TOWARDS SNR-LOSS RESTORATION IN DIGITAL HEARING AIDS

Rob A.J. de Vries and Bert de Vries

GN ReSound-Beltone Netherlands B.V.
Hurksestraat 42, 5652 AL Eindhoven
The Netherlands, devriesr@beltone.com

ABSTRACT

Loss of understanding speech in noise, known as signal to noise ratio (SNR) loss, is the largest problem of most hearing aid users today. This paper discusses the most important noise reduction (NR) agents available in current hearing aids and their ability to improve SNR-loss. Then, it is investigated how the many different NR agents can best be integrated to maximize results. A global minimization shows that a (Generalized Sidelobe Canceller) beamformer followed by single microphone NR agents is the theoretically optimal configuration. Since the performance of the different NR agents depends on the environment, it is suggested to use an environmental classifier to select the optimal combination of NR agents. Either by a fixed set of rules or by using a self training network catered to the specific SNR-loss of the hearing impaired.

1. INTRODUCTION

Hearing-impaired listeners often experience two problems: Hearing loss, which is an increase in hearing threshold and SNR-loss, which is loss of understanding high level speech in noise. SNR-loss is measured by the speech reception threshold (SRT), which is the SNR required for 50% correct word recognition. Hearing loss is typically caused by conductive loss or loss of outer hair cells. SNR-loss is normally caused by loss of inner hair cells. On average, a hearing loss of 30 to 60 dB is accompanied by a 4 to 7 dB SNR-loss [1]. However, accurate estimates of the SNR-loss can only be obtained by specific testing, since hearing loss and SNR-loss are basically independent characteristics.

By applying the appropriate gain in different bands and using superb circuitry, most modern hearing aids can readily restore hearing loss. Moreover, by making previously inaudible speech cues audible without much distortion, they significantly improve the SRT for low level speech in noise. However, restoring or even improving SNR-loss is far more difficult and far from solved. Killion [1] even states that currently no known one-microphone hearing aid manages to improve SNR-loss. The difficulty lies in the fact that filtering the noise often goes at the expense of speech cues. Moreover, techniques like spectral enhancement fail to give significant improvements, because the diminished frequency resolution caused by a loss of inner hair cells can not be restored [2]. So, although hearing-impaired listeners can now clearly hear what people say, their major complained is that they still can not understand them [1]. Killion [1] and Smoorenburg [2] therefor conclude that the main focus of hearing instruments should be on improving SNR-loss by reducing noise without jeopardizing speech cues.

Although improving SNR-loss (intelligibility/SRT) is one of the main goals of noise reduction in hearing aids, two other important objectives are; Comfort, reducing noise and possibly increasing the overall SNR without significantly improving SNR-loss, is still very much worthwhile when it makes the sound more pleasing. This has the added benefit that it makes it less tiring for the wearer to concentrate on the signal for a longer time. Thereby improving intelligibility indirectly. Natural sound, improving SNR-loss at the cost of very unnatural sounds is not desired by everyone. Another thing to keep in mind is the very low computational power available in hearing aids. E.g. one usually has only about one to two MIPS available for noise reduction.

This paper focusses on the main approaches to noise reduction currently used in hearing aids and how they can be integrated to restore SNR-loss as much as possible.

2. MULTIPLE OBSERVATIONS BASED NOISE REDUCTION

The most successful techniques to improve the SNR and SRT use microphone arrays. Simply because these can exploit spatial differences between the desired- and interfering sources, in addition to temporal- and spectral differences used in single observation based systems. Barring ultra-low power wireless communications, two closely-spaced microphones in a single behind-the-ear hearing aid is the preferred configuration; It avoids cumbersome wiring and is cosmetically appealing.

One popular configuration allows the user to switch between an omni- and a directional microphone, generally improving SNR with 4 dB and SRT up to 7 dB. To overcome the disadvantage of the fixed, usually cardiode, polar pattern of a directional microphone, both GN ReSound and Phonak use a ”delay-subtract” like beamformer as in figure 1. Here, $x_1 = s + n_1$ denotes the front microphone signal with s the target signal coming from the front and n_1 the collection of all sounds coming from other directions. Similarly, the rear microphone signal is defined as $x_2 = s_2 + n_2$, where $s_2 = F^{-1}\,s$. Ideally, the filter F^{-1} is a pure delay equal to d/c seconds where d is the inter-microphone distance and c the speed of sound. By adapting the delay one can place the null in the polar plot at the strongest interfering source. Constraints are

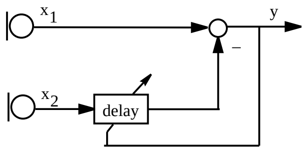

Fig. 1. Delay-subtract beamformer

needed in this adaptation to ensure that signals from the front are not cancelled. Adaptive microphone matching is also desired to ensure good performance over time.

Another popular scheme, though not yet available in hearing-aids, are variations on the Generalized Sidelobe Canceller (GSC) [3] shown in figure 2. In the first stage of the GSC ”delay-sum” and delay-subtract beamformers are used to generate the signal reference $x_4 = s + n_1/2 + F\, n_2/2$ and the noise reference $n_3 = n_1 - F\, n_2$, respectively. In a second stage adaptive noise cancellation is performed by minimizing $E\{|x_4(t) - G(t) * n_3(t)|^2\}$, where the asterix denotes convolution and $E\{\,\}$ expectation. The delays are used to implement non-causal filters and F and G are FIR’s of 10 to 50 taps. Since n_3, ideally, does not contain the target signal, the filter G will adapt to cancel the noise only. Due to miss-steering, microphone mismatch, etc, the target signal will usually ”leak” into the noise reference. Then G will also adapt to cancel the target signal. To avoid this, a speech detector is often employed to freeze adaptation in the presence of speech, which is assumed to be the target signal. The GSC can improve SNR by 8 dB and SRT up to 11 dB.

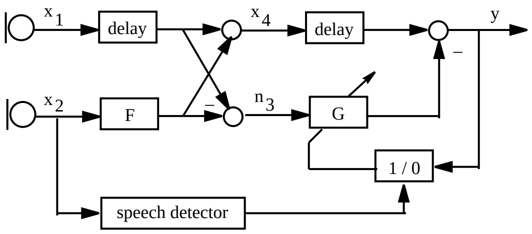

Fig. 2. Generalized sidelobe canceller

Although there are many other approaches to beamforming, these quickly require too much computational power. Increasing the number of microphones is possible through special add-on’s like Etymotic Research’s ArrayMic^{TM} and Starkey’s Radiant Beam Array which interface to the hearing aid via a telecoil.

Adaptive beamformers like the GSC will generally improve SNR and SRT the most. However, at cocktail parties, a fixed hyper cardoide beamformer will often give better results. An omni-directional pattern is often desired in quite and necessary on the street to hear uncoming traffic. So, the optimal beamformer will be environment dependent.

3. SINGLE OBSERVATION BASED NOISE REDUCTION

Noise reduction based on a single observation remains very relevant since many hearing aid models, such as in-the-canal types, do not have enough space to place two microphones at sufficient distance. In hearing aids, computational simplicity is of the essence, and consequently frequency-dependent attenuation for bands with low signal-to-noise ratio is the dominant technique in the hearing aids industry. Figure 3 illustrates modulation based filtering, a technique that is being used by many manufacturers, including Phonak, Siemens and GN ReSound. Modulation filtering can be applied independently in few to many frequency bands. The background is that the difference between the maximum and minimum of the envelope (the modulation index, abbrev. MDX) is generally much larger for speech signals (about 20 dB) than for noise sources. If a small MDX is measured, the signal in the band is attenuated through an expansion circuit. For large MDX, speech is assumed present and the expansion circuit is not activated. The modulation index is a good estimator for the SNR, since the ambient noise level can be estimated by tracking the lowest excursions of the envelope. The expansion circuit is basically a linear approximation (with saturation) of the Wiener filter,

$$G = \frac{SNR}{1 + SNR} \tag{1}$$

which is the optimal linear mean square error (LMSE) noise reduction filter. The response time of the circuit to changes in the envelope of the signal, as measured by attack and release times of the expander, is of crucial importance. If the attack time (attack of attenuation as a result of small MDX value) is in the order of seconds, the circuit will not have time to suppress a noisy speech signal during the low excursions of the envelope. The envelope of speech follows the syllable rate of speech which is maximal at about 4 [Hz]. Essentially, in this case the circuit acts as a speech detector that activates an expansion circuit during speech absence. If the attack time is much smaller, say in the order of tens of milliseconds, then the circuit will be quick enough to suppress local time-frequency regions with low SNR in the noisy speech signal. In this case, modulation filtering becomes a variant of spectral subtraction, a more ambitious technique aimed at improving SNR during speech presence. Unfortunately, fast gain changes in low SNR leads to a very annoying artifact that is called musical noise. Ephraim and Malah have developed a SNR tracking lowpass filter that reduces the musical noise phenomenon by smoothing the gain changes during low SNR, [4]. Since perceptual quality is essential for hearings aids, we highly recommend this technique when fast attack times are desired in modulation filtering or spectral subtraction. While slow-acting modulation filtering has become very popular in the hearing aids industry, fast-acting spectral subtraction has not found much application yet because the artifacts have proven to be too bothersome.

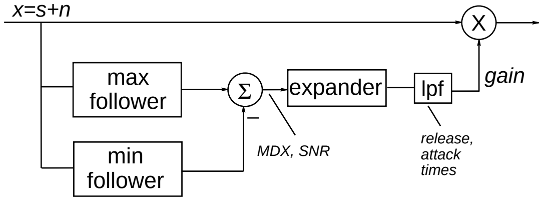

Fig. 3. Modulation index filtering

4. SPEECH ENHANCEMENT

While spectral subtraction and its variants are often described as a speech enhancement method, technically this term should be reserved for techniques that actually work on applying gain to specific speech segments or features. The perceptual results of speech enhancement have an even worse reputation than spectral subtraction. We mention two variants that have found applications in industrial hearing instruments. GN ReSound enhances the contrast (spatial highpass filtering) of the magnitude spectrum in their Canta-7 instrument, so as to favor speech features (high energy formants) over noise. Siemens has introduced the ConTrast system that attempts to enhance low energy transient speech segments such as plosives (/p/,/t/) and fricatives (/f/,/s/). The ConTrast system works independently in each of the higher 6 frequency bands (not below 500 Hz).

5. PERSPECTIVES ON INTEGRATION

We have just reviewed a number of techniques to improve the effective SNR for a speech/noise mixture, including beamforming, spectral subtraction, low-level expansion and speech enhancement. Can we just put all these algorithms in cascade (or parallel to each other) and expect to perceive the added benefits of each processing agent? It is easy to generate examples where the various noise reduction (NR) agents seem to act in opposite ways. For instance, a low energy level (unvoiced) following a high engery vowel will be suppressed by the spectral subtraction technique (since the local SNR for low-level inputs is low), whereas a speech transient boosting technique such as Siemens' ConTrast Enhancement may try to amplify the same segment. Apparently some kind of integration strategy is needed in order to maximize the effect of multiple active NR agents.

5.1. Global optimization

One approach to obtain an integration strategy for different NR agents is as follows. Solve the global optimization problem, using information that is not available in practice if necessary. Then approximate the optimal solution as good as possible by integrating different NR agents. The global objective chosen in this paper is to obtain an as good as possible approximation $\hat{s}$ of the target signal s in the MMSE sense, by linearly filtering the two microphone signals x_1 and x_2.

Consider the notation $X(f)^H = (X(f)^*)^T$ where $X(f)$ is the Fourier transform of a signal vector $x(t)$, the asterix denotes complex conjugate and T transpose. Further, let $S(f)$ be the target signal (that is unknown in practice), $X(f) = [X_1(f),\ X_2(f)]^T$ the input observation vector and $\mathbf{H}(f) = [H_1(f) H_2(f)]^T$ the filter gain vector in the frequency domain. Then the estimated signal is given by

$$\hat{S}(f) = \mathbf{H}(f)^T\, X(f) \tag{2}$$

and the optimal filter minimizing the error

$$E\left\{ \left| S(f) - \hat{S}(f) \right|^2 \right\} \tag{3}$$

is given by the non-causal Wiener filter for two observations:

$$\mathbf{H}_{opt}(f) = \left(\Gamma_{XX}^{-1}(f)\, \Gamma_{XS}(f) \right)^* \tag{4}$$

where $\Gamma_{XX}(f) = E\{X(f)X(f)^H\}$ is the power spectral density (psd) matrix of $X(f)$ and $\Gamma_{XS}(f) = E\{X(f)S(f)^H\}$ is the cross-psd between $X(f)$ and $S(f)$.

We will now show that $\hat{S}_{opt}(f) = \mathbf{H}_{opt}(f)^T\, X(f)$ can be rewritten as a cascade of the GSC given in section 2 followed by (1). To do this, we need the following definitions which are similar to those given in section 2. To simplify notation we will often drop the argument (f) in the rest of this section.

$$\begin{array}{lllccc}
X_1 & = & S + N_1, & X_2 & = & S_2 + N_2 \\
S & = & F\, S_2, & N_3 & = & N_1 - F\, N_2 \\
X_4 & = & S + N_4, & N_4 & = & \frac{N_1}{2} + F\, \frac{N_2}{2} \\
U & = & [N_3\, X_4]^T, & \hat{\tilde{S}}_{opt} & = & \tilde{\mathbf{H}}_{opt}^T\, U \\
\mathbf{G}_{opt} & = & \frac{\Gamma_{N_4 N_3}}{\Gamma_{N_3 N_3}}, & \tilde{\mathbf{H}}_{opt} & = & \left(\Gamma_{UU}^{-1}\, \Gamma_{US} \right)^* \\
Y & = & S + M, & M & = & N_4 - \mathbf{G}_{opt}\, N_3
\end{array} \tag{5}$$

where F is assumed to be known, $\mathbf{G}_{opt}$ and Y are the optimal filter and output of the GSC of section 2, respectively, and $\tilde{\mathbf{H}}_{opt}$ is the optimal filter minimizing the error $E\{|S - \tilde{\mathbf{H}}^T\, U|^2\}$. Assuming that S is independent of N_1 and N_2, it follows trivially from (5) that

$$\tilde{\mathbf{H}}_{opt} = \frac{\Gamma_{SS}}{\Gamma_{SS} + \Gamma_{MM}} \left[\begin{array}{c} -\mathbf{G}_{opt} \\ 1 \end{array} \right] \tag{6}$$

Hence, $\hat{\tilde{S}}_{opt}$ clearly consists of the GSC followed by (1).

From (5) it follows that

$$T_1 = \left[\begin{array}{cc} \frac{1}{2} & 1 \\ -\frac{1}{2F} & -\frac{1}{F} \end{array} \right], \quad \left[\begin{array}{c} X_1 \\ X_2 \end{array} \right] = T_1 \left[\begin{array}{c} N_3 \\ X_4 \end{array} \right] \tag{7}$$

$$T_2 = \left[\begin{array}{cc} 1 & -F \\ \frac{1}{2} & \frac{F}{2} \end{array} \right], \quad \left[\begin{array}{c} N_3 \\ X_4 \end{array} \right] = T_2 \left[\begin{array}{c} X_1 \\ X_2 \end{array} \right] \tag{8}$$

where (7) only holds when F is invertably stable. In this case it follows from (2) - (8) that

$$\hat{S}_{opt} = \mathbf{H}_{opt}^T\, X = \mathbf{H}_{opt}^T\, T_1\, U = \tilde{\mathbf{H}}_{opt}^T\, U = \hat{\tilde{S}}_{opt}$$

where the third equality follows from the fact that T_1 is a linear, stable and invertably stable filter. When F is not invertably stable, which is often the case in practice, it follows from (5), (6), (8), the independence of S from N_1 and N_2 and by very careful rewriting, that

$$\begin{aligned}
\mathbf{H}_{opt}^* &= \frac{\Gamma_{S_2 S_2}}{\Gamma_{S_2 S_2}\Gamma_{N3N3} + \Gamma_{N_1 N_1}\Gamma_{N_2 N_2} - \Gamma_{N_1 N_2}\Gamma_{N_2 N_1}} \\
&\quad \left[\begin{array}{c} |F|^2\Gamma_{N_2 N_2} - F^*\Gamma_{N_1 N_2} \\ -|F|^2\Gamma_{N_2 N_1} + F^*\Gamma_{N_1 N_1} \end{array} \right] \\
&= \frac{\Gamma_{SS}}{\Gamma_{SS} + \frac{|F|^2}{\Gamma_{N3N3}} \left(\Gamma_{N_1 N_1}\Gamma_{N_2 N_2} - \Gamma_{N_1 N_2}\Gamma_{N_2 N_1} \right)} \\
&\quad \left[\begin{array}{cc} 1 & \frac{1}{2} \\ -F^* & \frac{F^*}{2} \end{array} \right] \left[\begin{array}{c} -\frac{\Gamma_{N_3 N_4}}{\Gamma_{N_3 N_3}} \\ 1 \end{array} \right] \\
&= \frac{\Gamma_{SS}}{\Gamma_{SS} + \Gamma_{N_1 N_1} - \frac{\Gamma_{N_1 N_3}\Gamma_{N_3 N_1}}{\Gamma_{N_3 N_3}}} T_2^H \left[\begin{array}{c} -\mathbf{G}_{opt}^* \\ 1 \end{array} \right] \\
&= \frac{\Gamma_{SS}}{\Gamma_{SS} + \Gamma_{N_4 N_4} - \frac{\Gamma_{N_3 N_4}\Gamma_{N_4 N_3}}{\Gamma_{N_3 N_3}}} T_2^H \left[\begin{array}{c} -\mathbf{G}_{opt}^* \\ 1 \end{array} \right]
\end{aligned}$$

which shows that $\hat{S}_{opt} = \hat{\tilde{S}}_{opt}$ even when F is not invertably stable.

We can conclude that there is no need to combine and/or optimize the single observation based NR strategies in some parallel way with the dual observation based NR strategies; The theoretical optimal NR strategy for two observations consists of the GSC followed by the single observation based Wiener filter (1). In practice, this "post-filter" is approximated by a combination of the single observation based NR agents presented in section 3. How this combination can best be chosen is discussed in the next section.

5.2. Divide-and-conquer

A computationally attractive way to integrate different NR agents, is based on the idea to divide the acoustic input space in a set of classes and apply an appropriate NR agent in each class. A NR algorithm such as spectral subtraction with fixed parameters is too simple to provide pleasing perceptual results for all possible acoustic inputs. We need a controller that turns spectral subtraction off if the acoustic conditions do not match the assumptions, e.g., if the ambient noise is non-stationary or speech-like such as at a cocktail party. The availability of a gating controller allows for simpler realization of the set of NR agents. This type of processing is akin to the mixture of experts or gating network approach that is popular in the machine learning community. If the classes are well chosen, this approach can lead to substantial computational savings.

5.2.1. *Environment classification*

The method is examplified by Figure 4 without the dotted parts. We have an adaptive beamformer and three single-observation NR agents that are specialized for car noise suppression and speech and music enhancement. A trained environment classifier (e.g., based on a hidden Markov model, see [5]) divides the acoustic input space into the classes speech, music and speech-in-car-noise (and a none-of-the-above class). The classifier output also controls the weight adaptation process of the beamformer (it can serve as a speech detector) and the amount of beamforming (frontal beamforming may be undesirable in a car). For high probability of music, the music enhancement agent is turned on. Selection of the other NR agents is similarly based on high classifier output for the corresponding class.

Simple realizations of the gating network approach are currently in the market. For example, Oticon's VoiceFinder and Phonak's AutoSelect are examples of 2-class environmental classifiers that control the impact of NR agents.

5.2.2. *Learning to restore SNR loss*

How many acoustic classes and subclasses do we need? There is speech, music, car noise, babble, speech-in-car-noise, music and speech, music and speech in car noise, car noise in traffic noise and so on. The list is endless. Moreover, the acoustic subspace that we call speech-in-car-noise (or any other class) may not be the appropriate class to activate a particular NR agent. Essentially, speech-in-car-noise is a label that we ascribe to a subspace of sounds that have similar perceptual characteristics, but this label may be useless relative to the applicability of our available NR agents. In a more advanced version of the gating network method, the appropriate acoustic classes are learned from a relevant set of examples, cf. Figure 4. In this system we have a set of available NR agents $G_k, k = 1, \ldots, K$, and a weighted total gain $\sum_k p_k G_k$. The weights p_k are the output of the environmental gating network and indicate the relevance of the kth NR agent. The gating network is trained to optimize a perceptual cost on the total NR network output and a target output database, e.g. sum of log-spectral differences on a Bark scale. For a hearing aids application, the optimal gain reduction (α) is a function of the SNR loss for the patient. In principle, the NR network in Figure 4 attempts to restore the patient-specific SNR loss. Note that we don't assign environmental classes a priori, but rather let the classifier determine the relevance of a NR agent by means of p_k. The NR network proposed here is not currently available in industrial hearing aids.

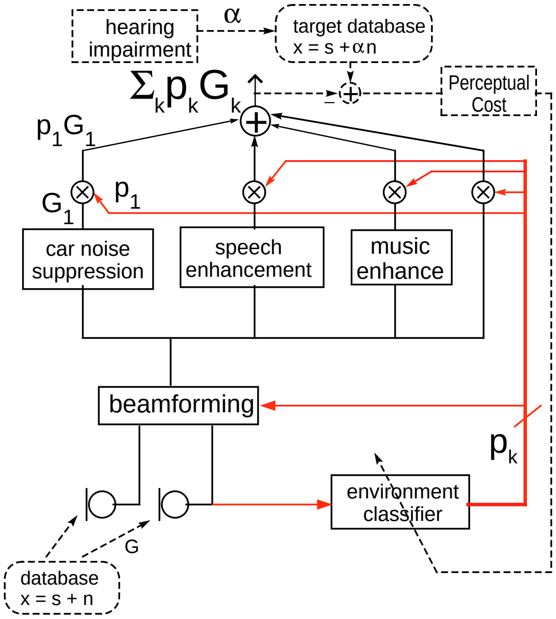

Fig. 4. Noise Reduction network that learns to restore SNR loss α

6. CONCLUSIONS

This paper discussed the possibilities of restoring SNR-loss, the main problem of today's hearing impaired. The best agents to restore SNR-loss are beamformers. Most current single-microphone NR agents improve comfort, but fail to improve SNR-loss in many environments. A theoretical analysis showed that parallel optimization of single- and multiple-microphone NR techniques is not necessary; The optimal configuration consists of a beamformer followed by a post processing stage of single-microphone NR agents. The best beamformer and the best NR agent for post processing are environment dependent. The contribution of the different NR agents should therefor be determined by an environmental classifier that either uses a fixed set of rules or is trained on a large data set. The latter is recommended because the relevant acoustic classes are difficult to determine a priori and may be many.

7. REFERENCES

[1] M.C. Killion, "SNR loss: I can hear what people say, but I can't understand them", The Hearing Review, vol. 4, no. 12, pp. 8, 10, 12 & 14, Dec. 1997.

[2] G. Smoorenburg, "Psychoacoustics of hearing loss", http://www.medical.hear-it.org/multimedia/smoorenburg.pdf, EHIMA, Brussels, May 1999.

[3] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming", IEEE Transactions on Antennas Propogation, vol. AP-30, no. 1, pp. 27-34, Jan. 1982.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

[5] S. Allegro, M. Büchler, S. Launer, "Automatic sound classification inspired by auditory scene analysis", Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis, http://www.ee.columbia.edu/crac/, Aalborg, Denmark, Sept. 2nd 2001.