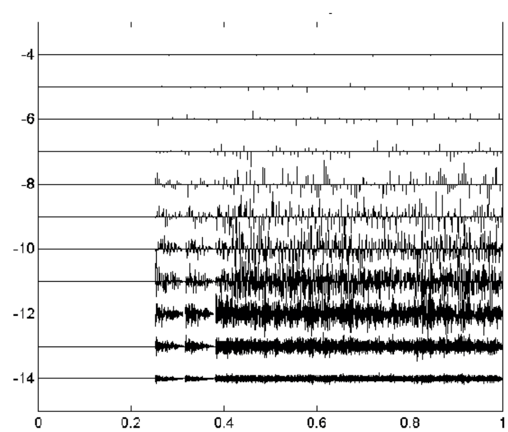

Figure 12. *Wavelet transform diagram of sound 2-3 (Left – robot sound).*