Hearing Research 310 (2014) 36–47

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Evaluation of the sparse coding shrinkage noise reduction algorithm in normal hearing and hearing impaired listeners

Jinqiu Sang [a], Hongmei Hu [a], Chengshi Zheng [b], Guoping Li [a], Mark E. Lutman [a], Stefan Bleeck [a,*]

[a] Institute of Sound and Vibration Research, University of Southampton, SO17 1BJ, UK
[b] Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form
15 January 2014
Accepted 24 January 2014
Available online 2 February 2014

ABSTRACT

Although there are numerous single-channel noise reduction strategies to improve speech perception in noise, most of them improve speech quality but do not improve speech intelligibility, in circumstances where the noise and speech have similar frequency spectra. Current exceptions that may improve speech intelligibility are those that require *a priori* knowledge of the speech or noise statistics, which limits practical application. Hearing impaired (HI) listeners suffer more in speech intelligibility than normal hearing listeners (NH) in the same noisy environment, so developing better single-channel noise reduction algorithms for HI listeners is justified. Our model-based "sparse coding shrinkage" (SCS) algorithm extracts key speech information in noisy speech. We evaluate it by comparison with a state-of-the-art Wiener filtering approach using speech intelligibility tests with NH and HI listeners. The model-based SCS algorithm relies only on statistical signal information without prior information. Results show that the SCS algorithm improves speech intelligibility in stationary noise and is comparable to the Wiener filtering algorithm. Both algorithms improve intelligibility for HI listeners but not for NH listeners. Improvement is less in fluctuating (babble) noise than in stationary noise. Both noise reduction algorithms perform better at higher input signal-to-noise ratios (SNR) where HI listeners can benefit but where NH listeners have already reached ceiling performance. The difference between NH and HI subjects in intelligibility gain depends fundamentally on the input SNR rather than the hearing loss level. We conclude that HI listeners need different signal processing algorithms from NH subjects and that the SCS algorithm offers a promising alternative to Wiener filtering. Performance of all noise reduction algorithms is likely to vary according to extent of hearing loss and algorithms that show little benefit for listeners with moderate hearing loss may be more beneficial for listeners with more severe hearing loss.

1. Introduction

For people with mild to severe hearing losses, current advanced hearing aids can help improve speech perception in quiet environments. However, one important reason why hearing-aid users often do not like to use hearing aids is that the current hearing aids do not work well in background noise (Alcantara et al., 2003; Dillon, 2001). Hearing-impaired (HI) people typically require a speech-to-noise ratio that is 3–6 dB higher than normal-hearing (NH) people to achieve the same degree of speech intelligibility (Alcantara et al., 2003; Plomp, 1994). Therefore, noise reduction strategies in hearing aids are one critical factor to help improve quality of life for HA users.

The most effective speech enhancement method to improve speech intelligibility today is through beamforming using microphone arrays (Kates and Weiss, 1996; Levitt, 2001; Schum, 2003), however, they work best with large microphone arrays. They also only work effectively when the target speech and interfering sounds are coming from different directions. However, due to the small size, usually only one or two microphones are placed in a HA and it is not possible to create large enough arrays. Therefore there is still a need to also develop better single channel noise reduction schemes. Currently, most HAs are equipped with a combination of single-channel noise reduction algorithms and beam-forming

Abbreviations: BKB, Bamford–Kowal–Bench sentence; CI, cochlear implant; CS-WF, a Wiener filtering approach with cepstral smoothing; HA, hearing aid; HI, hearing impaired; MAP, maximum a posteriori; NAL, National Acoustics Laboratory procedure; NH, normal hearing; NPSD, noise power spectral density; SCS, sparse coding shrinkage; SNR, signal-to-noise-ratio; SPP, speech presence probability; SRT, speech reception threshold; SSN, speech shaped noise

* Corresponding author. Tel.: +44 (0)2380596682; fax: +44 (0)2380593190.
E-mail addresses: s.bleeck@soton.ac.uk, rdsjq66@126.com (S. Bleeck).

0378-5955/$ – see front matter
http://dx.doi.org/10.1016/j.heares.2014.01.006

strategies (Widrow and Luo, 2003), and together they determine the final noise reduction performance of a HA. There are also situations in which only single-channel strategies can be used, for example in telephone speech or in HAs that are placed entirely in the ear canal.

The scope of the present work is limited to the effects of single-channel noise reduction algorithms on speech intelligibility in noise, more specifically in situations when the noise has the same long-term frequency spectrum as the speech signal. That is one of the most challenging situations for noise reduction algorithms. The scope does not include assessment of speech quality or trading off speech intelligibility against speech quality improvement.

Previous research (Dahlquist et al., 2005; Levitt, 1993; Levitt et al., 1993; Weiss and Neuman, 1993) has shown that a higher signal-to-noise ratio (SNR) in HAs does improve speech quality, but does not necessarily lead to benefits in understanding speech for a hearing-impaired listener. Noise reduction algorithms lower the noise level thereby reducing the loudness and annoyance of the interfering noise. The lower noise level is less distracting and is a factor contributing to the reported improvements in sound quality. Noise reduction strategies often do not improve speech intelligibility, because the processing can remove essential parts of the signal or distort the speech in a way that reduces intelligibility. The main exception to this generalization is when speech and noise have different frequency spectra and simple filtering can reduce the remote masking effect of noise without reducing the speech signal adversely.

Reviews of single-channel noise reduction algorithms with NH listeners to date have concluded that no speech intelligibility improvement occurs (for example Hu and Loizou, 2007), except for the algorithms which have *a priori* knowledge of the statistics of the speech and/or background noise (Kim and Loizou, 2010). However, such algorithms are neither robust nor practical in real acoustic environments. On the other hand, noise reduction algorithms have shown inconclusive positive and negative effects with HI listeners (Arehart et al., 2003; Dahlquist et al., 2005; Elberling et al., 1993; Harlander et al., 2012; Jamieson et al., 1995; Levitt, 2001; Levitt et al., 1993). Although there have been studies with positive effects on intelligibility for HI listeners, this may have been due to the noise having a different spectrum from the speech and is thus easier to reduce (Arehart et al., 2003). When noise spectrum is similar to speech spectrum, the general finding is that algorithms improve speech quality, without improving speech intelligibility. For example, Elberling et al. (1993) evaluated three spectral subtraction algorithms in babble noise and reported that the algorithms reduced the noise level but did not improve speech intelligibility in either NH listeners or HI listeners. Most recently, Harlander et al. (2012) evaluated model-based versus non-parametric monaural noise reduction approaches with HI listeners in stationary noises, non-stationary noises and one quasi-stationary noise. They found that none of the algorithms improved speech intelligibility, although the two model-based noise reduction algorithms improved speech quality (Harlander et al., 2012). The main exception to the generalization that single-channel noise reduction algorithms do not improve speech intelligibility occurs in studies of cochlear implant (CI) users. For example, a nonlinear spectral subtraction algorithm (Lockwood and Boudy, 1992) that did not show speech intelligibility improvement for HA users (Dahlquist et al., 2005) showed intelligibility improvement for CI users for the same speech shaped noise and for the same sentence tests (Verschuur et al., 2006). A very similar effect was shown in an independent study (Yang and Fu, 2005).

When developing noise reduction algorithms for HI listeners, all hearing loss factors should be taken into account, and compensated for, where possible. For people with sensorineural hearing losses, hearing loss factors include threshold elevation, loudness recruitment, reduced frequency selectivity and reduced temporal resolution. Automatic gain control can compensate for threshold elevation and loudness recruitment, but there are currently no appropriate solutions to compensate reduced frequency selectivity and reduced temporal resolution. Some researchers have attempted to compensate for reduced frequency selectivity with spectral sharpening but this did not improve intelligibility (Baer et al., 1993). A possible solution to reduce the effects of reduced frequency and temporal resolution is to extract and preserve key speech information while at the same time reducing the remaining speech and the overall noise. This way, there will be less self-masking and noise-masking of speech components yet essential speech information may be preserved after noise reduction. Of course, this begs the question of how to identify and preserve the key speech information. That is the focus of the present work. To this end, we investigate here a sparse coding shrinkage (SCS) noise reduction algorithm to extract key information from noisy speech. The approach exploits the principle that the speech signal is highly redundant and information is distributed sparsely in a noisy speech signal. By increasing the sparseness of a noisy speech signal, there is a large likelihood that intelligibility is improved (Li et al., 2012). The algorithm assumes a super-Gaussian (sparse) distribution of the principal components in clean speech and works by applying sparse shrinkage on the principal components. SCS was first proposed by (Hyvärinen, 1999) for image noise reduction (Hyvärinen et al., 1998) and later also for speech enhancement in noise (Hu et al., 2013, 2011; Li, 2008; Li and Lutman, 2008; Potamitis et al., 2001; Sang et al., 2011a,b; Zou et al., 2008). Sparse coding has shown significant benefit for cochlear implant users (Li and Lutman, 2008) and this suggests that there may be potential benefits of SCS for HA users too.

The performance of the SCS algorithm is compared with a state-of-the-art Wiener filtering approach: CS-WF (Breithaupt et al., 2008; Gerkmann and Martin, 2009; Gerkmann and Hendriks, 2012). Wiener filtering approaches can reach optimal performance when the speech and noise both have Gaussian distribution. However, in real environments, neither noise nor speech is usually Gaussian. As SCS has been developed to estimate the speech components with the assumption of super-Gaussian distribution, we hypothesize that SCS might perform better than CS-WF especially for HI listeners with reduced frequency and temporal resolution, who we propose would benefit from removal of redundant parts of the speech signal as well as noise.

The SCS was also compared with unprocessed speech as baseline performance in a noisy environment without any algorithms applied. Previous research demonstrated that noise reduction algorithms might reduce speech intelligibility for HI listeners (Dahlquist et al., 2005). The comparison with unprocessed speech is used to investigate whether there is any benefit of noise reduction algorithms for HI listeners. Babble noise and speech shaped noise were chosen as the additive noise due to their similar average long term spectrum when compared with the speech signal. In much of the previous research, speech intelligibility is quantified in terms of percentage of identified words (or syllables) correct. Percentage intelligibility is often measured at fixed input SNRs. As such intelligibility measures choose fixed input SNR during speech recognition tests, a relatively low input SNR might show poor performance of both unprocessed speech and enhanced speech while a relatively high input SNR might already show high performance of unprocessed speech and no need of noise reduction algorithms. Therefore such intelligibility measures are inherently limited by floor or ceiling effects. An alternative measure of speech