The Benefit of Speech Enhancement to the Hearing Impaired

N. Fink
Department of Bio-Medical Engineering
Faculty of Engineering
Tel-Aviv University, ISRAEL

C. Muchnik
Department of Communications Disorders
Faculty of Medicine
Tel-Aviv University, ISRAEL

M. Furst
School of Electrical Engineering
Faculty of Engineering
Tel-Aviv University, ISRAEL

Abstract— Most modern hearing-aids include different types of speech enhancement algorithms. Yet, decreased speech intelligibility in background noise is a common complaint of most hearing impaired even when speech enhancements algorithms are functional. Generally, the hearing-aid industry chose those algorithms that were proven to be most adequate to normal hearing subjects. However, it is not clear that an algorithm that is beneficial to normal hearing will increase the intelligibility of the hearing impaired as well, and vice-versa.

We have recently developed a single-channel speech enhancement technique that is based on an ear model comprising outer-hair cell functionality. The algorithm was evaluated in systematic speech intelligibility test of Hebrew words. Hearing impaired subjects, who used either a hearing aid or a cochlear implant, demonstrated a significant improvement in their performance with the algorithm. On the other hand, normal hearing subjects demonstrated no improvement in their performance on the same task. We, therefore, suggest that speech enhancement algorithms for the hearing-impaired should be different from those that are beneficial to normal hearing subjects.

Keywords-component: speech in noise, Cochlear reconstruction algorithm, speech enhancement, speech intelligibility

I. INTRODUCTION

Loss of hearing is a major health problem with serious social implications. Many who have suffered a hearing loss (HL) feel restricted socially and professionally. One of the most common complaints among patients with cochlear hearing loss is difficulty in understanding speech in a noisy environment with or without their hearing assisting devices (hearing aid HA or cochlear implant CI). Current HAs work well in quiet environments and provide the hearing impaired (HI) with improved understanding of auditory signals. Yet hearing assisting devices are less efficient in noisy environments.

Speech Enhancement (SE) algorithms, often called noise reduction (NR) algorithms, aim to improve sound quality and speech recognition in noise. Single-channel based SE algorithms for the HI are efficient in improving speech recognition at positive SNRs. At positive SNRs normal-hearing (NH) subjects have no difficulty in recognizing speech and therefore do not need SE. Multi-channel based SE algorithms are more efficient than single-channel but among their drawbacks are large sized HAs (such as BTE) or the need of binaurally assisting devices.

Sound qualities of SE algorithms have been broadly evaluated but few studies have focused on speech recognition. Some of these studies evaluated the recognition of NH with single-channel SE algorithms [1-3] and reported improvement in speech recognition of up to 33% at SNR of -5 dB with an auditory masked threshold in conjunction with noise suppression AMT-NS [3], partial improvement with the AMT-NS approach of up to 5% at SNRs of 0 dB and 5 dB, none at SNR of -5 dB [2] or no improvement at all with 4 families of SE algorithms: spectral subtractive, sub-space, statistical model and Wiener-filter tested at SNRs of 5 dB and 0 dB [1]. HA-users evaluation of single-channel SE algorithms reported partial improvement with the AMT-NS technique of 2% at SNRs of 5 dB, 0 dB and -5 dB [2]. Studies involving single channel SE algorithms on CI users reported improvement in sentence recognition by 8-21% at positive SNRs (0-9 dB) with spectral subtraction [4] and by 20% at SNR of 5 dB with the subspace algorithm [5]. Alternatively, with a multi-channel SE algorithm based on blind source separation, bilaterally CI recipients improved recognition by 40% at SNR of 0 dB [6].

Recently 5 promising algorithms for speech enhancement were selected and implemented on a common real-time hardware/software platform [7]. Two SE algorithms were single-channel based (perceptually optimized spectral subtraction and Wiener-filter-based noise suppression) and three were multi-channel based (Broadband blind source separation, Spatially preprocessed speech-distortion-weighted multi-channel Wiener filtering – MWF, Binaural coherence dereverberation filter). Listening tests were conducted by different research groups at different sites. Tests were performed with NH and bilaterally HI subjects with flat and sloping mild HL. Three perceptual measures were used: speech reception threshold (SRT), listening effort scaling and preference rating. In a multitalker babble noise, resembling an office scenario (pseudo-diffuse), only one algorithm, the spatially preprocessed speech-distortion-weighted multi-channel Wiener filtering, provided an SRT improvement (of 6-7 dB) relative to the unprocessed condition.

To conclude, Single-channel SE algorithms have not demonstrated persuasive speech recognition improvement
for the HI (HA-users or CI-recipients). Furthermore, single-channel SE algorithms that may benefit a NH subject at negative SNRs may not benefit a hearing-impaired even at positive SNRs.

In the present work a single-channel SE algorithm based on an ear model was used to evaluate speech recognition on a large number of normal hearing, HA users and CI recipients.

II  METHOD

A. Cochlear Representation Algorithm

The cochlear representation algorithm (CRA) is based on a full computational model of the cochlea [8]. The model relies less on heuristics (and parameter estimation) than conventional algorithms that are not based on a hearing model. The CRA integrates outer hair cells (OHC) activity in a one-dimensional cochlear model. In response to a word input to the cochlea, a representation of the basilar membrane's velocity is displayed in color as a function of time and distance from the stapes (fig 1). Alternating the OHC redundancy variable can result in a normal cochlea (fig 1 top panel) or a damaged cochlea response (fig 1 bottom panel). The dynamic properties of a normal cochlear model output are used to reconstruct noisy speech signals with improved SNR [9]. As displayed in figure 2, the CRA method represents the auditory signal (fig 2b), distinguishes between noised and un-noised speech fragments in the input signal according to threshold considerations (fig 2c), and applies a corresponding masker (fig 2d). Reconstruction of the masked signal results in an enhanced signal (fig 2f) in comparison to the noisy input (fig 2a).

B. Psycho-acoustical Experiments

1. Subjects

A total of 99 subjects participated in the experiment according to the following subdivision: (1) NH; (2) HA users; and (3) CI users. The number of participants in each group, their mean age and gender is summarized in Table 1.

![Figure 1: Representation of a normal (top) and damaged (bottom) cochlea response to an input signal (word).](image)

![Figure 2: (a) A Time-domain representation of the word shen at SNR 0 dB. (b) representation of the energy of the signal along the cochlea as a function of time. (c) the algorithm progresses along the time domain in narrow time windows and identifies areas of speech (dark red). (d) A mask with a shape of the identified speech is derived, masking out areas that were not identified as speech. (e) The masker is applied to the representation of the signal and reconstructed back to the time domain (f).](image)

<table>
<thead>
<tr>
<th></th>
<th>HA users</th>
<th>CI users</th>
<th>Normal Hearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>subjects</td>
<td>33</td>
<td>22</td>
<td>41</td>
</tr>
<tr>
<td>Age range</td>
<td>23-64</td>
<td>14-82</td>
<td>17-52</td>
</tr>
<tr>
<td>Age (mean±std)</td>
<td>51±14</td>
<td>49±18</td>
<td>35±14</td>
</tr>
<tr>
<td>Gender</td>
<td>(14F, 19M)</td>
<td>(9F, 13M)</td>
<td>(11F,30M)</td>
</tr>
</tbody>
</table>

Table 1: Age and gender of the different groups participating in the experimental task.

2. Word Database and Noise Types

The database consists of the Hebrew adaptation to the AB List (HAB). The AB list, a set of monosyllabic real words, comprises consonant-vowel consonant (CVC) words. The list was designed so that the different phonemes in English shall be equally distributed
Throughout the entire list [10]. The AB list is commonly used in hearing tests as it reduces the effect of word frequency and/or word familiarity on test scores. Corresponding lists are produced for other languages and accents [11]. The HAB list was designed for Hebrew natives, and it consists of 15 lists of 10 monosyllabic words such as “sir”, “kir” [12]. A single female speaker recorded the HAB list with a sampling rate of 44.1 kHz.

Gaussian white noise was added to the database in various SNRs (0, 5, 10, 18, 24 and 30 dB). The clean and noisy HAB lists were band-pass filtered between 500 Hz and 8 kHz. The filtered lists were applied to the CRA. Eventually, each input word had a corresponding reconstructed word. All together the complete database consisted of several HAB word lists (each list comprising of the same 150 words) in various treatments: noise type (white noise), noise level (SNRs 0,5,10,18,24,30 and quiet) and treatment (CRA, none).

3. Procedure

Each subject passed a standard audiometric test procedure. The NH group consisted of subjects with 15 dB HL (PTA of 500, 1000, 2000 Hz). The hearing of the patients with HAs was assessed with and without their aid.

The psychoacoustical experiment was word recognition in an open-set. Subjects were seated in a sound proof room, wearing their HA/CI in front of a loudspeaker. They were randomly introduced with words chosen from the recorded database. The level of the tested words was adjusted for most comfortable level (MCL).

Each subject was tested in 8 experimental sessions of 10-20 words each. Following each word, the subject was asked to repeat the write the word he/she heard.

HI subjects were tested with SNR levels of 30, 24, and 18 dB, while the NH subjects were tested also with lower SNRs, with a minimum of SNR of 0 dB. HI subjects were not able to perform the experiment in the very low SNRs.

III RESULTS

Mean speech recognition score of 22 CI users is depicted in Fig 3 and of 33 HA users is in Fig 4.

Figure 3: Mean speech recognition score for quiet and for different noise levels for the CI users (n=22). White bars represent score with the CIs without the CRA. Black bars represent score with CI+CRA. Values above bars represent mean speech recognition score. Error bars are displayed. * t-test, p<0.05.

For decreasing SNRs, speech intelligibility for the CI deteriorated with and without the CRA, from 64% at quiet to 29% at SNR18 dB for no treatment and from 57% to 46% with the CRA. Similarly for the HA, Speech intelligibility deteriorated with and without the CRA, from 67% at quiet to 42% at SNR 18 dB for no treatment and from 66% to 53% with the CRA. The use of the CI or HA alone is characterized by a gradual decline in performance as SNR decreases, pointing to the difficulty of understanding speech as the energy of the competing noise increases. In SNRs 18, 24, 30 dB (for CI) and SNR18 dB (for HA), there was a significant improvement in performance relative to the situation when CRA was not applied (t-test, p<0.05). Significant dis-improvement (t-test, p<0.05) in word recognition with the CRA in comparison to without the CRA was found at quiet conditions for the CI.

Figure 4: Mean speech recognition score for quiet and for different noise levels for the HA users (n=33). Significant improvement (t-test, p=0.049) in word recognition with the CRA in comparison to without the CRA was found at SNR of 18 dB.

The results of the NH group are depicted in Fig 5. From the 41 normal hearing subjects, 11 were tested at high SNRs (similar to the HI groups) and reached a 'ceiling effect'. Additional 30 subjects were tested at low SNRs (0, 5 and 10). Both groups were tested similarly in quiet conditions and performed similarly with and without the CRA (99% intelligibility). For decreasing SNRs, Speech intelligibility deteriorated with and without the CRA, from 98% at SNR30 to 53% at SNR0 for no treatment and from 95% to 50% with the CRA.

Figure 5: Mean speech recognition score for quiet and for different noise levels for NH (n=41).
Non-significant difference in performance with and without the CRA was achieved (0%, -3%, -4%, -4%, +1%, 0% and -3% for quiet and SNRs 30, 24, 18, 10, 5 and 0 dB respectively, where negative values depict poorer performance with the CRA rather than with no treatment). These findings are consistent with a comparative intelligibility study of single-microphone noise reduction algorithms on NH [1-2].

IV DISCUSSION

The psychoacoustical experiments revealed a significant improved performance of HI subjects with CRA - the speech enhancement algorithm tested. On the other hand, such an improvement was not observed in NH listeners. It is, therefore, most likely that NH and HI subjects use different strategies in processing and recognizing noisy speech.

Subjects suffering from cochlear lesions and regularly wearing their HAs acquire new acoustic cues that potentially reorganize their cerebral cortex. Alternatively, similar subjects that do not wear their HAs on a daily basis cannot acquire new acoustic cues. Gatehouse [13] has used the term “acclimatization effect” to explain the difference in speech recognition scores between the aided ear and the un-aided ear.

We hypothesize that the different performance in recognition of the different groups is due to different strategies used by these groups, when processing the input from the cochlea. We assume that the strategy used by NH for speech recognition is optimal usage of speech representation redundancy. The speech can be recognized in different overlapping spectral ranges, and the decision is based on one or more frequency range. HI Subjects whom improved their speech recognition with the CRA relative to without have probably been accustomed to the distorted signal produced by their HA/CI which emphasized certain frequency ranges and ignored other frequencies where these subjects did not have residual hearing. Other HI subjects who did not benefit from the CRA relative to their HA or CI, probably still use the strategies of NH which look at the whole frequency range of the speech signal. Speech enhancement techniques reduce the competing noise by estimating the noise spectrum and subtracting it from the noisy speech signal. This generates a distorted signal that emphasizes part of the spectrum where the noise is minimal, but diminishes other parts of the speech signal spectrum. A NH that listens to the enhanced speech cannot use his regular strategy since some of the speech spectrum is missing. On the other hand, the speech enhancement algorithm improves the strategy used by some of the HI, since it cleans and emphasizes the part of the spectrum they perceive and use for recognition.

REFERENCES