A Smartphone-based Digital Hearing Aid to Mitigate Hearing Loss at Specific Frequencies

Wei Wang, Zhilu Chen, Baoyuan Xing
Electrical and Computer Engr Dept
Worcester Polytechnic Institute
weiwang,zchen2,bxing@wpi.edu

Xiaochen Huang, S Han, Emmanuel Agu
Computer Science Dept
Worcester Polytechnic Institute
xhuang3, shan, emmanuel@cs.wpi.edu

Abstract

Hearing Loss is one of the three most common chronic conditions among the elderly. In many cases, an individuals hearing is only impaired at certain (not all) frequencies. Analog hearing aids boost all sound frequencies equally including frequencies in which the individuals hearing is good, causing discomfort to the user. Digital hearing aids can amplify only the specific frequencies at which a persons hearing is impaired. In this paper, we describe the design, implementation and evaluation of a smartphone digital hearing aid app. Our digital hearing aid implementation has two parts: speech processing in the frequency domain and sound classification. We used Weighted Over-Lap Add (WOLA) filter bank to decompose microphone sounds into different frequency bands that are then amplified in the frequency domain. Mel-frequency cepstral coefficients (MFCC) of input sounds are computed and used as features for sound classification by the Gaussian Mixture Model (GMM) machine learning model. Our digital hearing aid app amplifies select frequency bands and correctly classifies speech in quiet and noisy environments. The results of a small user evaluation of our prototype are also promising.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design

Keywords

Digital hearing aids, smartphone, sound classification

1 Introduction

Thirty five million (or 11%) americans were hearing-impaired by the year 2008 [11]. However, due to several factors including cost, only 25% of people with hearing loss currently use hearing aids. Due to custom Digital Signal Processing (DSP) chips required to process sounds in real time, the average price of a hearing aid is $1,601.

A hearing-impaired patient is frequently unable to hear sounds at certain frequencies. Analog hearing aids amplify sounds at all frequencies uniformly, which improves the user's hearing at the inaudible frequencies but makes audible frequencies too loud. Digital hearing aids can apply different amplification (or reduction) factors to different frequency bands. For instance, a person with sensitivity to high-pitched sounds may configure their digital hearing aids to compress high frequencies but not other frequencies.

Digital hearing aids have many desirable features that enhance the intelligibility of speech [10]. *Noise reduction* enables human speech to be heard clearly in noisy environments by reducing noise in specific frequency bands while preserving overall sound quality. *Programmability* enables the hearing aid to be customized to accommodate user-specific impairments, and to be reprogrammed if the user's hearing changes over time. *Feedback reduction* ensures that already amplified sounds emanating from the receiver are not mistakenly amplified further, causing the user some discomfort. *Directional microphones* preserve sounds from the speaker's forward direction (facing listeners) while suppressing sounds from other directions.

In this paper, we present a software-only design and implementation of a digital hearing aid app on Google's Android platform. A digital hearing aid app is accessible at low cost to economically disadvantaged populations worldwide. A hearing-impaired user running our app simply connects their earphones to their Android phone, and is able to listen to environmental sounds, live conversations, phone calls and music all amplified to compensate for hearing loss in specific frequencies. While Android and iPhone hearing aid apps such as SoundAmp R [15] and HearYouNow [8] have begun to emerge, their inner workings have not been published. To the best of our knowledge, this is the first exposition and of a hearing aid app.

2 Hearing Aid App Architecture

The architecture of our hearing aid app is shown in figure 1. The digital hearing aid system is split into two parts: *speech processing* in the frequency domain and *sound classification* to classify input sounds into speech and speech with noise categories. The acoustic signal is read in by the micro-

1st Workshop on Mobile Medical Applications14, November 6, 2014, Memphis, TN, USA.

http://dx.doi.org/10.1145/2676431.2676435

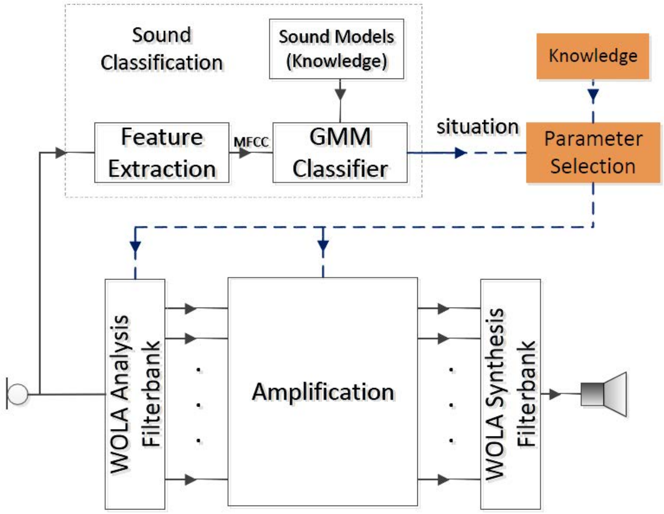

Figure 1. Architecture of our hearing aid app

phone. WOLA filter banks then split the sound up into different frequency bands, which are then amplified (reduced) by the amplification in the specific frequency ranges at which the user's hearing is impaired. Finally, the WOLA synthesis filter bank reconstructs the acoustic signal from the amplified sub-band signals, which is sent to the receiver for play out.

Hearing aids users usually encounter different sound environments each day such as conversations in quiet or noisy environments, in music theaters or in road traffic noise. A digital hearing aid classifies the user's listening environment into different listening situations to which different adaptations are employed. For instance, if the digital hearing aid's sound classifier detects noise in the sound, it applies a noise reduction algorithm to reduce the noise. In our system, 12-dimension Mel-Frequency Cepstral Components (MFCCs) of input sounds are computed and used as features of the acoustic signal. Then a fast GMM-based model classifies two listening situations: clean speech and speech with noise. In our current prototype, noise reduction and the configuration unit for programming the hearing aid with different parameters in different situations, has been left as future work.

3 Speech Processing in Frequency Domain

Speech processing in the frequency domain decomposes the acoustic signal into different frequency sub-bands that are then amplified with different gains. Digital filter bank systems are widely used for frequency-selective signal decomposition in modern hearing instruments. Signals are processed in decimated sub-band signals for energy-efficiency [7, 9]. Complex-modulated (DFT) lter banks are highly efficient decimating lter bank system [6, 20]. The DFT filter bank uses a polyphase implementation of the Finite Impulse Response (FIR) lters (Figure 2).

The Weighted Over-Lap Add (WOLA) filter bank is a low-power, low delay DFT structure [20], widely used in digital hearing aids. The WOLA analysis filter bank splits input microphone sounds into 8 sub-band signals. Finally, the WOLA synthesis filter bank is used to reconstruct the amplified acoustic signals from sub-band frequencies.

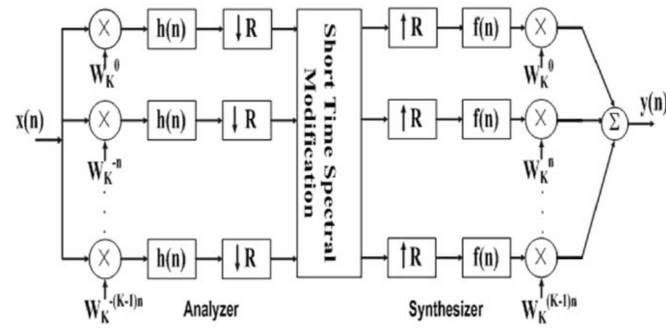

Figure 2. Complex modulator model of the DFT filter bank

4 Sound Classification

Sound classification has been widely researched in different applications, such as speech/music discrimination [12], and modeling sound events on mobile phones [13]. These applications use classifiers such as decision trees in [12], which are too complex for hearing aids. A simpler classification scheme may be used in hearing aids, since it only needs to identify two contexts - speech in quiet and speech in noise. MFCC features have become very popular for sound classification in digital hearing aids [4, 1]. Many applications use the GMM classifier to achieve good sound recognition accuracy [21]. We utilize 12-dimension MFCCs as features in the GMM classifier for sound classification in our digital hearing aid app.

4.1 Front-End Processing

The audio stream from microphone is segmented into frames of uniform duration. Features for classification are extracted while processing each individual frame. We use independent non-overlapping frames of 8 ms. Our sampling rate is 16000, so there are 128 samples in one frame from which 12-dimension MFCCs are extracted. Figure 3 is the flowchart of the MFCC.

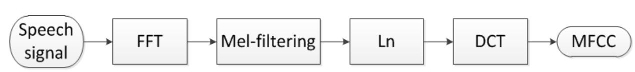

Figure 3. MFCC Flow Diagram

The frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response very closely [5]. The MFCC maps frequency space to mel-space. The mel-filters are evenly distributed in mel-space, focusing the analysis more on low-frequency bands. Fig. 4 shows the mappings of mel-space to frequency space.

4.2 Fast Gaussian Mixture Model

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of different Gaussian distributions [17]. We use 8 Gaussian components, each with a mean, variance and weighting. M is the number of Gaussian components. The Gaussian Mixture

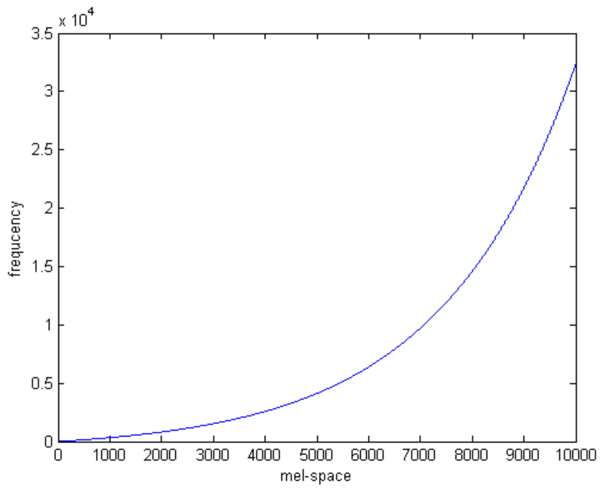

Figure 4. Mel-space to frequency space mappings

Density (GMD) for a D-dimensional random vector can be calculated as [17]:.

$$p(\mathbf{x}/\lambda) = \sum_{i=1}^{M} C_i b_i(\mathbf{x}) \ (\sum_{i=1}^{M} C_i = 1) \tag{1}$$

Where x is a D-dimensional random vector, b(x) is the component density, and C_i is the mixture weight. Each component density is a D-dimensional Gaussian function of the form

$$b(\mathbf{x}) = \frac{1}{(2\pi)^{D/2} \mid \mathbf{\Sigma_i} \mid} e^{-\frac{1}{2}(\mathbf{x}-\mu_\mathbf{i})'\Sigma_i^{-1}(\mathbf{x}-\mu_\mathbf{i})} \tag{2}$$

with mean vector μ_i and covariance matrix Σ_i. The complexity of exponential calculations is very high. To reduce complexity, the Gaussian Mixture Density (GMD) is calculated in the log domain, thus avoiding the complex exponential calculation. The GMD calculation has two steps: Mahalanobis distance calculation and exponential-logarithm calculation. In log form, the GMD is expressed as

$$\tilde{p}(\mathbf{x}/\lambda) = log \sum_{i=1}^{M} C_i b_i(\mathbf{x}) \tag{3}$$

$$= log \sum \{exp[-\frac{1}{2}(\mathbf{x} - \mu_\mathbf{i})'\Sigma_i^{-1}(\mathbf{x} - \mu_\mathbf{i}) \tag{4}$$

$$+ log \frac{C_i}{(2\pi)^D \mid \mathbf{\Sigma_i} \mid}]\} \tag{5}$$

The first part in equation 4 is the Mahalanobis distance calculation which is shown in equation 6. The second part in equation4 is called the exponential-logarithm

The exponential-logarithm calculation can be derived through logadd calculation defined as shown in 8, which

$$-\frac{1}{2}(\mathbf{x} - \mu_\mathbf{i})'\Sigma_i^{-1}(\mathbf{x} - \mu_\mathbf{i}) \tag{6}$$

$$log\{\sum exp(a_i)\} \tag{7}$$

could be calculated by Taylor Expansion calculation, which is shown in equation 7.

$$\begin{aligned} LogAdd(a, b) &= log(e^a + e^b) \\ &= a + log(1 + e^{b-a}) \end{aligned} \tag{8}$$

5 Hearing Aid App Implementation

The hearing aids app was developed using the Android Software Development Kit (SDK). We tested our app on a Samsung Galaxy S2 smartphone running Android operating system version 2.3.3. The entire app including the GUI, speech processing and sound classification was written in Java.

During frequency domain speech processing, PCM formatted data is placed in a circular queue buffer. Once full, the buffer is input to the WOLA unit 4 samples at a time. After processing one set of inputs, the WOLA unit outputs 4 samples to the output buffer queue, which are then played out using the receiver as shown in Figure 5.

The software components for sound classification and the interactions between them are shown in Figure 6. An environmental model training system is pre-built based on extracted features. In essence, two models are trained using the Gaussian Mixture Model with 8 mixture number using 12-dimension MFCC features calculated from raw data. The training data is recorded using a 16000 Hz sampling rate, 16-bit mono audio samples from the smartphone microphone. The app's sound classification module exploits the pre-built models and is able to classify the environment when users are speaking. When launching the app for sound classification, the microphone on the smartphone records input sound for 4 seconds using a 16000 Hz sampling rate of 16-bit mono audio samples. Each frame consists of 128 samples and 12th order MFCC for each frame is computed. The classifier identifies the probability of each frame. After accumulating 500 frames, the classifier displays the decision on the User Interface. Screenshots of the app is shown in Figure 8.

6 Evaluation

6.1 Simulation of the WOLA algorithm

For validation purposes, we simulate the WOLA algorithm using MATLAB. As shown in Figure 7, applying different gains to different channels enlarges the sound in different frequencies.

6.2 Evaluation of MFCC Features and GMM-Based Classifiers

The evaluation of the MFCCs as features and GMM-based classifiers is performed on a database that consists of two classes of sounds: clean speech and speech with noise.

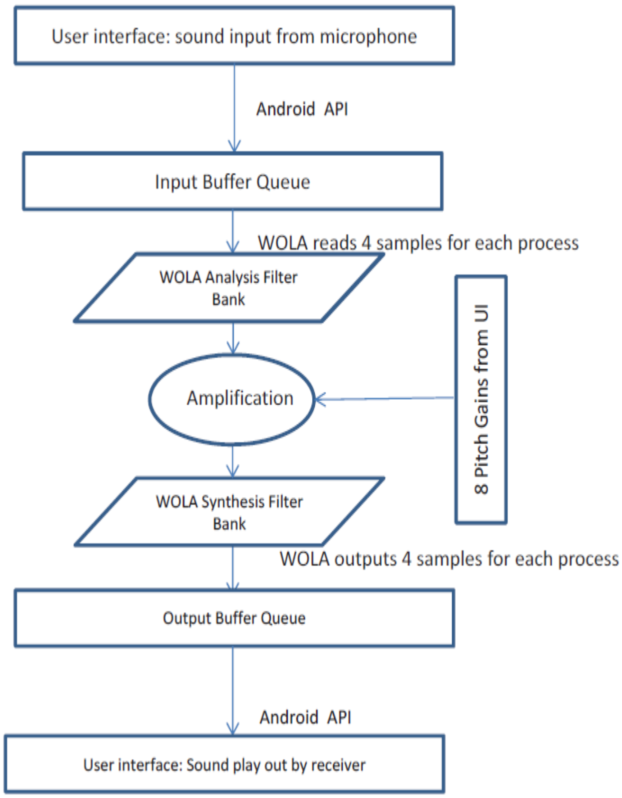

Figure 5. Implementation of Speech Processing in the Frequency Domain

The clean speech comprises of speech spoken by different people in different situations, such as a living room, a cafeteria or making a speech. The noisy speech is generated by randomly mixing selected files from the clean speech library with different noises. We use 4 minutes speech for training and 4 seconds of speech for testing. We randomly select 30 different speech samples for testing. This tests resulted in 29 correct testing sets (out of 30) yielding a recognition rate of 96.7%.

6.3 User Studies

6.3.1 Speech Processing in Frequency Domain

Ten graduate students (6 males and 4 females) at Worcester Polytechnic Institute (WPI) participated in our user study to evaluate our hearing aid app. We tested our app with subjects with no hearing impairments since access to deaf subjects requires full IRB approval at a hospital. We shall do so in future. Even subjects without impairments can detect amplification (or compression) of different sound frequencies. Each subject listened to the sound generated from the app for about 2 minutes and then listened to the original sound, where the frequency spectrum of each sound is amplified in three different sub-bands. All subjects reported that they could detect frequency changes between the original sound and the processed sound. They also felt that the apps signal processing calculations were a little slow and that the app consumed a lot of power.

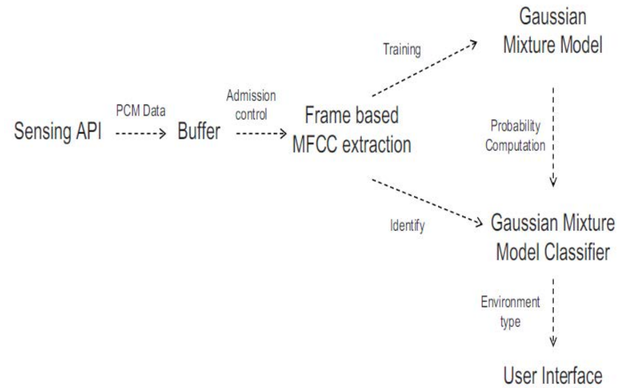

Figure 6. Implementation of Sound Classification

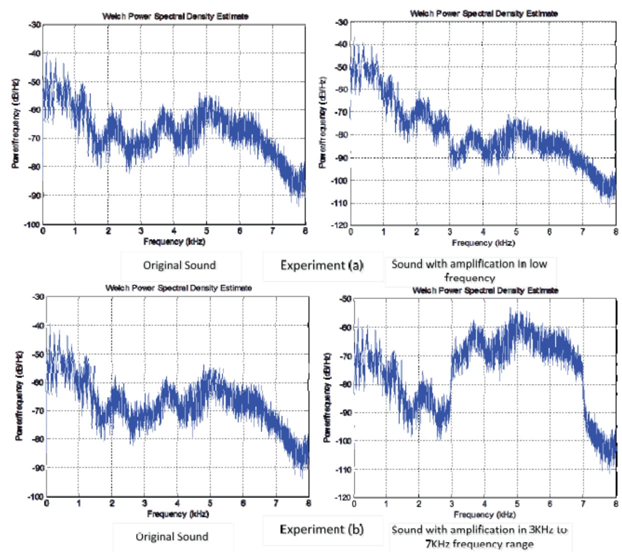

Figure 7. WOLA Simulation Results

6.3.2 Sound Classification

Our hearing aid app can also detect environmental noise. We classify environments into two categories: quiet environment, and noisy environment. Sixteen subjects installed and used our app in different scenes. The first test scene was a conference room (no people) to illustrate scenes with very little noise (about 50 decibels). Pairs of subjects talked to each other while the app was turned on in order to capture the sound. After recording for 10 seconds, the app accurately determined that the input sound signal was ”voice in quiet environment” on 14 of the 16 phones, while the other two phones incorrectly determined that the input signal was ”voice in noisy environment”. A second set of tests were done in a noisy outdoor plaza with a fountain, music band playing and students chatting (¿ 75 decibels). The app on all 16 cell phones correctly determined that this was ”voice in noisy environment”.

Several factors influence the accuracy of sound classification including the distance between the microphone and the

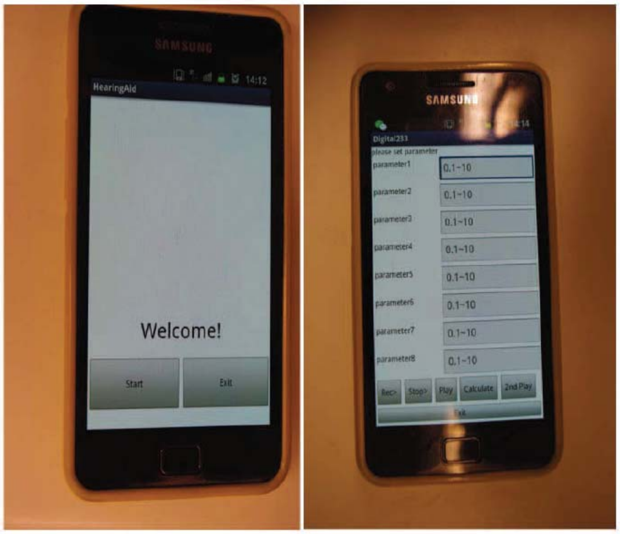

Figure 8. App Screenshots

speech/noise source, and the direction which the microphone is facing relative to the source. Voices from other humans may also be regarded as noise, leading to misclassification of the actual sound environment. In future work, we plan to investigate some of these factors.

7 Related Work

Hearing aids: have been researched and manufactured for over 25 years [7, 10]. Analog hearing aids [18] and custom digital hearing aids [10] have been proposed.

Speech processing: prior work exists in speech processing in noisy environments [9], segmentation of audio signals into speech or music [12], energy-efficient implementations of speech classification algorithms [4] and speech production from MFCCs [19].

Sound classification in ubiquitous computing: has emerged due to the popularity of smartphones. Systems for scalable large scale classification of audio environments have been proposed [13, 14]. Proposed techniques can identify speakers [13], use sounds to classify parts of a city [3], infer conversation episodes within an audio stream [16] and detect the stress of speakers from their speech [2].

8 Conclusions and Future Work

In this paper, we introduced our digital hearing aids system on Android smartphones. We describe the WOLA filter bank used in our system to decompose input sounds into frequency sub-bands, permitting the amplification of specific frequencies in which a user's hearing is impaired. We successfully implement sound classification on Android phones using MFCC sound features and the GMM classification algorithm. The environmental sound recognition results are accurate both in quiet and noisy environments. In future, using our sound classification module, the hearing aid app will automatically detect environmental noise levels and configure the hearing aid with appropriate parameters.

Our frequency domain processing is currently a bit slow due computational intensity of the algorithms implemented. To improve performance, we shall investigate fixed point arithmetic for our calculations and using the smartphone's GPU to accelerate parallelizable algorithms. Algorithms such as WOLA filter bank lend themselves to parallelization. We shall also investigate efficient noise reduction algorithms. Finally, we intend to collaborate with hospitals to evaluate our app with hearing impaired patients. The overall goal of such tests will be to establish the utility, accessibility and acceptance of the hearing aid app for real patients. Since hearing impaired patients are considered vulnerable populations for the purposes of Institution Review Board (IRB) approval, great care will be taken to handle such tests correctly.

9 References

[1] E. Alexandre, R. Gil, L. Cuadra, L. Alvarez, and M. Rosa. Speech/music/noise classification in hearing aids using a two-layer classification system with mse linear discriminants. In *Proceedings of European Signal Processing Conference,*. Switzerland, 2008.

[2] K.-H. Chang, D. Fisher, J. Canny, and B. Hartmann. Hows my mood and stress?: an efficient speech analysis library for unobtrusive monitoring on mobile homes. In *Proceedings BodyNETs*, page 7177, 2011.

[3] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao. Automatically characterizing places with opportunistic crowdsensing using smartphones. In *Proceedings ACM Ubicomp*, page 481490, 2012.

[4] R. Dong, D. Hermann, E. Cornu, and E. Chau. Low-power implementation of an hmm-based sound environment classification algorithm for hearing aid application. In *Proceedings EUSIPCO*, 2007.

[5] M. frequency cepstrum. http:/en.wikipedia.orgwiki/mel-frequency_cepstrum.

[6] H. G. Gckler and A. Groth. *Multiratensysteme*. Schlembach-Fachverl, 2004.

[7] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass. Signal processing in high-end hearing aids: state of the art, challenges, and future trends. *EURASIP Journal on Applied Signal Processing*, page 2915.

[8] HearYouNow. https:/itunes.apple.comusapphearyounow-your-personal-soundid569522474?mt=8.

[9] E. Hnsler and G. Schmidt. *Speech and audio processing in adverse environments*. Springer, 2008.

[10] J. M. Kates. *Digital hearing aids.* Cambridge University Press., 2008.

[11] S. Kochkin. Marketrak viii: 25-year trends in the hearing health market. *Hearing Review*, 16(11):1231, 2009.

[12] Y. Lavner and D. Ruinskiy. A decision-tree based algorithm for speech/music classification and segmentation. *EURASIP Journal on Audio, Speech, and Music Processing*, 2, 2009.

[13] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *Proceedings MobiSys*, page 165178. ACM, 2009.

[14] H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings ACM Sensys*, page 7184, 2010.

[15] S. R. https:/itunes.apple.comusappsoundamp-rid318126109?mt=8.

[16] M. M. Rahman, A. A. Ali, K. Plarre, M. Alabsi, E. Ertin, and S. Kumar. mconverse: Inferring conversation episodes from respiratory measurements collected in the field. In *Proc Wireless Health*, 2011.

[17] D. Reynolds. *Gaussian mixture models*. Encyclopedia of Biometric Recognition 2, 17.36, 1468, Reading, Massachusetts, 2008.

[18] R. E. Sandlin. *Textbook of Hearing Aid Amplification 2e*. Singular Publishing Group.

[19] Z. Tychtl and J. Psutka. Speech production based on the mel-frequency cepstral coefficients. In *Proceedings EUROSPEECH1999,, Budapest, Hungary*. ISCA, 1999.

[20] P. P. Vaidyanathan. *Multirate systems and filter banks.* Pearson Education India, 1993.

[21] J.-J. Xiang, M. F. Mckinney, K. Fitz, and T. Zhang. Evaluation of sound classification algorithms for hearing aid applications. In *Proceedings IEEE ICASSP*, page 185188, 2010.