Audio Engineering Society

Convention Paper 8912

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

Mobile Audio Measurements Platform: Towards Audio Semantic Intelligence into Ubiquitous Computing Environments

Vrysis Lazaros[1], Dimoulas Charalampos[2], Kalliris George[3] and Papanikolaou George[4]

[1] Aristotle University of Thessaloniki, Dipl. Eng., Phd researcher, Thessaloniki, 54124, Greece
lvrysis@auth.gr

[2] Aristotle University of Thessaloniki, Dr. Eng., Lecturer, Thessaloniki, 54124, Greece
babis@eng.auth.gr

[3] Aristotle University of Thessaloniki, Dr. Eng., Assist. Prof., Thessaloniki, 54124, Greece
gkal@jour.auth.gr

[4] Aristotle University of Thessaloniki, Dr. Eng., Prof., Thessaloniki, 54124, Greece
pap@eng.auth.gr

Abstract

The current paper presents the implementation of a mobile software environment that provides a suite of professional-grade audio and acoustic analysis tools for smartphones and tablets. The suite includes sound level monitoring, real-time time-frequency analysis, reverberation time and impulse response measurements, whereas feature-based intelligent content analysis is deployed in terms of long-term audio events detection and segmentation. The paper investigates the implementation of a flexible and user-friendly environment, which can be easily used by non-specialists, providing professional functionality and fidelity of specific-purpose devices, and eliminating the mobile-interfacing and hardware limitations. Emphasis is given to the integration of additional capabilities that will offer valuable amenities to the user, having to do with the management of measurement sessions and intelligent cloud-based semantic analysis.

1. INTRODUCTION

Recent developments of microelectronics have brought enough processing power and hardware capabilities to mobile devices that make them comparable to personal computers. This unleashes the potential of mobile software, allowing heavy-duty tasks, such as semantic audio analysis, to be processed in real time. Among others, mobile audio sensors can be exploited along with their corresponding software control and signal processing functions in sophisticated audio analysis applications, including the implementation of mobile sound level meters. Recently, a related mobile software application has been successfully been implemented [1] [2]. Further feature expansions, comparisons with professional sound level meters and in real world sound-level monitoring scenarios were conducted, whereas remarkable evaluation-performance and flexibility were observed.

The current paper also presents application upgrades towards automated sound analysis and measurements management, including new features that aim at the direction of semantic audio concepts. The presented toolbox has been developed using the iOS Software Development Kit (SDK) [3] and has been primarily materialized on iPhone, iPod Touch and iPad terminals. However, the proposed framework can be easily adapted to a variety of smartphones and mobile terminals containing reliable audiovisual sensors that can be suitably controlled with the use of the corresponding SDKs.

Mobile application stores provide similar software, such as the "Audio Tools" [4] and "Audio Tool" [5] applications, available on the iOS and Android store respectively, but the majority of them are entertainment-oriented, lacking in terms of scientific documentation, measuring accuracy and advanced data management. This software integrates a suite of professional-grade acoustic analysis tools, including sound level metering (SLM), real-time analysis (RTA), reverberation time (RT) and impulse response (IR) estimations, providing similar functionality and fidelity to professional devices, and eliminating the hardware and the mobile-interface limitations. Joint time-frequency analysis and semantic audio-pattern markup tools are employed in long term analysis mode (LTA), in order to extract level-statistics and visualizations. Additionally, an advanced measurements data-handling interface has been developed introducing cloud-based management and analysis concepts.

Figure 1 SLM Module

2. OPERATIONAL FEATURES

The platform is divided into five main modules as Table 1 shows. Each one of them is capable of carrying out specific measurements.

Mode	Description
SLM	Sound Level Meter
RTA	Real Time Analyzer
RT	Reverberation Time Meter
IR	Impulse Response Analyzer
LTA	Long Term Analysis

Table 1 Measurement modes

2.1. Sound level meter

Sound level meter (SLM) (Figure 1) module provides a faithful reproduction of a classic digital SPL meter. The Graphical User Interface (GUI) has two main labels that show the current and peak SPL values in addition to a digital volume unit (VU) meter. Max, min, average SPL, session time and duration are also displayed. The user can select the decay time of the metering (Table 2) and the filter being used (A, B, C, D and Z) [8][6].

2.2. Real time analyzer

Real time analyzer (RTA) module offers graphical and numerical representation of the SPL measurement for each octave band. The user can select the averaging

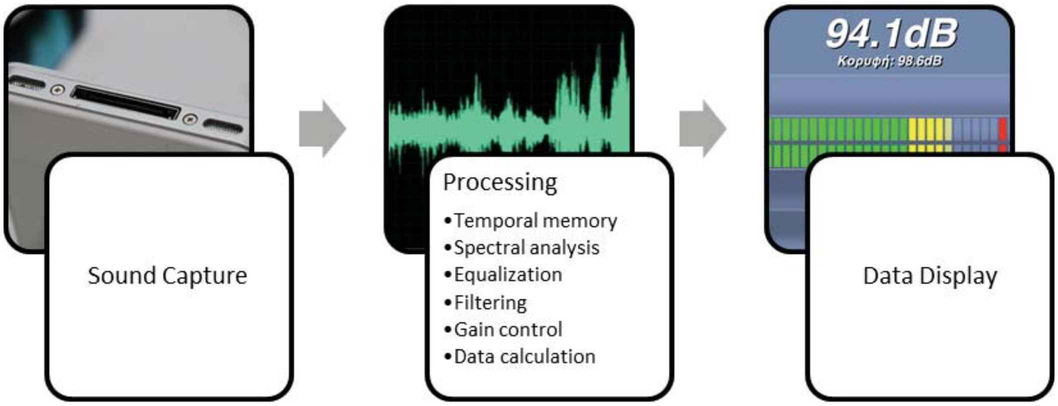

Figure 2 SLM & RTA modules processing flow

mode (slow, impulsive) and the filter being applied (A, B, C, D and Z).

2.3. Reverberation time

Reverberation time (RT) module can be used for executing reverberation time measurements. In order to operate, the user should use an external sound source by connecting it to the device's 3.5mm jack, as audio output, while the built-in microphone is used as audio input. The module offers a wizard, which helps the user to ensure that the measurement can be successful. It monitors the background noise level and the sound level generated by the application's noise playback. Reverberation time is measured for the 125Hz, 250Hz, 500Hz, 1kHz, 2kHz, 4kHz and 8kHz octave bands.

Mode	Time
Impulsive	35ms
Fast	125ms
Slow	1s

Table 2 Decay modes [6]

2.4. Impulse response

Impulse response (IR) module is used for measuring a room's impulse response. This measurement can be performed using an external impulse sound source, such as little dynamite explosion, balloon burst or gun shot. The software monitors the sound levels, captures the impulse response and saves its waveform into a file. In addition to this, it computes room acoustical parameters following the ISO 3382 standard such as ETC, EDT, T15, T30, C50, C80 and CT [7].

2.5. Long term monitoring and semantic analysis

Along with the above mentioned low-level measurement modes the software also provides long-term audio analysis capability, based on semantic audio processing concepts. This, higher-level, mode brings real time audio-pattern recognition; visually resulting into an event detection markup timeline.

A dynamic audio-samples database is used as a pattern-storing matrix, which is configurable by users. Samples can be added, by making a simple recording, and deleted as well.

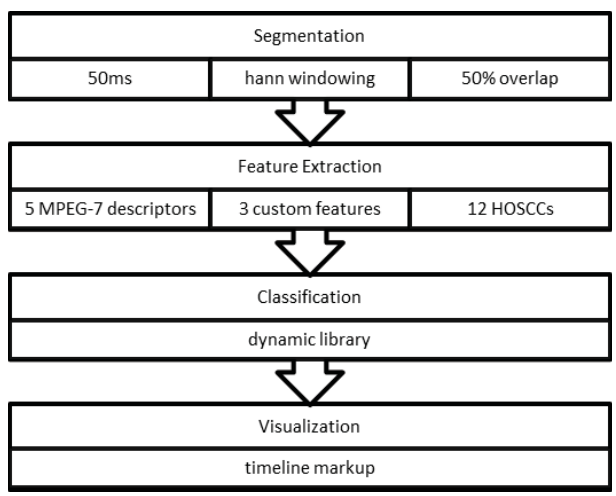

Figure 3 Semantic analysis processing flow

2.6. Session management

The application implements a user-friendly measurement session manager (Figure 4). Each measurement can be easily stored on the device's memory and recalled anytime. The application stores the session's measurement data, title, location, user's comments, while automatically determines the position utilizing the device's GPS system, captures a timestamp, and offers a handy interface in order to capture a picture of the measurement's location. Measurement's data can be easily shared via e-mail.

In addition to this, a cloud-based session manager handles all the users' data, aiming at building a user-generated, spatiotemporal digital map used for storing measurements. Users can store, update and retrieve raw audio data and its corresponding analysis output.

All measurements uploaded to the cloud are accessible by anyone who uses the application. By exploiting the device's GPS sensor and cellular data capabilities, the application can easily classify and group measurements by geographical location and kind. Thus, a user can instantly check and confirm the correctness of a specific measurement by comparing it to similar ones, provided by other users. He can even get the desired data without making a measurement.

Figure 4 Session management interface

3. DEVELOPMENT

The development of the application is designed according to the waterfall development model; the coding is platform-specific, while extended emphasis is given to performance optimizations.

The application is structurally implementing a modular design so that new modules can be easily added, and is split into two main components, the back end and the frond end. The back end performs the audio playback, recording, measurement and calibration procedures, while the front end adds the user interface and the audio session management functionality. The VU GUI is implemented using the OpenGL libraries while the code is written in C++ and Objective C languages in order to maximize performance, allowing algorithms to be processed in real time.

The back-end component utilizes the Audio Toolbox Framework provided by the iOS SDK and uses the Audio Queues Services that offer low-level data access to the input buffers of the device's audio input device. In general, the samples being captured are stored temporally using 64-bit floating-point variables; the signal is being processed spectrally, calibrated and filtered. Frequency analysis is performed using 2048 and 4096-point RDFT-FFT algorithm, calibration is applied in order to cancel the non-linearity of the device's audio interface and filtering is performed by applying functions that are defined by the A/B/C/D weighting filters [6].

SLM & RTA modules implement the processing flow described in Figure 2; spectral power calculations are done according to the Welch's method [10]. Windowing is performed using the Hann function, considering parameters such as spectral leakage and processing gain [11]. Window overlapping is set to 50%.

RT performs reverberation time measurements employing the interrupted noise method [12]. The application utilizes a band-pass pink noise generator in order to produce portions of octave-band filtered pink noise with duration of several seconds, which end abruptly. The response of the room to this noise signal, i.e. the reverberation, is recorded and the sound level is monitored. The module calculates the RT30, while RT60 is being estimated using linear interpolation [9].

IR module captures and analyzes impulse response measurements. The impulse response is recorded into a wav file and asynchronously the room acoustic measures are calculated [14] based on Schroeder's curve [13].

LTA module processing flow is based on segmentation, feature extraction and classification,

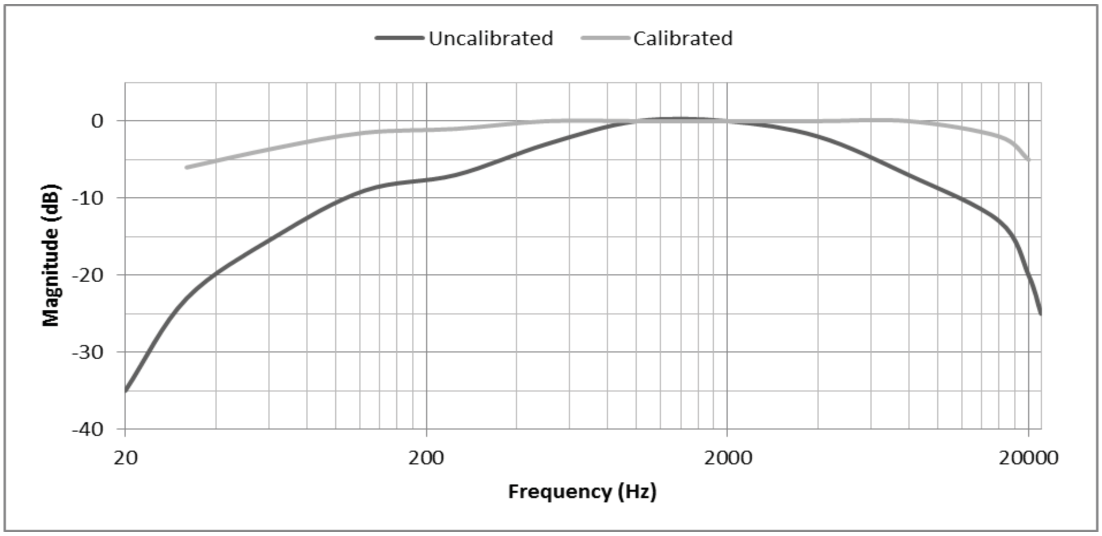

Figure 5 iPhone's audio input interface frequency response

as Figure 3 presents. Taking into consideration performance issues, two sets of features have been deployed. First feature set includes five MPEG-7 low-level descriptors and three more custom features (Table 4), while the second consists of twelve Hybrid Octave Scale Frequency Cepstral Coefficients (HOSCCs, Table 3), which act in the same way as Mel Frequency Cepstral Coefficients (MFCCs) do, reducing computational overhead by utilizing already calculated values [17].

For the derivation of the HOSCCs, 12 filter bank log-energies were used. The method includes the following steps. First, a 2048-point FFT transform was computed for every frame, its outcome was filtered by the hybrid filterbank, the output data was spaced adequately and finally was decorrelated by a discrete cosine transform (DCT) [18].

Band #	Frequency Range (Hz)
1	50-70
2	70-110
3	110-150
4	150-190
5	190-270
6	270-350
7	350-710
8	710-1420
9	1420-2840
10	2840-5680
11	5680-11360
12	11360-22.050

Table 3 Hybrid filterbank bands

MPEG-7 Descriptors
Audio Power
Audio Spectrum Centroid
Bandwidth
Audio Spectral Spread
Spectral Flatness Measure
Custom Audio Features
A/B/C/D Filtered Loudness
A/B/C/D Filtered Octave Band Levels
Crest Factor Level

Table 4 Audio features

4. CALIBRATION

An important aspect for improving the accuracy of the application's measurements is the calibration of the device's audio input. The iPhone's audio input interface provides a non-flat frequency response, as expected, because is mainly intended for voice capturing (it looks like that some kind, similar to A-Weighting, of filtering is applied – Figure 5).

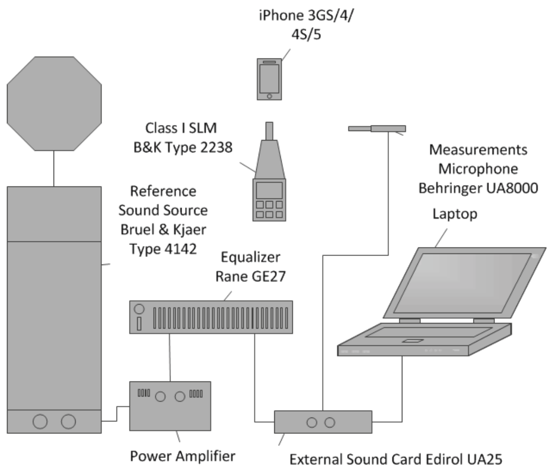

Figure 6 Calibration scheme

For measuring purposes, a flat frequency response is indispensable, so, first of all, an equalization process was held as Figure 6 shows. White noise was produced by a pre-calibrated reference sound source into an

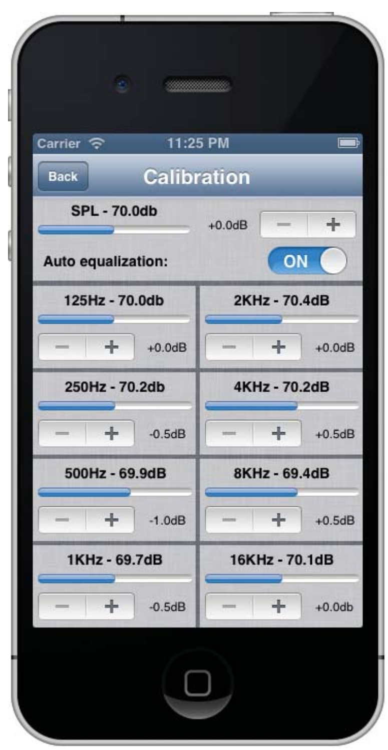

Figure 7 Manual calibration interface

anechoic chamber, and thus, the gain for each FFT bin was adjusted in order to flatten the frequency response curve. After equalization, the overall gain was corrected by comparing the application's reading to a class 1 Sound Level Meter. Both devices placed in one-meter distance from the source, while the output of the loudspeaker was a 1kHz tone at 80dB SPL(A).

The application is properly pre-calibrated for the iPhone (3GS, 4, 4S and 5), the iPad (1st, 2nd, 3rd and 4th generation) and iPod touch (4th and 5th generation) devices. However, if the user wishes to use an external microphone, the software provides the appropriate interface for manual calibration, which includes 10 octave-band and overall gain controls (Figure 7).

5. CONCLUSIONS

The toolbox meets its design specifications implementing a robust environment for modern audio measurement concepts. It provides a new form for sound analysis and data acquisition that can facilitate similar works and research.

Preliminary results show that software can be used by non-experts, while the usage adaptation time is short. Mobile audio platform can be used for instant measurements prior to advanced ones.

Audio recording interface delivers an improved, in comparison to stock, frequency response (+/-2dB @40-18kHz) while the dynamic range is 60dB, measuring from 35 to 95dB SPL(A).

New concepts, like spatiotemporal measurements mapping combined with row audio input and its semantic output, cloud based measurement storage, and global sharing options have been introduced.

Semantic analysis starting from spectral and power based event recognition for industrial measurements can lead to the development of audio quality and speech eligibility metrics for room acoustics.

6. FUTURE WORK

Future software enhancements will focus on improving measurement functionality and expanding cloud-based analyzing capabilities.

Further evaluation on the accuracy of the measuring processes, a research in finding a suitable, in terms of size and performance, external microphone, that will upgrade the audio recording quality are scheduled. Measuring methods will be enriched, bringing more options for the IR module [15], and specifically the MLS and Sine Sweep methods [19], [20].

Advanced measurement session management and cloud-based services will be expanded providing more sophisticated long-term audio event detection, summarization and extensive semantics processing [21]. Intelligent annotation and management of the provided multi-modal measurement sessions can be deployed, thus mapping the relationships between the involved audio levels-features with the audiovisual meta-information and the semantic tags. In this context, photos and video captures can be utilized along with speech-comments for icons and ear-cons meta-data extraction, facilitating content-based measurement management.

Ground-truth knowledge-basis can be gradually constructed and validated, materializing user semantic interaction, along with personalization and pervasive computing services.

7. REFERENCES

[1] L. Vrysis, C. Dimoulas, G. Kalliris, Papanikolaou G., "SPL Meter: Sound Pressure Level Meter software development on a smartphone (iPhone)", Helina 6th National Conference Acoustics 2012, Kerkyra, Greece, 2012

[2] L. Vrysis, F. Katsimbas, "SPL MASTER: Sound Pressure Measurement Software for the iPhone" Audio capturing, editing and managing methods, Interdisciplinary graduate program "Advanced Computer and Communication Systems", Thessaloniki, 2011.

[3] Apple «iOS Developer Library», https://developer. apple.com/library/ios/ (2013, March)

[4] Studio Six Digital "Audio Tools", http://www. studiosixdigital.com/audiotools/ (2013, March)

[5] Bofinit Corporation "Audio Tool", http://sites. google.com/site/bofinit/ (2013, March).

[6] ANSI Standards S1.4-1983 and S1.42-2001

[7] ISO 3382 - 1:2009, Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces.

[8] G. Papanikolaou, "Electroacoustics", University Studio Press (2005)

[9] G. Papanikolaou, "Electroacoustic Measurements", University Studio Press (2008)

[10] P.D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Transactions on Audio Electroacoustics, pp. 70–73 (1967).

[11] F.J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform", IEEE Proceedings, vol. 66 (1), pp. 51-83 (1978).

[12] K. Jambrosic, M. Horvat, H. Domitrovic, "Reverbaration time measuring methods", SFA Acoustics 2008, Paris, France, 2008

[13] M. Schroeder, D. Rossing, W.M. Hartmann, D.M. Campbell, N.H. Fletcher, Springer Handbook of Acoustics, Springer (2007)

[14] E. Nilsson, "Room Acoustic Measures for Classrooms", Internoise 2010, Lisbon, Portugal (2010)

[15] G. Stan, J. Embrechts, D. Archambeau, "Comparison of different impulse response measurement techniques", Journal of the AES, vol. 50 (4), pp. 249-262 (2002).

[16] S. Ntalampiras, I. Potamitis, N. Fotakakis, "Acoustic Detection of Human Activities in Natural Environments", JAES, vol. 60 (9), pp. 686-695 (2012, September)

[17] C. Dimoulas, G. Kalliris, "Investigation ! f wavelet approaches for joint temporal, spectral and cepstral features in audio semantics", 134th AES Convention, Rome, 3-7 May, 2013.

[18] T. Ganchev, "Contemporary Methods for Speech Parameterization", Springer (2011)

[19] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique", AES 108th Convention, n. 5093 (2000).

[20] A. Farina, "Advancements in impulse response measurements by sine sweeps", AES 122nd Convention, n. 7121 (2007).

[21] F. Rumsey, "Semantic Audio: Machines Get Clever with Music", JAES, vol. 59 (11), pp.882-887 (2011, November)