SPEECH INTELLIGIBILITY ENHANCEMENT BY EQUALIZATION FOR IN-CAR APPLICATIONS

Enguerrand Gentet[1,2], *Bertrand David*[1], *Sébastien Denjean*[2], *Gaël Richard*[1], *Vincent Roussarie*[2]

[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[2]Groupe PSA, Chemin de Gisy, 78943 Vélizy-Villacoublay, France

ABSTRACT

In this paper, we propose a speech intelligibility enhancement method for typical in-car applications in noisy environments. While traditional speech enhancement algorithms aim at increasing the Signal to Noise Ratio (SNR), the goal here is to increase intelligibility by applying dedicated voice transformation techniques without changing the original SNR. The proposed method consists in an adaptive equalizer which reallocates the energy of frequency bands to maximize the Speech Intelligibility Index (SII) under the constraint of a fixed perceived loudness. The validation of the algorithm is carried out by means of a perceptual test derived from the Hearing in Noise Test (HINT) using four typical in-car noises of different driving conditions. The results obtained demonstrate the merit of the algorithm for low-frequency noises, that correspond to usual driving conditions, but also show the limit of the algorithm on noises with a spectrum more spread out induced by rain.

***Index Terms*—** near-end listening enhancement, speech intelligibility index, sentence recognition in noise

1. INTRODUCTION

Speech is nowadays present in a number of in-car applications ranging from hands-free communications, radio programs and podcasts to speech synthesis messages from the various car devices (board, navigation system,...). However, despite the steady car manufacturing progress, significant noise still remains in the car interior. It may come from the engine, the air friction, the rolling or from any other source of noise. In any case it often leads to a loss of intelligibility of speech signals.

Typical methods for increasing in-car listening comfort involve on the one hand classical noise reduction methods (to directly process corrupted speech) and mechanical noise control methods (to reduce the production of noise at its source). In our case, we rather try to mitigate the noise impact by transforming the speech signal before it is transmitted or played in the noisy environment. More precisely, we are aiming at efficiently process speech signals in order to improve their intelligibility without modifying their perceived loudness. There is a clear growing interest for this *near-end listening enhancement* problem [1–13] for which a dedicated international challenge was recently organised [14].

The main idea of our approach is to apply a dynamic equalizer to the speech signal in order to optimize an objective measure of intelligibility. A general approach to measure intelligibility is to make a correlation-based comparison between the spectro-temporal representation of the clean and degraded versions of the speech signal [15–17]. In contrast, the SNR-based measures need the noise as a separate signal for the estimation of intelligibility [18, 19]. In both case, the computation can be done on small signal chunks [17, 19] or on longer segments [15, 16, 18]. Typical car noises we are interested in are stable, mostly stationary and can usually be measured in absence of speech with the in-car microphones. That is why we selected the widely used SNR-based measure *Speech Intelligibility Index* (SII) [18] computed on complete signals.

This strategy has already been followed in previous studies using linear [6] or non-linear approximations [11, 12]. Besides being sub-optimal, these approaches rely on a power limitation constraint which cannot guarantee that the perceived loudness is not increased by the processing. The main contribution of our approach is then to propose an exact solution to a direct speech intelligibility optimization problem under the constraint of a fixed perceived loudness.

Our results with this new power constraint show smaller but still noticeable improvement of the SII in different noisy car contexts. Besides, our results permit to highlight the functional properties of the previous approaches based on approximations [6, 11, 12]. Then subjective listening tests confirm the efficiency of the proposed method in low frequency noises but also show its limitation for wide-band noises.

The paper is organized as follows. In section 2, we detail the optimization of the speech intelligibility index and the speech signal processing method. The subjective validation test protocol and results are discussed in section 3. We finally suggest some conclusions and perspectives of this work.

2. SPEECH INTELLIGIBILITY INDEX OPTIMIZATION

2.1. SII computation

The main hypothesis of the SII is that speech is composed of i^{max} frequency channels carrying independent information. The standard ANSI/ASA S3.5 [18] provides a computational method for several frequency band decompositions. In our study, the results for a third-octave band decomposition ($i^{max} = 18$) are presented but a critical band decomposition ($i^{max} = 21$) is also well suited. Equivalent spectrum levels expressed in decibels (dB) are obtained by integrating the periodogram of a signal on each sub-band, divided by its frequency bandwidth b_i. From the equivalent speech spectrum levels E_i and the equivalent noise spectrum levels N_i, two coefficients per sub-band are computed: the audibility and distortion coefficients.

The audibility coefficients represent the portion of audible spectrum above a so-called perturbation spectrum. The equivalent perturbation spectrum levels D_i are directly derived from the noise and the listener threshold of hearing, and thus they will be considered as constant. For normal-hearing listeners in a noisy environment, the D_i are often equal to the equivalent masking spectrum levels Z_i obtained by applying a perceptual masking model to the equivalent noise spectrum levels $\{N_j\}_{j\leq i}$. The audibility coefficients A_i are then computed as follows:

$$\mathrm{A}_i(E_i) = \min(\max(\frac{E_i - (D_i - 15\,\mathrm{dB})}{30\,\mathrm{dB}}, 0), 1)\,. \tag{1}$$

The distortion coefficients L_i represent the loss of intelligibility that occurs when the equivalent speech spectrum levels are excessively greater than some average speech reference levels U_i given by the standard ANSI/ASA.

$$\mathrm{L}_i(E_i) = \min(1 - \frac{E_i - (U_i + 10\,\mathrm{dB})}{160\,\mathrm{dB}}, 1)\,. \tag{2}$$

All sub-bands do not carry the same average amount of speech information. This is characterized by a Band Importance Function (BIF) which is applied to weight each sub-band. Several BIF are suggested in the standard depending on the speech material used. In this work, we select the default function which was designed for average speech. The SII is then a weighted sum of all these coefficients :

$$\mathrm{SII}(\{E_i\}) = \sum_{i=1}^{i^{max}} f_i(E_i)\,, \tag{3}$$

with
$$f_i(E_i) = I_i \cdot \mathrm{A}_i(E_i) \cdot \mathrm{L}_i(E_i)\,. \tag{4}$$

where I_i are the BIF weighting coefficients.

2.2. SII Optimization

In order to take into account the varying sensitivity of the human ear to different frequencies we use the dB(A) weighting coefficients H_i for estimating the signal level as follows :

$$E^{dBA} = 10 \cdot \log(\sum_i g_i(E_i))\,, \tag{5}$$

with
$$g_i(E_i) = b_i \cdot 10^{(E_i + H_i)/10}\,. \tag{6}$$

Let E_0^{dBA} be the reference level, the optimization problem can then be written as:

$$\{E_i^{opt}\} = \underset{\{E_i\}}{\arg\max} \sum_i f_i(E_i)\,, \tag{7}$$

subject to
$$\sum_i g_i(E_i) = G = 10^{E_0^{dBA}/10}\,. \tag{8}$$

The exponential functions g_i are continuous and convex on $\mathbb{R}$ and it can be shown that the f_i are:

- constant and minimal on $]-\infty, d_i^- = D_i - 15\,\mathrm{dB}]$ so either $E_i^{opt} > d_i^-$ either the i^{th} sub-band is deactivated (e.g. $E_i^{opt} = -\infty$),
- decreasing on $[d_i^+ = D_i + 15\,\mathrm{dB}, +\infty[$ so $E_i^{opt} \leq d_i^+$,
- continuous and concave on $[d_i^-, d_i^+]$,
- non-differentiable on $d_i^u = U_i + 10\,\mathrm{dB}$ if $d_i^- < d_i^u < d_i^+$.

Let us note Ω_2 the set of sub-bands where $d_i^- < d_i^u < d_i^+$. These sub-bands have two search intervals where f_i is concave and differentiable: $[d_i^-, d_i^u]$ and $[d_i^u, d_i^+]$. The set Ω_1 composed of the remaining sub-bands have only one search interval where f_i is concave and differentiable: $[d_i^-, d_i^+]$. For each sub-band of both sets, either E_i^{opt} belongs to one of these search intervals, or $E_i^{opt} = -\infty$ which means that the sub-band must be deactivated. As a result we have $3^{|\Omega_2|} \cdot 2^{|\Omega_1|}$ sub-problems and we use the Lagrange multiplier search method [20] to solve everyone of them and select the best solution.

Each sub-problem is solved as follows. Let Ω_{deact} be the set of deactivated sub-bands, the remaining sub-bands composes the set $\Omega_{act} = (\Omega_2 \cup \Omega_1) \setminus \Omega_{deact}$ and their search interval are noted $[l_i, u_i]$ with $l_i \in \{d_i^-, d_i^u\}$ and $u_i \in \{d_i^u, d_i^+\}$. Let λ denote the Lagrange multiplier for equation 8, v_i for $E_i \geq l_i$, and w_i for $E_i \leq u_i$. The Karush-Kuhn-Tucker (KKT) conditions for the problem can be written as follows:

$$\sum_{i \in \Omega_{act}} g_i(E_i) = G\,, \tag{9}$$

$$\forall i \in \Omega_{act}\,, \quad l_i \leq E_i \leq u_i\,, \tag{10}$$

$$-f_i' + \lambda \cdot g_i' - v_i + w_i = 0\,, \tag{11}$$

$$v_i \cdot (l_i - E_i) = 0\,, \tag{12}$$

$$w_i \cdot (E_i - u_i) = 0\,, \tag{13}$$

$$v_i \geq 0\,, \tag{14}$$

$$w_i \geq 0\,. \tag{15}$$

For $i \in \Omega_{act}$ let $\overline{E_i}(\lambda)$ be the solution to $-f_i' + \lambda \cdot g_i' = 0$ i.e.:

$$\overline{E_i}(\lambda) = \begin{cases} 10 \cdot \log(\frac{I_i}{3 \cdot \ln 10 \cdot \lambda \cdot b_i}) - H_i & \text{if } d_i^u \geq u_i\,, \\ \frac{160 + d_i^u + d_i^-}{2} - \frac{10}{\ln 10}\,\mathrm{W}(\frac{24 \cdot \lambda \cdot b_i}{I_i / \ln^2(10)} 10^{\frac{2H_i + 160 + d_i^u + d_i^-}{20}}) & \text{if } d_i^u \leq l_i\,. \end{cases}$$

Note the use of the Lambert W function when $d_i^u \leq l_i$. The following equations satisfy all the KKT conditions except (9):

$$E_i(\lambda) = \begin{cases} l_i & \text{if } \overline{E_i}(\lambda) \leq l_i \\ \overline{E_i}(\lambda) & \text{if } l_i < \overline{E_i}(\lambda) < l_i \\ u_i & \text{if } \overline{E_i}(\lambda) \geq u_i \end{cases}\,, \tag{16}$$

$$v_i(\lambda) = \begin{cases} -f_i'(l_i) + \lambda \cdot g_i'(l_i) & \text{if } \overline{E_i}(\lambda) \leq l_i \\ 0 & \text{if } \overline{E_i}(\lambda) > l_i \end{cases}\,, \tag{17}$$

$$w_i(\lambda) = \begin{cases} 0 & \text{if } \overline{E_i}(\lambda) < u_i \\ f_i'(u_i) - \lambda \cdot g_i'(u_i) & \text{if } \overline{E_i}(\lambda) \geq u_i \end{cases}\,. \tag{18}$$

Therefore, the problem is solved by using an iterative procedure to identify a λ that yields $\sum_{i \in \Omega_{act}} g_i(E_i(\lambda)) = G$.

2.3. Voice processing

From the dB(A) level of a signal E_0^{dBA} we can compute the optimum equivalent spectrum levels $\{E_i^{opt}\}$ and design a frequency equalizer that aims to transform the long term equivalent speech spectrum into the optimum equivalent spectrum. For generalization purposes we consider that we do not have access to the speaker long term equivalent speech spectrum levels but only to normalized reference levels U_i' obtained from the signal dB(A) level as follows:

$$U_i' = U_i - U^{dBA} + E_0^{dBA}\,, \tag{19}$$

The computation of the equivalent equalization levels F_i is detailed in equation 20 and the resulting frequency equalizer is applied on the whole signal using a Short Time Fourier Transform (STFT) overlap-add method.

$$F_i = E_i^{opt} - U_i'\,. \tag{20}$$

For deactivated sub-bands ($E_i^{opt} = -\infty$) we fix $E_i^{opt} = -60\,\mathrm{dB}$.

2.4. Quantitative results analysis

We selected typical car noises recorded with an acoustic head (HMS IV, HEAD acoustics *GmbH*) in three different driving conditions: high-speed noise (HS), low-speed noise (LS) and low-speed noise with rain (LS+R). All the corresponding equivalent masking spectrum levels Z_i are available on figure 1. Due to space limitations we only show herein the results for the LS noise but all the following comments also apply on the two other noises. The SNR used thereafter is defined as the ratio between the level of the normalized reference speech and the level of the noise :

$$\mathrm{SNR} = 10 \cdot \log\left(\sum_i b_i \cdot 10^{U_i'/10} / \sum_i b_i \cdot 10^{N_i/10}\right). \qquad (21)$$

The optimum equivalent speech spectrum levels as a function of SNR can be seen on figure 2 for the LS noise. The optimum spectra seem to have logical patterns : for low SNRs, many sub-bands are deactivated (hatched on the figure) and all the energy is allocated to the high frequencies where the masking noise is the lowest. For higher SNRs, each sub-band is progressively activated as soon as there is enough energy to be redistributed above the masking noise.

We compare below our proposed optimal solution of the dB(A) constrained SII optimization problem to several existing methods based on the optimization of an approximated SII curve [6, 11, 12]. To that aim, we compute the SII obtained by all optimization approaches and compare them to the SII of an average speech spectrum for a wide range of SNR. The figure 3 shows the SII improvements according to the SNR in the LS noise for each solution.

As expected, all methods do increase the SII for nearly all SNR but the approximated methods do not perform well on the full SNR range. It can be noticed that Taal & al. approach [9] is close to our optimal approach for low SNR but is clearly less efficient at higher SNR while on the contrary the other methods [6, 12] are only nearly optimal at high SNR. Nevertheless, our proposed optimal solution is more costly since it is based on an exact solution of the direct SII optimization problem. Clearly, the Stanton & al. approach is an excellent alternative choice for typical in car applications since it is fast and nearly optimal for SNR above -35dB.

3. SUBJECTIVE VALIDATION TEST

3.1. Experimental Setup

The speech material is composed of 200 sentences from the Canadian French *Hearing In the Noise Test* (HINT) [21] and Fournier sentences recorded by the French *Collège National d'Audioprothèse*, pronounced by an unique male speaker. In order to have a speech material of equal difficulty in the considered noises we equalized the difficulty of sentences as suggested by Nielsen et al. [22] in a synthetic stationary noise (noted EQ) whose spectrum is an average of our in-car noises. The equalization has been done in two iterations with two groups of six subjects each. As a result of this equalization, a correction in dB is assigned to every sentence : difficult sentences are slightly amplified while the easy ones are slightly attenuated. Finally, 10 lists of 20 sentences have been created. Eight lists are composed of the sentences with the lowest correction (less than 3 dB) in a way that the mean variance is minimized. The other sentences are divided in two lists that will be used to train the subjects to the task.

Traditionally, the noise used in *Sentence Recognition in Noise* (SRN) test is synthesized from a *Long Term Average Speech Spectrum* (LTASS) from the speaker in order to have almost the same SNR in each sub-band but in our case, the noises are given by our applicative context (see figure 1). HS and LS are both stationary and are spectrally similar except that LS is slightly flatter; the main difference is their global level, respectively 92 dB and 87 dB. The presence of raindrop impacts adds high frequency noise components, that is why the LS and LS+R spectra are identical in the low frequencies but differ in the high frequencies. The presence of rain

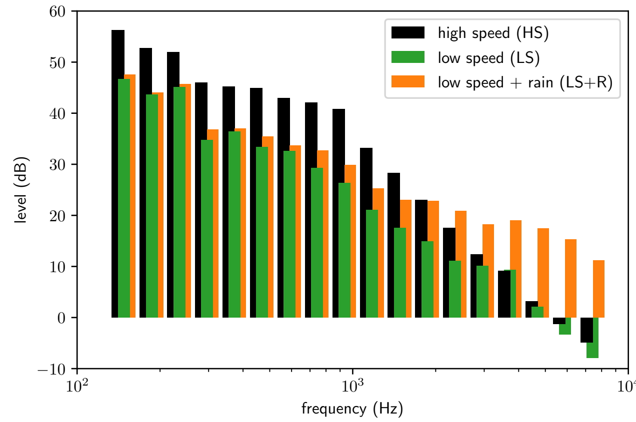

Fig. 1. Long term equivalent masking spectrum levels Z_i of three car noises.

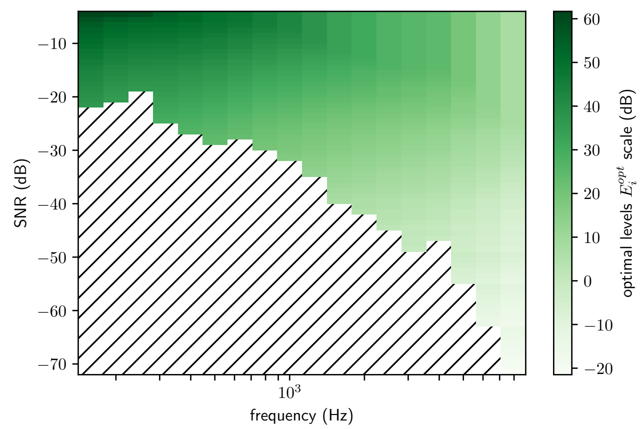

Fig. 2. Optimum equivalent speech spectrum levels E_i^{opt} as a function of SNR for the LS noise.

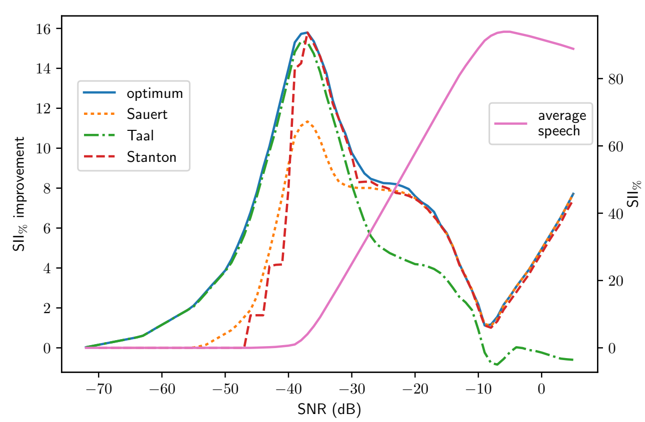

Fig. 3. SII improvement as a function of SNR for each optimization approach in the LS noise. The figure has twin y-axes: the first axis is related to the improvement and the second one is related to the reference average speech $\mathrm{SII}_{\%}$ (SII expressed in percent).

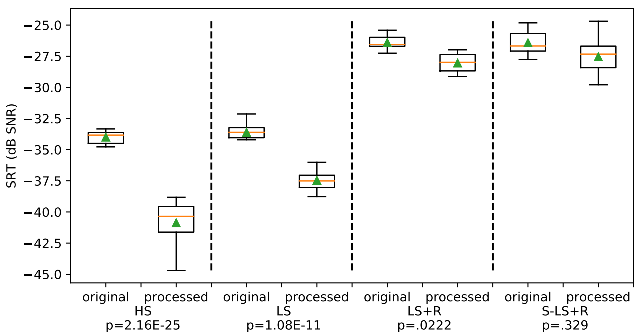

Fig. 4. Box plots of the mean SRT estimates.

also adds non-stationarity to the noise which may induce an uncontrolled bias in the perceptual tests. To assess the potential effect of non-stationarity, we generated a synthetic stationary noise (S-LS+R) from the same long term spectrum as LS+R. For the tests, all stimuli are played using a programmable equalizer (HEAD acoustics *GmbH* PEQ V) and calibrated headphones (Sennheiser^{TM} HD 650).

The *Speech Reception Threshold* (SRT) is the speech level that corresponds to 50% intelligibility and the adaptive process used for its estimation is the one from Brand et al. [23]. For a specific list, the first sentence is presented at low level and is increased until the listener is able to repeat at least one word. The 19 following sentences are presented only once and their level depends on the previous answer. If a subject repeats less (resp. more) than 50% of a sentence the next presentation level is increased (resp. decreased) proportionally to the gap. Also the more the level oscillates the smaller the step. The exact formula used is the one suggested in [23].

The repetition of one word from a sentence presented at a specific level is considered as an independent Bernoulli trial of probability expressed as in [23] by a sigmoid whose inflection point is characterized by two parameters : its slope and its center. At the end, the probability parameters are estimated using a maximum likelihood estimator on the resulting Bernoulli process and the SRT corresponds to the center of the sigmoid.

The tests have been conducted on 13 normal-hearing listeners whose hearing sensitivity has been verified by pure-tone audiometry. The subjects who participated to the equalization phase have not been part of the main SRN test.

First the SRT estimate is applied on the two training lists in the EQ noise so the subject gets used to the task. The SRT estimate is then conducted on the eight balanced lists: two for each noise either with the original voice or the processed voice. Every list, every sentence and every noise condition are presented in a pseudo-random order using a balanced latin square to avoid any order effect.

3.2. Results

The mean SRT estimates are displayed with box plots on figure 4. First it can be noticed that the SRT is improved by the processing for all situations. However, performances vary across noises as the SRT have a mean decrease of 6.9 dB for the HS noise and only 3.9 dB for the LS noise. For the LS+R and S-LS+R noises, the mean SRT decrease is even smaller with respectively 1.7 dB and 1.1 dB.

To assess the significance of the obtained result, we ran an *Analysis Of VAriance* (ANOVA) with repeated measure and the two fixed factors were the noise condition (4 possible values), the processing (2 possible values) and the interaction between these two. The results show an overall statically significant differences : noise (p = 4E-61), processing (p = 3E-29) and the interaction noise*processing (p=3e-17). A post-hoc analysis has been run on the interaction to confirm where the differences occurred and there are two main observations. First, the SRT improvement is significant in the HS and LS noises but not in the LS+R and S-LS+R ones (for a level of significance $\alpha = 0,01$), the p-values are on figure 4. Finally, the SRT in dB SNR is not significantly different between the HS and LS noises for the unprocessed speech while it is for the processed speech: the processing has better performance in the HS noise than the LS noise.

3.3. Discussion

In re-allocating the signal energy in low-noise sub-bands, we improve the perception of some frequency content at the expense of others but which are masked anyway. This process greatly improves the intelligibility when the noise spectrum is located in a specific area, here in the low frequencies for the HS and LS noises. But as soon as the noise spectrum becomes flatter or more wide-band, the improvement is less noticeable and no more significant. This result is easily understandable as there are no more noise free bands to re-allocate the energy in, it is then harder to efficiently filter the signal.

As shown in figure 3 the theoretical intelligibility improvement in the LS noise is maximum for SNRs around -37 dB specifically where the SRT was estimated. However these levels are lower than those encountered in real-life applications since they ask for a strong and sustained effort to understand the sentences. For SNR greater than -30 dB the theoretical range of improvement decreases as the SNR increases. The SRT improvement in more realistic in-car use cases may then be quite smaller. The SRT indicator may always be used in SRN tests but its improvement at low SNR does not prove its improvement at reasonable listening levels. A thorough validation of intelligibility improvement would need a different and specific subjective listening test at higher SNR for all the approaches which claim to improve intelligibility in noise.

4. CONCLUSION

We expressed the speech intelligibility enhancement task as a constant dB(A) constrained SII optimization problem for which an exact solution was proposed. We have shown that despite the new power constraint the SII optimization approaches still significantly improve the speech intelligibility in low-frequency in-car noises where it is indeed straightforward to reallocate the signal energy in noise free sub-bands : we observed a mean SRT improvement of 3.9 dB (resp. 6.9 dB) for a low-speed (resp. high-speed) car noise. However for noises with flatter spectrum the mean SRT improvements fall between 1.1 dB and 1.7 dB and are no more significant.

Solving the exact maximization problem of the SII also showed that for our car noises the approximations from earlier studies are nearly optimal in specific SNR ranges and less efficient in others. More precisely, some approximated methods are more efficient in low SNR and other in high SNR. Since these approaches have a lower computational complexity, our proposed method may be used to guide the choice of the best suited approximation for a given application.

We have also addressed a criticism of SRT-based validation procedures to assess intelligibility improvements of SNR-based processing that requires suitable subjective listening tests at reasonable listening levels.

Future work will be dedicated to the extension of the method for measures based on short time frames to allow for a dynamic local optimization to adapt to the mean energy variation between phonemes. Another perspective will be to combine our approach with dynamic compression.

5. REFERENCES

[1] R. Niederjohn and J. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," *IEEE Trans. on ASLP*, vol. 24, no. 4, pp. 277–282, 1976.

[2] I.V. McLoughlin and R.J. Chance, "Lsp-based speech modification for intelligibility enhancement," in *DSP Proc.*, 1997, vol. 2, pp. 591–594.

[3] M.D. Skowronski and J.G. Harris, "Applied principles of clear and lombard speech for automated intelligibility enhancement in noisy environments," *Speech Comm.*, vol. 48, no. 5, pp. 549–558, 2006.

[4] S.D. Yoo, J.R. Boston, A. El-Jaroudi, CC. Li, J.D. Durrant, K. Kovacyk, and S. Shaiman, "Speech signal modification to increase intelligibility in noisy environments," *J. ASA*, vol. 122, no. 2, pp. 1138–1149, 2007.

[5] H. Brouckxon, W. Verhelst, and B.D. Schuymer, "Time and frequency dependent amplification for speech intelligibility enhancement in noisy environments," in *Interspeech*, 2008.

[6] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations," in *Eusipco*, 2010, pp. 1919–1923.

[7] Y. Tang and M. Cooke, "Energy reallocation strategies for speech enhancement in known noise conditions," in *Interspeech*, 2010.

[8] K. Gibak and P.C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE Trans. on ASLP*, vol. 18, no. 8, pp. 2080–2090, 2010.

[9] C.H. Taal, R.C. Hendriks, and R. Heusdens, "A speech preprocessing strategy for intelligibility improvement in noise based on a perceptual distortion measure," in *ICASSP*, 2012.

[10] TC. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Interspeech*, 2012.

[11] C.H. Taal, J. Jensen, and A. Leijon, "On optimal linear filtering of speech for near-end listening enhancement," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 225–228, 2013.

[12] R. Stanton, N.D. Gaubitch, P. Naylor, and M. Brookes, "A differentiable approximation to speech intelligibility index with applications to listening enhancement," in *AES Conference on Audio Forensics*, 2014.

[13] K. Nathwani, G. Richard, B. David, P. Prablanc, and V. Roussarie, "Speech intelligibility improvement in car noise environment by voice transformation," *Speech Comm.*, vol. 91, pp. 17–27, 2017.

[14] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility-enhancing speech modifications: the hurricane challenge.," in *Interspeech*, 2013, pp. 3552–3556.

[15] R. L. Goldsworthy and J. E. Greenberg, "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," *J. ASA*, vol. 116, no. 6, pp. 3679–3689, 2004.

[16] J.M. Kates and K.H. Arehart, "Coherence and the speech intelligibility index," *J. ASA*, vol. 117, no. 4, pp. 2224–2237, 2005.

[17] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. on ASLP*, vol. 19, no. 7, pp. 2125–2136, 2011.

[18] ANSI, "S3. 5-1997, methods for the calculation of the speech intelligibility index," *NY: ANSI*, vol. 19, pp. 90–119, 1997.

[19] K.S. Rhebergen, N.J. Versfeld, and W.A. Dreschler, "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," *J. ASA*, vol. 120, no. 6, pp. 3988–3997, 2006.

[20] K.M. Bretthauer and B. Shetty, "The nonlinear knapsack problem–algorithms and applications," *European J. of Operational Research*, vol. 138, no. 3, pp. 459–472, 2002.

[21] V. Vaillancourt, C. Laroche, C. Mayer, C. Basque, M. Nali, A. Eriks-Brophy, S.D. Soli, and C. Giguère, "Adaptation of the hint (hearing in noise test) for adult canadian francophone populations," *Int. J. Audiology*, vol. 44, no. 6, pp. 358–361, 2005.

[22] J.B. Nielsen and T. Dau, "Development of a danish speech intelligibility test," *Int. J. Audiology*, vol. 48, no. 10, pp. 729–741, 2009.

[23] T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *J. ASA*, vol. 111, no. 6, pp. 2801–2810, 2002.