Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2016 August ; 2016: 85–88. doi:10.1109/EMBC.2016.7590646.

Smartphone-Based Noise Adaptive Speech Enhancement for Hearing Aid Applications

Issa Panahi [IEEE Senior Member], Nasser Kehtarnavaz [IEEE Fellow], Linda Thibodeau

Issa Panahi, Nasser Kehtarnavaz and Linda Thibodeau are with the University of Texas at Dallas, Richardson, TX 75080, USA;

Abstract

It is well established that the presence of environmental noises has a negative impact on the performance of hearing aid devices. This paper addresses a noise adaptive speech enhancement solution for the purpose of improving the performance of hearing aid devices in noisy environments. Depending on three noise types of babble, machinery, and driving car, the parameters of a recently developed speech enhancement algorithm are appropriately adjusted to gain improved speech understanding performance in noisy environments. This solution is implemented on smartphone platforms as an app and interfaced with a hearing aid device. A clinical testing protocol is devised to evaluate the performance of the app in participants with normal hearing and hearing impairments. The clinical testing results have indicated that statistically significant improvement in speech understanding is gained between the unprocessed and processed conditions using the developed noise adaptive speech enhancement solution.

I. Introduction

According to the World Health Organization, more than 5% of the world's population, which accounts for 360 million people (328 million adults and 32 million children), suffer from disabling hearing loss [1]. In the US, the number of people suffering from hearing loss is estimated at 35 million [2]. It is well established that the performance of hearing devices, including hearing aids (HA), cochlear implants (CI), and sound amplifiers (SA), is adversely impacted in the presence of environmental noises. The literature offers many studies where advanced speech enhancement algorithms are developed to improve the signal or speech processing pipelines of hearing devices in the presence of environmental noises for the purpose of gaining better hearing capability by users of these devices.

Recently, attempts have been made in devising speech enhancement signal processing pipelines by adapting the speech enhancement to specific environmental noises that are identified automatically. Examples of such noise adaptive speech processing pipelines can be found in [3–5]. In commercial HAs, speech enhancement or noise reduction algorithms have been deployed to improve speech understanding in noise, but their proprietary nature limits researchers' ability to evaluate and improve these algorithms.

This paper presents a noise adaptive speech enhancement pipeline that is implemented on smartphones and is clinically tested on participants with normal hearing and those already using hearing aids. The smartphone-based solution developed in this work is devised such that it has a broad impact on the hearing aid research as it offers researchers with a widely used and portable platform to clinically evaluate signal processing algorithms for hearing aid applications.

The rest of the paper is organized as follows: Section II provides an overview of the algorithms or components involved in the noise adaptive speech enhancement pipeline implemented on smartphone platforms. The clinical setting used to test this pipeline is then described in section III followed by the experimental results and their discussion in section IV. Finally, the conclusion is stated in section V.

II. Overview of Noise-adaptive Speech Enhancement Pipeline

Figure 1 illustrates the major components or algorithms that go into a typical noise adaptive speech processing pipeline. The key attribute of this pipeline is that its components are devised to be computationally efficient so that the entire pipeline is enabled to run in real-time on modern smartphones.

A. Noise Classification

For the speech enhancement algorithm or component of the pipeline to adapt its parameters to different environmental noises, it is necessary to identify or classify the environmental noise considering that different noise types have different statistical characteristics.

Similar to the pipeline discussed in [3], a Voice Activity Detector (VAD) is used to separate pure noise signals from noisy speech signals. Then, in the absence of speech signals, a noise classification algorithm is used to identify the noise type. A typical noise classification algorithm consists of two components: a feature extractor and a classifier. The literature includes many works where various signal features and classifiers have been utilized to achieve environmental noise classification. In [6], a computationally efficient and effective noise classification scheme was introduced which was later on implemented on the smartphone platform in [7]. This classification scheme uses subband features consisting of band-periodicity and band-entropy features. Band-periodicity features capture the periodicity aspect of noise signals whose characteristics remain more or less stationary over time; whereas band-entropy features capture the non-stationary characteristics of noise signals. An overview of these features is stated next.

Band-periodicity and band-entropy features are computed from signal segments of duration S seconds. Each segment is divided into L overlapping frames of length N , with the l^{th} frame specified by $F_l := \{z_n \mid z_n \in \mathbb{R}, n = 1, \dots, N\}$, where z_n represents the n^{th} sample in the frame. Assuming the sampling rate of fs , the frequency range $[0, fs/2]$ is divided into B non-overlapping subbands. The cross-correlation between every two consecutive frames, that is F_l and F_{l-1} in each band, is computed and the peak value of the cross-correlation is denoted by $P_{b,l}$ where b and l represent the band and frame index, respectively. The band-periodicity feature in band b is then defined as:

$$BP_b = \frac{1}{L} \sum_{l=1}^L P_{b,l}, \quad b = 1, \dots, B \quad (1)$$

where L is the total number of frames over duration S . The band-entropy feature in each band over duration S is defined as:

$$BE_b = \frac{1}{L} \sum_{l=1}^L H_{b,l}, \quad b = 1, \dots, B \quad (2)$$

where $H_{b,l}$ represents the entropy of the l^{th} signal frame in band b [6]. Considering B bands, a feature vector of dimension $2 \times B$ is thus extracted to capture an input noise signal characteristics.

The extracted feature vector is then fed into a Random Forest (RF) classifier to find a matched noise class to the incoming noise signal frames. An RF classifier is an ensemble of a number of classification trees. Each tree is trained independently from the other trees using a randomly selected (with replacement) subset of a training set.

In this study, the classification was applied to three commonly encountered noise environments of babble (non-stationary), driving car (semi-stationary), and machinery (stationary). Interested readers are referred to [7] for more details of the developed noise classification scheme which is used as part of the speech enhancement described next.

B. Noise Adaptive Speech Enhancement

The developed noise adaptive speech enhancement (SE) consists of two stages. The first stage uses the two microphones on a smartphone to increase or improve the signal to noise ratio (SNR) of noisy speech signals using a Block-based Normalized Least Mean Square (BNLMS) error algorithm. In the second stage, a single microphone SE approach with a tuning factor is considered to estimate the a-priori SNR using the Minimum Mean Square Error-Log Scale Amplitude (MMSE-LSA) estimation technique. The value of the tuning factor is obtained experimentally and set based on the type of noise identified by the noise classifier discussed in the previous section.

B.1. SNR improvement—The widely used Normalized Least Mean Square (NLMS) error algorithm is known for its robustness and stable performance. However, due to its sample-based computational approach, it is not suited for real-time implementation on smartphones that require processing to be carried out on a frame-based basis. In [8], a BNLMS algorithm was introduced which is used here as the first stage of our speech enhancement. The adaptation of the filter weights for the BNLMS algorithm is given by [8]:

$$w(l+1) = w(l) + \mu \sum_{n=1}^N x_2(n;l)e(n;l) \quad (3)$$

where $w(l)$ denotes the filter weights for a frame l of length N , μ step size, n sample index in a frame and $e(n, l) = x_1(n, l) - x_2(n, l)$ with x_1 and x_2 representing the signals from the two microphones.

B.2. SNR estimation—In the second stage, a single microphone MMSE-LSA estimator is used which can be stated as [9]:

$$\widehat{X}(k, l) = G(\xi(k, l), \gamma(k, l)) Y(k, l) \quad (4)$$

where $\widehat{X}$ is the estimated speech magnitude spectrum, Y is the received noisy speech spectrum, k represents the frequency bin index ξ & γ denote the a-priori and a-posteriori SNR, and $G(\cdot)$ is a non-linear gain in the form of a lookup table or curve which is setup by assumed statistical models of speech and noise.

For SNR estimation, one popular approach is to estimate the a-priori SNR with an estimated background noise power. This is known as the decision-directed approach and is expressed as follows [9]:

$$\xi(k, l) = \alpha \frac{|\widehat{X}(k, l-1)|^2}{|\widehat{N}(k, l-1)|^2} + (1 - \alpha) \max \left[\frac{|Y(k, l)|^2}{|\widehat{N}(k, l)|^2} - 1, 0 \right] \quad (5)$$

where $\widehat{N}$ denotes the magnitude spectrum of the noise signal, and α is a smoothing parameter varying from 0 to 1.

In [10], our research team has developed a modified estimation approach by minimizing the mean-square error between the true speech spectrum and the estimated speech spectrum containing residual noise. The modified estimation is given by [10]

$$\xi(k, l) = \alpha \frac{|\widehat{X}(k, l-1)|^2}{(1 \pm \rho) |\widehat{N}(k, l-1)|^2} + (1 - \alpha) \max \left[\frac{|Y(k, l)|^2}{(1 \pm \rho) |\widehat{N}(k, l)|^2} - 1, 0 \right] \quad (6)$$

This modification incorporates a tuning factor ρ . By changing ρ , a tradeoff between a-priori SNR overestimation and underestimation is established. By increasing ρ , overestimation of the noise power may occur, leading to less residual noise in the enhanced speech. As ρ is decreased, underestimation of the noise power may occur, leading to a large portion of the noise remaining in the enhanced speech. When $\rho = 0$, (6) becomes the same as (5). For more details on this modification and its block-based real-time implementation, interested readers are referred to [10].

III. Clinical Setup and Method

This section describes the clinical setup and method that were used for testing the developed smartphone-based noise adaptive speech enhancement. Ten participants with bilateral moderately sloping sensorineural hearing loss, ages 20 to 91 years, were fitted with Phonak Audeo V90-312 receiver-in-the-ear HAs in which the microphones were disabled. Ten

participants with normal hearing, ages 20 to 30 years, were fitted with the same HAs for 15dB HL (hearing level) thresholds. The on-ear gain and output were set to within 5dB of NAL-NL1 (a software program used for hearing aid prescription) as verified with a Verifit audioscan fitting system.

As illustrated in Fig. 2, the participants were seated outside of a soundbooth. Each wore the bilateral HAs and a Phonak ComPilot II (a commercially available device worn around the neck allowing streaming audio from smartphones to be transmitted wirelessly to HAs). The smartphone (or Roger Pen – a commercially available device with audio enhancement capability) was placed midway between the two speakers separated by 6.3 feet. A speech signal was presented from the front speaker and three types of background noise (babble, machinery, and driving car) were presented from the rear speaker.

The participants were asked to repeat the HINT [11] sentences presented at 65dBa (a-weighted decibels) in the three background noise stimuli. Noise signals were presented at -5dB or 0dB SNR depending on the participants' performance during a practice session. The three noise types and these three listening stimuli were presented in randomized order: (1) Unprocessed smartphone-stimuli denoting no speech enhancement prior to transmission to ComPilot II; (2) Enhanced smartphone-stimuli denoting speech enhancement by the developed algorithm running on the smartphone prior to transmission to ComPilot II; and (3) Roger Pen-stimuli denoting its proprietary enhancement prior to transmission to ComPilot II. To avoid any delay with the Bluetooth wireless connectivity, as noted in Fig. 2, the participants were seated outside the soundbooth during testing.

IV. Experimental results and discussion

The entire pipeline was coded in C which was then integrated into an Android smartphone (Nexus 6) using the steps provided in [12]. The code shell in this reference was used for the microphone interfacing and the GUI (graphical user interface). The software tools that were used to achieve the smartphone implementation included the IDE (Integrated Development Environment) of Android Studio and the Android SDK (Software Development Kit). To support C codes within Android smartphones, the Android NDK (Native Development Kit) was used.

Initially the developed speech enhancement pipeline was tested via simulation to examine its performance based on the widely used speech quality measure PESQ [5]. As done in [10], a frame shift time of 10ms with a sampling frequency of 16kHz was used making $N=160$. Two SNR situations corresponding to -5dB and 0dB or two different and challenging noise levels were considered, and the PESQ measure was computed and averaged before and after the speech enhancement. Figure 3 illustrates three bar charts corresponding to the three noise types examined. As can be seen from this figure, the PESQ measure after the speech enhancement was improved by more than 40% for each of the three noise types.

Next, the outcome of our clinical study is reported. As shown in Fig. 4, the participants with normal hearing and hearing impairments consistently indicated on average 30% improvement in speech recognition when activating the speech enhancement compared to

the unprocessed condition. The participants indicated the greatest improvement in babble noise. No statistically significant differences were seen between the participant groups ($p=.781$) or between the smartphone enhanced and the Roger Pen enhanced situations ($p=.496$), indicating that the developed speech enhancement pipeline performed as well as a commercially available device. These results indicate that using the smartphone platform offers a viable approach for improving speech understanding in environmental noises. Sample audio clips of the noisy and enhanced speech signals can be downloaded and heard from this link <http://www.utdallas.edu/~kehtar/NIH-Project>.

V. Conclusion

This paper has presented a speech processing pipeline to enhance speech signals in the presence of three types of environmental noises for hearing aid applications. This pipeline has been implemented to run in real-time on the smartphone platform and to interface with commercially available hearing aids. Through both simulation testing and clinical testing, it has been demonstrated that the developed pipeline leads to significant improvements in speech understanding in both normal hearing and impaired hearing users when it is activated on the smartphone platform. The developed solution is general purpose in the sense that it can be utilized to carry out other signal processing algorithms to enhance the hearing capability of hearing aid users.

Acknowledgments

This work was supported by the National Institute of the Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health under the award number 5R56DC014020-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors wish to thank the graduate students Fatemeh Saki, Tyler Caldwell, Chandan Reddy, Yu Rao, Abhishek Sehgal, and Elizabeth Buell, for their contributions in conducting the experiments.

VII. References

- [1]. <http://www.who.int/mediacentre/factsheets/fs300/en/>
- [2]. <http://www.hear-it.org/35-million-Americans-suffering-from-hearing-loss>
- [3]. Gopalakrishna V, Kehtarnavaz N, Mirzahasanloo T, and Loizou P, "Real-time automatic tuning of noise suppression algorithms for cochlear implant applications," IEEE Trans. Biomed. Eng, vol. 59, pp:1691–1700, 2012. [PubMed: 22481802]
- [4]. Gopalakrishna V, Kehtarnavaz N, Loizou P, and Panahi I, "Real-time automatic switching between noise suppression algorithms for deployment in cochlear implants," Proc. of 32nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC'10, pp. 863–866, 2010.
- [5]. Loizou P, Speech Enhancement: Theory and Practice, CRC Press, 2013.
- [6]. Saki F, and Kehtarnavaz N, "Background noise classification using random forest tree classifier for cochlear implant applications," IEEE Int. Conf. on Acoustics, Speech, and Signal Process. ICASSP, pp. 3591–3595, 2014.
- [7]. Saki F, Sehgal A, Panahi I, and Kehtarnavaz N, "Smartphone-based real-time classification of noise signals using subband features and random forest classifier," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP, China, 3 2016.
- [8]. Zhao D, Lu X and Xiang M, "Block NLMS cancellation algorithm and its real-time implementation for passive radar," Proc. of IET Int. Conf, pp. 1–5, 4 2013.
- [9]. Ephraim Y and Malah D, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process, vol. 33, no. 2, pp. 443–445, 4 1985.

- [10]. Rao Y, Hao Y, Panahi I, and Kehtarnavaz N, "Smartphone-based real-time speech enhancement for improving hearing aids speech perception," to appear in Proceedings of IEEE EMBC, Orlando, 8 2016.
- [11]. <http://www.californiaearinstitute.com/audiology-services-hint-bay-area-ca.php>
- [12]. Kehtarnavaz N, Parris S, Sehgal A, Smartphone-Based Real-Time Digital Signal Processing, Morgan and Claypool Publishers, 2015.

Figure 1.
Components of the developed noise adaptive speech enhancement pipeline

Figure 2.
Clinical Setup

Figure 3.
Improvements in speech quality measure PESQ

Figure 4.
Clinical testing results