AN EXPERIMENTAL FRAMEWORK FOR THE DERIVATION OF PERCEPTUALLY-OPTIMAL NOISE SUPPRESSION FUNCTIONS

Adrien Daniel[1]*, Ludovick Lepauloux*[2]*, Christelle Yemdji*[1]*, Nicholas Evans*[1] *and Christophe Beaugeant*[2]

[1]EURECOM, Multimedia Department, Sophia-Antipolis, France
[2]Intel, Mobile and Communications Group, Sophia-Antipolis, France
firstname.lastname@{eurecom.fr,intel.com}

ABSTRACT

This paper presents a novel experimental framework designed to derive, through subjective testings, noise suppression functions which are perceptually optimal under specific experimental conditions. Noisy speech sequences are continuously processed according to a gain curve function of the *a priori* SNR that listeners are required to adjust two points at a time with respect to specified perceptual criteria. An experiment based on this framework is reported testing one specific combination of speech and noise signals. The specified perceptual criterion was the suitability for a phone conversation. The resulting mean experimental gain function shows a statistically significant deviation from an ideal Wiener filter.

Experiments based on this framework are repeatable, suit untrained listeners and are considerably faster than conventional subjective testing methods, without the necessity to place restrictive assumptions on the assessed noise suppression function.

Index Terms— Subjective testing, auditory perception, noise reduction, speech enhancement, Wiener filter

1. INTRODUCTION

In the context of speech enhancement, noise reduction involves the recovery of a speech sequence from a recording corrupted with additive noise. Most approaches work in the spectral domain and involve: (i) the estimation of the noise power spectral density (PSD), and (ii) the attenuation of noise according to an estimated gain function. This paper is concerned only with the latter.

Several theoretically derived gain functions have been proposed [1]. Among the most popular are spectral subtraction [2], Wiener filtering [3], Bayesian Estimators [4, 5], and subspace-based methods [6]. While applied to human communications, such approaches do not always reflect subjective preferences.

An extensive body of work [7–14], has investigated the influence of more perceptual aspects. These approaches, however, are generally based on a single aspect of sound perception: audibility of noise, and rely on the masking properties of the human auditory system. They do not address other aspects such as intelligibility or ease of listening, or investigate the trade-off between the tolerance of residual noise and speech distortion.

One recent exception [15] reports the estimation of preferred noise suppression functions for users of cochlear implants. Participants were required to listen to a noise-corrupted speech signal processed in real time according to a parametric Wiener filter whose parameters they adjusted to obtain optimal quality. Compared to previous perceptual approaches, where noise suppression functions are derived theoretically according to models of the human auditory system, such entirely subjective approaches tend to reflect more reliably true perceptual preferences. In the described experiments, however, estimated gain functions are restricted to those defined by a given mathematical expression—only parameters in such expressions are freely explored. In addition, subjective testing is notoriously time consuming and expensive.

This paper presents a flexible framework to subjective testing which aims to estimate perceptually-optimal gain functions without any restrictive assumptions of their form. Though perhaps somewhat trivial in its nature, the new approach is thoroughly justified, allows for the estimation of gain functions in a rapid and repeatable fashion and ensures a satisfactory level of statistical significance from tests with a relatively small dataset of speech signals and number of untrained listeners. This is illustrated in Section 3 which reports an experiment based on this framework.

2. FRAMEWORK

The following describes the general principle behind the new approach, some constraints which allow for each listening test to involve only two degrees of freedom, and the full experimental procedure used to estimate a perceptually-optimal gain function.

2.1. Principle

The framework described here aims to estimate via a method of adjustment a set of noise suppression or gain functions which are perceptually optimal for the respective experimental conditions. Gain functions are estimated for a given set of experimental conditions, defined in this paper by a specific noise type, the speaker gender and the average signal-to-noise ratio (SNR). In any given test involving one experimental condition, participants are required to optimize a gain curve (see Figure 1, left window) according to some prescribed perceptual criteria. Each test is composed of a number of trials in which speech signals corrupted by noise are heard repeatedly while subject to real-time filtering according to the current gain curve. Via a number of given points, the gain curve is adjusted by the participant to obtain an output speech which is 'best' according to the specified criteria. Finally, test results are averaged across several participants to obtain a listener-independent, perceptually optimized gain curve for the given experimental condition.

However, crucial to the new approach proposed here and to reduce bias as much as possible, participants do not adjust the gain curve directly, but instead by using an indirect, blind, adjustment method, with the two-dimensional graphical user interface (GUI) depicted in Figure 1 (right window). In each trial, each of the two dimensions in this triangle area is mapped to the gain value of one specific point on the curve. The positioning of the pointer in this

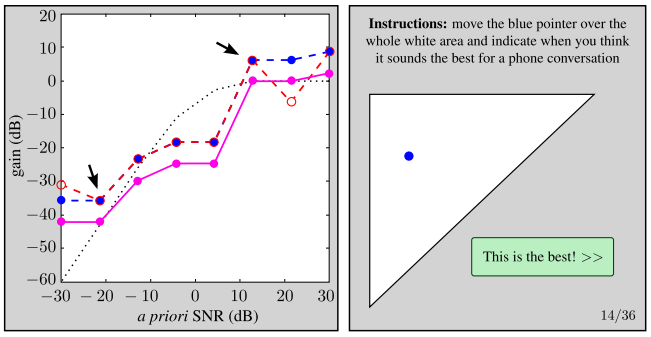

Fig. 1. Left window (hidden to the participant): gain curve under adjustment (dashed line with open ticks), monotonically increasing constrained gain curve (dashed line with filled ticks), level-corrected gain curve (solid line with filled ticks) which is used at last to compute the gain filter, Wiener filter (dotted line). The two arrows point to the pair of gain points under adjustment. Right window: GUI manipulated by the participant. The pointing area is on a dB scale.

area is thus equivalent to the adjustment of gains associated with two specific SNR values. When the participant considers a given pointer position as best-matching the specified perceptual criteria, they validate their selection and the next trial is presented. The pairs of gain points considered in each trial are selected randomly while ensuring that all points are adjusted an equal number of times.

Since this approach involves the adjustment of only two points at a time, it is far less complicated than the simultaneous adjustment of every point and is thus suitable to untrained listeners. The approach also reduces preconceived ideas regarding the profile of the gain curve, since the underlying noise suppression process is hidden. As described next, however, subjective testing in this way is slightly more troublesome than might appear at first. In practice, various constraints are necessary to ensure repeatable and consistent results.

2.2. Processing and constraints

The filtering process is implemented in the spectral domain using the short time Fourier transform (STFT) and overlap-add linear convolution [16]. Assuming that noisy speech signals result from additive noise:

$$Y_{(k,m)} = X_{(k,m)} + D_{(k,m)}, \tag{1}$$

where X is the clean speech signal and D is the noise signal. k and m are the STFT frequency bin and temporal frame indexes respectively.

Let $G(\xi)$, a function of the *a priori* SNR ξ, be a real-valued gain function resulting from the current adjustment of the gain curve by a participant, and let the filtering output be given by:

$$\hat{X}_{(k,m)} = G(\xi_{(k,m)})Y_{(k,m)}, \text{ with } \xi_{(k,m)} = \frac{|X_{(k,m)}|^2}{|D_{(k,m)}|^2}. \tag{2}$$

ξ can either be exact or estimated using state-of-the-art noise estimation algorithms [1, 17], depending on the objectives of the testing under conception.

During its adjustment by a participant, two constraints are imposed on the gain curve described by $G(\xi)$.

Monotonicity constraints: first, since the lower the SNR in a given frequency bin the higher the relative noise energy, and therefore the higher should be the associated suppression, the gain curve is forced to be monotonically increasing. Let us denote the N-point gain curve under adjustment by a set of couples $(\xi_p, g_p), p \in \mathbb{P} = \{1, ..., N\}$, where ξ_p is the *a priori* SNR associated with point p, and g_p its corresponding gain value, and such that $\forall p, q \in \mathbb{P}, p < q \Leftrightarrow \xi_p < \xi_q$. Let us also define $p_1, p_2 \in \mathbb{P}$ as the specific pair of points under adjustment, such that $p_1 < p_2$. Restriction $g_{p_1} \leq g_{p_2}$ is physically ensured by preventing participants from moving the pointer in the area where $g_{p_1} > g_{p_2}$, hence the triangular shape of the exploratory area in Figure 1 (right window). Each update involving point p_1 and p_2 is then made subject to the following monotonicity constraints:

$$\begin{cases} \forall p \in \mathbb{P} \mid p < p_1,\ g_p \leftarrow \min(g_p, g_{p_1}), \\ \forall p \in \mathbb{P} \mid p > p_2,\ g_p \leftarrow \max(g_p, g_{p_2}), \\ \forall p \in \mathbb{P} \mid p_1 < p < p_2,\ g_p \leftarrow \min(g_{p_2}, \max(g_p, g_{p_1})). \end{cases}$$

The application of such update rules is illustrated in Figure 1 (left window) by the differences between the two dashed lines. The profile with open, red ticks corresponds to a gain curve under adjustment, without any constraints, whereas the dashed line with filled blue ticks corresponds to the constrained gain curve which ensures a monotonic increase. Notice that as a consequence of these rules, while only two points p_1 and p_2 are freely adjusted at a time by the participant, in reality other points get indirectly adjusted accordingly when necessary. This is a key point to the efficiency of this method.

Energy conservation constraints: second, since otherwise participants would be free to adjust the output level by vertically offsetting the gain curve, the curve is forced to a certain vertical offset. Such a control of the output level is critical to a meaningful perceptual test, as for instance, some participants might otherwise unconsciously try to lower artifacts below hearing threshold by reducing gain values altogether. The resulting profile of the gain curve might not be optimal anymore for more realistic (higher) output levels such as in phone conversations. More generally, if we repeat an experiment, nothing guaranties that the participant will aim twice at the same output level. Hence for any given experimental condition, the output level must be a fixed parameter, and the offset of the gain curve must be adjusted accordingly. An important consequence is that the curves obtained from all participant are then optimally valid at the same output level and are thus comparable. The output level is determined by the total output energy resulting from the filtering process of equation (2), using equation (1):

$$\begin{aligned} E_{\text{out}} = \sum_{m=1}^{M}\sum_{k=1}^{K}|\hat{X}_{(k,m)}|^2 &= \sum_{m=1}^{M}\sum_{k=1}^{K}G(\xi_{(k,m)})^2|X_{(k,m)}|^2 \\ &+ \sum_{m=1}^{M}\sum_{k=1}^{K}G(\xi_{(k,m)})^2|D_{(k,m)}|^2 \\ &+ \sum_{m=1}^{M}\sum_{k=1}^{K}2G(\xi_{(k,m)})^2\text{Re}\{X_{(k,m)}D^*_{(k,m)}\}, \end{aligned}$$

where quantity $\sum_m \sum_k G(\xi_{(k,m)})^2|X_{(k,m)}|^2$ constitutes the useful output energy. The output level is controlled via an overall gain g_{offset} correcting the offset of the gain curve. It is desirable to let participants adjust independently the amount of residual noise in the output, however choosing g_{offset} so as to maintain constant the total output energy would mechanically result in reducing the proportion of useful energy in the output when increasing residual noise. The desired behavior is obtained by maintaining constant the useful output energy only. Let us call E_{useful} the targeted useful energy constant. $G(\xi)$ linearly interpolates on a log-scale the N-point gain function $\{(\xi_p, g_p)\}$, with g_1 and g_N acting as minimum and maximum threshold values, respectively. Going a step further in the

perception of the output signal, it is the loudness-weighted useful energy which is maintained constant via g_{offset}:

$$\sum_{m=1}^{M}\sum_{k=1}^{K}[W_k^{40}\times|X_{(k,m)}|]^2[g_{\text{offset}}\times G(\xi_{(k,m)})]^2 = E_{\text{useful}},$$

where W_k^{40} are equal loudness contour weighting coefficients defined according to ISO 226:2003 standard [18, 19] at 40 phons. This leads to:

$$g_{\text{offset}} = \sqrt{E_{\text{useful}}/\sum_{m=1}^{M}\sum_{k=1}^{K}[W_k^{40}\times|X_{(k,m)}|\times G(\xi_{(k,m)})]^2}, \quad (3)$$

which is continuously updated over modifications of the profile of the gain curve by the participant.

For computational reasons, however, SNR values $\xi_{(k,m)}$ are rounded to nearest decibel to form a smaller set of non-repeated integer values $\{\Xi_s, s = 1, ..., S\}$, where S depends on the former distribution of SNR values. The total weighted useful energy E_s associated with each Ξ_s is computed and stored beforehand:

$$E_s = \sum_{m=1}^{M}\sum_{k\in\mathbb{K}_m^s}(W_k^{40}\times|X_{(k,m)}|)^2,$$

where $\mathbb{K}_m^s$ are sets made of frequency bins such that $[\xi_{(k,m)}] = \Xi_s$. The overall gain is then real-time computed as (in decibels):

$$G_{\text{offset}} = 10\log_{10}\frac{E_{\text{useful}}}{\sum_{s=1}^{S}E_s\times G(\Xi_s)^2}, \quad (4)$$

and finally added to each point on the gain curve to give the offset-controlled gain curve illustrated in Figure 1 (left window) by the solid line with pink filled ticks. Because in practice S is very much smaller than $K \times M$, the computational cost is reduced in equation (4) compared to that of equation (3).

2.3. Procedure

For each experimental condition, each participant optimizes a separate gain curve for R realizations of noisy speech sharing the same characteristics. Such repeated measures of the same experimental condition reduce intra-listener measurement errors. Hence, the study of C experimental conditions implies the adjustment by each participant of $R \times C$ curves.

A single test is composed of a set of trials whose results produce a single one of the $R \times C$ gain curves for one participant. Initially, all curves are set to a flat profile (i.e., $\forall p, q \in P, g_p = g_q$). Each trial involves the adjustment of a single pair of points (p_1, p_2) affecting the profile of the curve. The set of trials involves a number of adjustments Q to any one point made simultaneously with another randomly chosen point. The order of each test is randomized while ensuring that any one point can be presented again only when all other points have been presented at least the same number of times. Hence after each whole adjustment of the gain curve, the participant has a new opportunity to correct the current value of each gain point, until reaching Q adjustments for all points. Finally, the $R \times C$ tests are randomly interleaved when presented to the given participant.

In addition, the initial position of the pointer in Figure 1 (right window) is randomized in each trial and the association of each axis in the triangle to gain points under adjustment is randomized; either the horizontal and vertical axes indicate increasing values of g_{p_1} and g_{p_2} respectively, or they indicate decreasing values of g_{p_2} and g_{p_1} respectively.

Such a robust experimental protocol is essential to give reliable gain curve estimates unbiased by order effects or current gain optima.

3. EXPERIMENTS

In the following we report an illustrative experiment which aims to test the proposed framework for a specific experimental condition. Results are compared to the noise suppression function of an ideal Wiener filter.

Fifteen participants volunteered for this experiment (four female, eleven male, aged 24-32). While two of them are authors of this paper, the remainder are naïve listeners not familiar with signal processing research. They received a 5-minute training session to familiarize themselves with the GUI before conducting the experiment. They were asked to select a pointer position in each trial which corresponded to their preferred speech characteristics for a supposed telephone conversation; they were not instructed to seek a noise-free output, but were instead asked to judge for themselves the trade-off between residual noise and speech distortion. Finally, they were asked to spend approximately 30 seconds exploring the triangular area in the GUI before making a final decision.

3.1. Setup

Experiments were conducted with three phonetically-balanced sentences (Harvard sentences) [20] of about 2–3 s from a single male speaker. Noise signals were one of three white noise realizations. All signals were sampled at a rate of 32 kHz and analyzed using 50% overlapping Hann windows of 32 ms.

Both speech and noise signals were bandpass filtered (50Hz–14kHz), then normalized to -26 dBov using the '14kBP' filter and speech voltmeter defined in ITU-T Recommendation G.191 [21]. Noise signals were normalized using the RMS long-term level, whereas speech signals were normalized using the active speech level. Three noisy speech stimuli were formed by corrupting each of the three speech signals with a different noise signal. No scaling gains were applied before summation, resulting in an average SNR of 0 dB. The *a priori* SNR $\xi_{(k,m)}$ used in equation (2) was an exact value, obtained directly from the noise and clean speech signals.

Each of the three noisy speech signals constitutes a repetition of the same experimental condition (i.e., $C = 1$ and $R = 3$). Therefore, each participant was required to adjust three gain curves. The use of different speech utterances avoids the influence of phonetic content. Each gain point was presented $Q = 3$ times within each curve adjustment. Gain curves were composed of $N = 8$ points of SNR values evenly spaced between -30 dB and 30 dB (inclusive). Gain values were allowed to lie in the [-60 20] dB range and the GUI pointing area had a precision of 1 dB on both axes. In total, each participant was presented with $C \times R \times \frac{1}{2}N \times Q = 36$ trials.

The perceptually-optimal gain curves derived through this above experiment were to be compared to that of a conventional Wiener filter [3]. The Wiener filter is used widely for noise reduction, is also a function of the *a priori* SNR and thus constitutes a meaningful comparison. To ensure such comparison was valid, the targeted constant of useful energy used in equation (4) was the useful output energy resulting from the Wiener filter H:

$$E_{\text{useful}} = \sum_{s=1}^{S}E_s\times H(\Xi_s)^2, \text{ with } H(\xi) = \frac{\xi}{\xi+1}.$$

SNR	-30	-21	-13	-4	4	13	21	30
Sig.	*	-	*	**	**	**	-	**

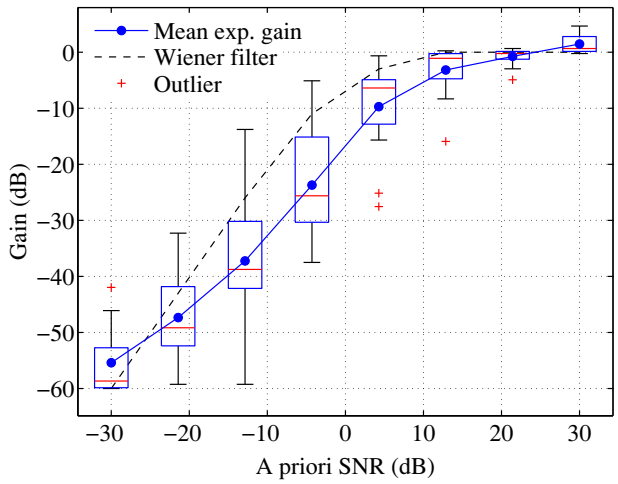

Fig. 2. Experimentally derived gain function with that of the Wiener filter. Boxplot whiskers represent the lowest/highest datum still within 1.5 times the interquartile range of the lower/upper quartile. Mean values are calculated from all subjects and repetitions. The upper table gives the statistical significance (after Bonferroni correction) of the deviation of the experimentally derived profile from the Wiener filter. Significance codes: *** <.001, ** .001–.01, * .01–.05.

All experiments were conducted in a quiet office. Stimuli were presented diotically with headphones[1] (Beyerdynamic DT 770 Pro) at an average output level of 65 dB SPL at each ear for the -26 dBov normalized clean speech stimulus alone.

3.2. Results and data analysis

Optimized gain values for each of the three speech utterances were averaged for each participant to form a single dataset on which statistical analyses were performed [22]. A mean experimental gain curve is also obtained by averaging gain values across the full number of participants, which reduces inter-listener measurement errors. The data distribution is plotted in Figure 2 (boxplots), together with the mean experimental curve (solid profile) and the Wiener filter for comparison (dashed profile). The experimental curve is more aggressive than the Wiener filter; it is mainly below, which indicates more suppression. This deviation from the Wiener filter is statistically significant, as discussed below.

Histograms and QQ-plots (not illustrated here) showed that data associated with gain points of -21, -13 and -4 dB SNR can reasonably be considered as normally distributed, while remaining data cannot. This non-normality of the data associated with extreme SNR values can probably be explained by the constraints imposed on the gain function (range and energy conservation). Therefore, depending on the normality of the data associated with each gain point, statistical significance of the deviation of the experimental curve from the Wiener filter was tested using either a parametric test (one-sample *t*-test), or a non-parametric test (one-sample Wilcoxon signed-rank test). p-values were corrected for multiple comparisons using Bonferroni correction. They are reported in the upper table of Figure 2 for each SNR value. As a result, the experimental mean curve deviates significantly from the Wiener filter, except for SNR values of -21 and 21, near to which the two curves intersect.

Outliers (i.e. markedly deviating values) were not removed because their origin is uncertain and cannot be proved to be measurement errors. It might be due to the non-normality of data associated with extreme SNR values (as highlighted above), but also to a mixture of distributions.

The value of gain function at g_1 where the SNR is equal to $\xi_1 = -30$ dB is somewhat peculiar. Below -30 dB SNR, certainly all speech components are masked and only noise is perceived. Therefore g_1 acts as a limiter in the gain curve. It essentially controls the amount of residual noise preferred when speech is absent. Informal listening tests revealed that, for the present setup, residual noise is barely audible when $g_1 = -40$ dB. Among all the adjusted curves from all participants (45 in total), only two preferred $g_1 \geq -40$ dB. Besides, we noted that after averaging the $R = 3$ repetitions for each participants, g_1 is systematically below -40 dB. This observation indicates the general preference to completely attenuate residual noise. The question is open as to whether or not this choice is dependent on the average SNR. Finally, informal listening of the noisy speech signals shows less musical noise in the resulting signal processed with the experimentally derived gain function than the Wiener gain function, perhaps at the expense of speech intelligibility, which was not a specific perceptual criterion in this experiment.

On average, each participant took about 25 minutes to complete the whole experiment. The ability of the proposed framework to show a statistically significant deviation with only 15 participants in such a short period of time demonstrates its strength. Besides, participants did not report any difficulty to understand the task or to perform it. A demonstration video together with sound samples are available on our website.[2]

4. CONCLUSION

This paper introduces an experimental framework for subjective testing and the derivation of perceptually-optimal noise suppression functions. We also report an example experiment using the proposed approach for a specific experimental condition (male speech corrupted with white noise). Listeners indicated a preferred noise suppression function which deviates in a statistically significant sense from an ideal Wiener filter. Thus, while a Wiener filter is optimal in the minimum mean-square-error sense, it is not necessarily optimal in a perceptual sense. Further work is ongoing to extend this study to assess the variability of the perceptually-optimal gain function for different noise types and average SNRs. With a wide range of results for different conditions, noise reduction algorithms could be adapted to switch between specific gain curves to the most appropriate for the prevailing noise conditions.

Carrying out such studies by means of conventional speech quality assessment methods (such as MOS [23] or MUSHRA [24] tests) would demand the comparison of hundreds or thousands of preprocessed speech signals or the restriction of considered gain profiles to specific forms. In contrast, listeners can conduct a complete experiment in less than 30 minutes using our approach and gain profiles are free from any restrictive assumptions.

[1] This corresponds to an headset phone situation. A monophonic presentation (handset situation) might yield different results.

[2] `http://audio.eurecom.fr/content/media`

5. REFERENCES

[1] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[3] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1218–1234, July 2006.

[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. 32, no. 6, pp. 1109–1121, Dec. 1984.

[5] P. J. Wolfe and S. J. Godsill, "Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement," in *11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Aug. 2001, pp. 496–499.

[6] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 334–341, July 2003.

[7] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 126–137, Mar. 1999.

[8] Y. Hu and P. C. Loizou, "A perceptually motivated approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 457–465, Sept. 2003.

[9] T. S. Gunawan, E. Ambikairajah, and J. Epps, "Perceptual speech enhancement exploiting temporal masking properties of human auditory system," *Speech Communication*, vol. 52, no. 5, pp. 381–393, May 2010.

[10] S. Gustafsson, P. Jax, and P. Vary, "A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 1998, vol. 1, pp. 397–400.

[11] P. J. Wolfe and S. J. Godsill, "Towards a perceptually optimal spectral amplitude estimator for audio signal enhancement," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, June 2000, vol. 2, pp. 821–824.

[12] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 857–869, Sept. 2005.

[13] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 700–708, Nov. 2003.

[14] C. H. You, S. N. Koh, and S. Rahardja, "An invertible frequency eigendomain transformation for masking-based subspace speech enhancement," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 461–464, June 2005.

[15] S. J. Mauger, P. W. Dawson, and A. A. Hersbach, "Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction," *The Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 327–336, Jan. 2012.

[16] C. Yemdji, M. I. Mossi, N. Evans, and C. Beaugeant, "A scalable frequency domain-based linear convolution architecture for speech enhancement," in *17th International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–6.

[17] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[18] International Organization for Standardization, "BS ISO 226:2003 - Acoustics - Normal equal-loudness-level contours," Sept. 2003.

[19] Y. Suzuki, V. Mellert, U. Richter, H. Mller, L. Nielsen, R. Hellman, K. Ashihara, K. Ozawa, and H. Takeshima, "Precise and full-range determination of two-dimensional equal loudness contours," *Tohoku University, Japan*, Mar. 2003.

[20] IEEE Subcommittee on Subjective Measurements, "IEEE recommended practices for speech quality measurements," *IEEE Transactions on Audio and Electroacoustics*, vol. 17, pp. 225–246, Sept. 1969.

[21] International Telecommunication Union, "ITU-T recommendation g.191 - Software tools for speech and audio coding standardization," Mar. 2010.

[22] D. C. Howell, *Statistical Methods for Psychology*, Wadsworth Cengage Learning, 2012.

[23] International Telecommunication Union, "ITU-T recommendation p.800 - Methods for subjective determination of transmission quality," Aug. 1996.

[24] International Telecommunication Union, "ITU-R recommendation BS.1534 - Method for the subjective assessment of intermediate quality levels of coding systems," Jan. 2003.