24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

VARIANCE OF SPECTRAL ENTROPY (VSE): AN SNR ESTIMATOR FOR SPEECH ENHANCEMENT IN HEARING AIDS

Fangqi Liu and Andreas Demosthenous

University College London, Department of Electronic and Electrical Engineering, London, UK.
email: fangqi.liu.14@ucl.ac.uk

Ifat Yasin

University College London, Department of Computer Science, London, UK.

In everyday situations an individual can encounter a variety of acoustic environments. For an individual with a hearing aid following speech in different types of background noise can often present a challenge. For this reason, estimating the signal-to-noise ratio (SNR) is a key factor to consider in hearing-aid design. The ability to adjust a noise reduction algorithm according to the SNR could provide the flexibility required to improve speech intelligibility in varying levels of background noise. However, most of the current high-accuracy SNR estimation methods are relatively complex and may inhibit the performance of hearing aids. This study investigates the advantages of incorporating a spectral entropy method to estimate SNR for speech enhancement in hearing aids; in particular a variance of spectral entropy (VSE) measure. The VSE approach avoids some of the complex computational steps of traditional statistical-model based SNR estimation methods by only measuring the spectral entropy among frequency channels of interest within the hearing aid. For this study, the SNR was estimated using the spectral entropy method in different types of noise. The variance of the spectral entropy in a hearing-aid model with 10 peripheral frequency channels was used to measure the SNR. By measuring the variance of the spectral entropy at input SNR levels between -10 dB to 20 dB, the relationship function between the SNR and the VSE was estimated. The VSE for the speech-in-noise was measured at temporal intervals of 1.5s. The VSE method demonstrates a more reliable performance in different types of background noise, in particular for low-number of speakers babble noise when compared to the US National Institute of Standards and Technology (NIST) or Waveform Amplitude Distribution Analysis (WADA) methods. The VSE method may also reduce additional computational steps (reducing system delays) making it more appropriate for implementation in hearing aids where system delays should be minimized as much as possible.

Keywords: Spectral entropy, signal-to-noise ratio, speech enhancement, hearing aids

1. Introduction

The signal-to-noise ratio (SNR) is one of the most fundamental metrics in noise level estimation; it is defined as a power ratio of noise and speech. SNR estimation of speech in noisy environments has been investigated over decades (e.g. [1]–[6]). Robust estimated SNRs could improve the performance of speech enhancement algorithms [4], particularly in hearing aids. The SNR estimation not only influences the performance of noise reduction algorithms [7], [8] but could also be used to iden-

tify the preferred noise reduction strength for the listener, which in turn may improve listening comfort as well as speech intelligibility [9], [10].

In general, SNR estimation methods can be separated into two categories: i) methods measuring the a-priori SNR according to the a-posteriori SNR focused on a relatively short time frame (approximately 40 ms) [5], [11] and ii) those focused on the average SNR across longer time frames (approximately 1 s). When compared with short-frame SNRs, the value of the average SNR across time frames has been shown to more accurately quantify the actual level of the non-stationary noise; the findings were also found to be more correlated with human speech perception [6]. One widely used method for estimating average SNR across time frames is the US National Institute of Standards and Technology (NIST) SNR measurement [12], which is based on a sequential Gaussian mixture estimation. NIST estimates the SNR according to the energy distributions of the signal and noise. The NIST method shows relatively reliable performance in different types of noise, but its SNR estimation accuracy is relatively low. Another SNR estimation method, the Waveform Amplitude Distribution Analysis (WADA) [4] assumes that the amplitude of speech and noise follow Gamma and Gaussian distributions. The WADA method measures the SNR by estimating the shaping parameter of the Gamma distribution, which is affected by the noise level. The algorithm shows good performance in stationary noise, but the computation of the parameter is relatively complicated. Recently, a deep neural network (DNN) based SNR estimation approach has been published by Papadopoulos *et al*. [3]. The SNR is estimated by using feature-trained models. Although the DNN approach shows high SNR estimation accuracy with different types of noise, it cannot be directly applied to hearing aids due to its high computational complexity. In hearing aids, the signal processing delay should be ideally minimized, since long system delays can reduce speech intelligibility [13]. The computational steps in most traditional SNR estimation methods (e.g. the NIST method) are complex and may introduce large system delays within the hearing aid.

This study focuses on developing an SNR estimation algorithm with high computation efficiency and relatively good accuracy that could be used in a hearing aid device. The algorithm is designed to deal with non-stationary noise such as babble noise (the noise generated by multiple speakers). The spectral entropy SNR estimation method is based on the fact that the spectral entropy of clean speech and the undesired noise are often very different (e.g. [14], [15]). Thus, the average SNR across time frames could be evaluated according to the respective changes in spectral entropy. However, unlike traditional spectral entropy based methods, which calculate spectral entropy either by using a fast Fourier transform (FFT) applied across short time frames [14] or apply a large number of filtering channels [15], the proposed method measures spectral entropy across a select number of peripheral filtering channels within the hearing aid. Most of the key information required to understand speech is encoded in the spectrum of the speech signal [16]. According to the Shannon information theory [17], the speech spectrum should have a lower entropy than the noise. Increasing the background noise level will corrupt the spectrum of the speech embedded in the noise. Thus, the higher the noise level, the flatter the spectrum of the noisy speech [18]. However, the value of instantaneous spectral entropy is not a reliable metric to be used for SNR estimation, since the spectral entropy of phonemes differs from one another and speech phonemes are connected with silence pauses [19]. The variance of the spectral entropy across time frames depends on the SNRs within frames. Therefore, it is more reliable to use the variance of the spectral entropy (VSE) among speech frames to represent the SNR level. If the VSE in any particular length of speech uniquely maps a particular SNR level, the SNR of an unknown portion of speech could be obtained via the estimated VSE/SNR relationship function; this could form the basis of a hearing-aid noise-reduction algorithm.

The rest of the paper is organized as follows: Section 2 describes the method of using VSE to estimate the SNR for a fixed number of frequency channels that could represent the first stage of processing in a hearing aid. Section 3 presents the results, the computation complexity and accuracy of the proposed method compared with the NIST and WADA approaches. Discussion and concluding remarks are presented in Section 4.

2. Method of SNR estimation

The basic idea of using the spectral entropy method to estimate the SNR is to establish the relationship function between the VSE and the SNR. Therefore, according to the simulated relationship function, the SNR level can be established using the measured VSE. If this approach is used within a hearing aid then the hearing aid could be configured to adjust a noise reduction algorithm depending on the estimated SNR. A flow chart showing how spectral entropy could be used to estimate the SNR in a given hearing aid is shown in Figure 1. The input speech signal is first processed by the peripheral filter bank. Then, the variance of the spectral entropy among channels is estimated. According to the simulated relationship function established for different types of noise, the SNR can be estimated using the measured VSE. The relationship functions can be stored in the hearing aid as a look-up table.

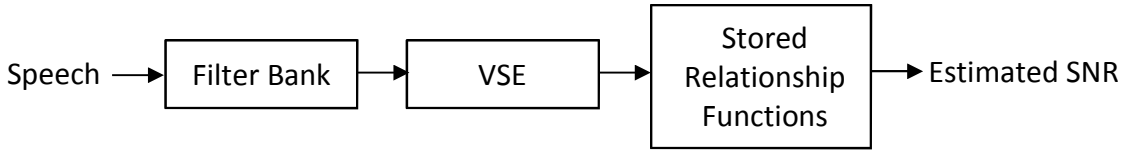

Figure 1: Flow chart of SNR estimation in a hearing aid.

2.1 Measuring the variance of spectral entropy

Unlike the method published by Shen *et al.* [14], which uses FFT to calculate the spectral entropy, the proposed method estimates the spectral entropy based on the signal of each peripheral filtering channel in a hearing aid. Since different frequency components of the speech are processed by different channels/filters (filter-bank) of the hearing aid, the signal level in each channel roughly represents the frequency response of the speech. To calculate the spectral entropy in hearing aids the first step is to obtain the probability density function (PDF) of each channel, that is:

$$p(t,i) = \frac{x(t,i)}{\sum_{i=1}^{n} x(t,i)} \tag{1}$$

where $p(t,i)$ is the PDF of the channel i at the time t, $x(t,i)$ is the amplitude of the signal in channel i at the time t, and n is the total number of the channels in the hearing aids. The corresponding entropy of each channel is:

$$h(t,i) = p(t,i) log_2 p(t,i) \tag{2}$$

where $h(t,i)$ is the spectral entropy of the channel i at the time t. The total spectral entropy at the time t is:

$$H(t) = \sum_{i=1}^{n} h(t,i) \tag{3}$$

However, the spectral entropy of a particular time t is not robust enough to reflect the SNR change of the speech in a time varying signal. In order to track the average SNR among frames, the VSE among frames is calculated:

$$v = E\left[\left(H - E(H)\right)^2\right] \qquad \text{where } H = \{H(1), H(2), H(3), \dots, H(L)\} \tag{4}$$

$$L = \frac{T}{fs} \tag{5}$$

where v is the VSE, L is the number of sampled points during the short time interval, T is the length of the time interval , and fs is the sampling rate.

2.2 Experiments setup

Four types of noise were used: pink noise, white noise, low-number of speakers babble noise (4-speaker babble noise) and high-number of speakers babble noise (24-speaker babble noise) were

evaluated in this study. The babble noise with fewer speakers is less stationary than with higher numbers of speakers [20]. 900 utterances spoken by 56 male and 56 female speakers from the AURORA resource database [21] were used in the present study. In all experiments, the root mean square level of speech was fixed at 60 dB to simulate the general speech level in real environments, while the noise level was increased from 40 dB to 70 dB with a step size of 1 dB to obtain the noisy speech (noise mixed with speech) at SNRs ranging between -10 dB and 20 dB. The time interval of both noise and speech utterances were fixed at 1.5 s. The sample rate was 44100 Hz. For testing, a computational model of the peripheral auditory filter bank (channels) for a hearing aid system was built using 2nd-order Butterworth bandpass filters. The channel number was set to 10 which is a number of frequency channels often used in hearing aids (hearing aids can have frequency channels numbering from 4-16) [22]). The characteristic frequency of each filter ranged from 250 Hz to 8000 Hz. This is the frequency range used in some hearing aids and covers the speech frequencies [23].

This study first evaluated the relationship function between VSE and SNR of an example utterance "2841", spoken by an adult male speaker. Then, the VSE/SNR relationship functions among 900 utterances were estimated in all four types of noise detailed above. Finally, the estimation accuracy of the spectral entropy-based SNR method was evaluated by measuring the mean absolute SNR estimation errors (MAEE). The testing results of estimation accuracy of all tested utterances were compared with the NIST and WADA methods.

3. Results

3.1 Relationship function (VSE vs. SNR) for the spoken utterance "2841"

Figure 2 (panels 2A, 2B, 2C) shows the spectral entropy of the example spoken utterance "2841" in 24-speaker babble noise at SNRs of -10 dB, 5 dB, and 20 dB, respectively. It can be seen that the spectral entropy values are very stable at an SNR of -10 dB (Figure 2A), but show large increased variation in spectral entropy at SNR levels of 5 dB and 20 dB (Figure 2C). Figure 2D shows the relationship function between VSE and the SNR. It can be seen that above about -5 dB, the VSE increases as the SNR increases.

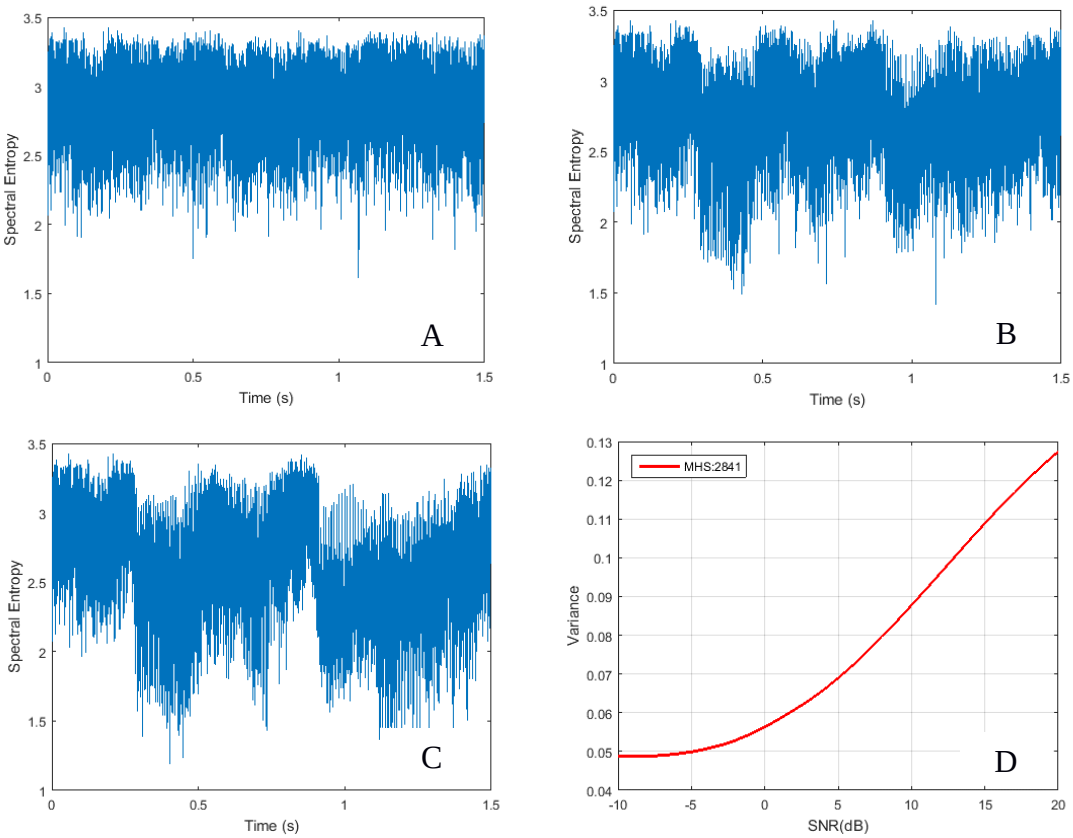

Figure 2: Spectral entropy of the utterance “2841” at SNRs of A: -10 dB; B: 5 dB; and C: 20 dB. D: Relationship function between VSE and the SNRs of utterance “2841” for SNRs ranging between -10 dB and 20 dB with a step size of 1 dB.

3.2 Simulating the relationship function among utterances

Figure 3 shows the relationship functions (variance vs. SNR) for 900 utterances in pink noise, white noise, 4-speaker babble noise, and 24-speaker babble noise. The average VSE among all utterances is always shown by the solid lines. Dashed lines represent the range of variation of the relationship function, which is obtained by using the mean value plus or minus the standard deviation of the VSE. It can be seen that at high SNR levels, the relationship function shows a large variation. In pink noise, white noise, and 24-speaker babble noise, the overall relationship function variation is less than for 4-speaker babble noise; the relationship function shows less stability across utterances in 4-speaker babble noise.

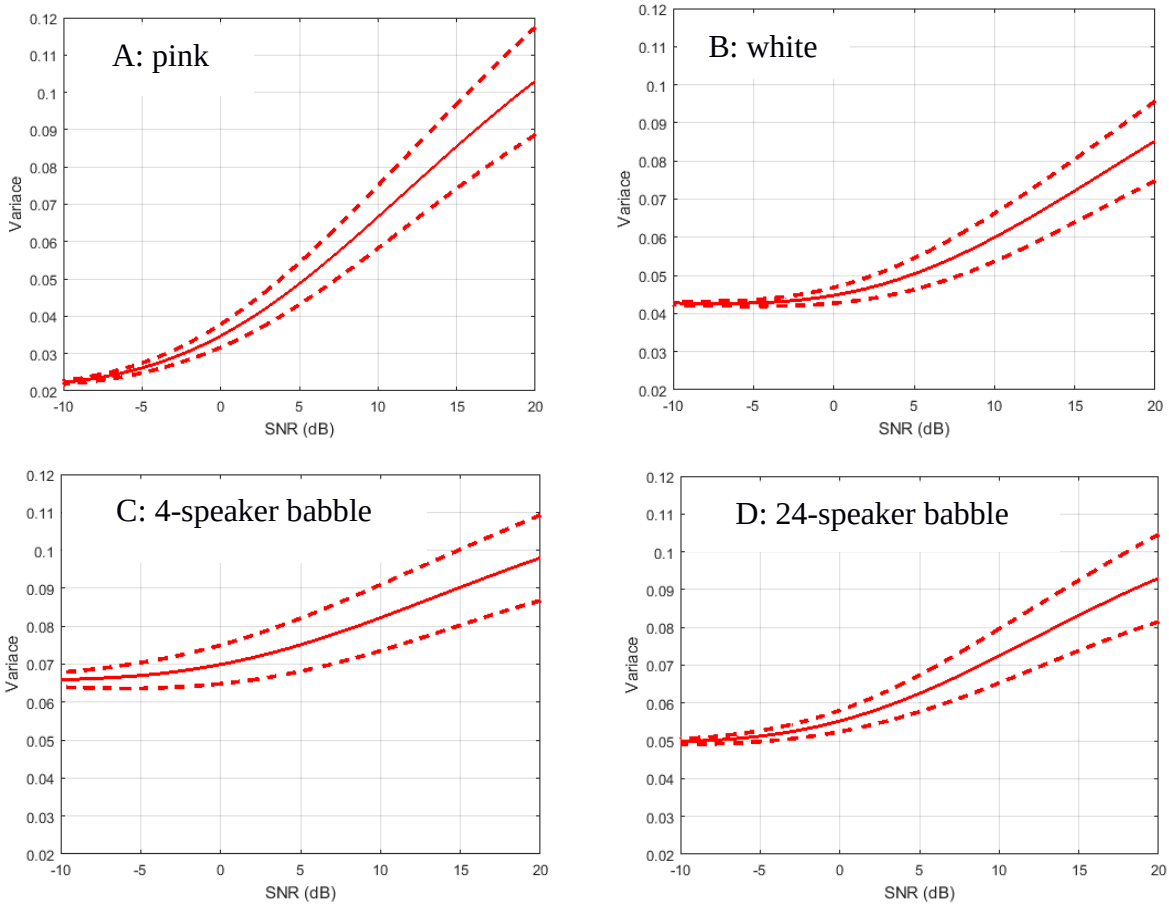

Figure 3: The mean value (solid line) and the standard deviation (dashed line) of the VSE-SNR relationship function in A: pink noise; B: white noise; C: 4-speaker babble noise; and D: 24-speaker babble noise for 900 speech utterances.

3.3 Computational complexity

For real-time signal processing, the spectral entropy needs to be computed sample by sample. In each channel of a hearing aid, the computation within channels could be processed in parallel. The VSE method needs two additive operations, one division, two multiplications, and one logarithm operation as shown in equations (1)-(3). The WADA based method needs a large group of sample points to make sure the amplitude of speech waveform follows the Gamma distribution. The DNN based method needs an additional memory unit to store the trained model and the size of the memory

should be relatively large to store all trained models. Therefore, when compared with other approaches, the computational cost of the VSE method would be lower. The VSE method may also result in reduced system delays when implemented in hearing aids compared to other methods. The spectral entropy method and estimation of the SNR from the VSE requires fewer computational steps.

Figure 4 shows the estimated SNRs using the VSE method plotted against the real (actual) SNR. Panels 4A, 4B, 4C and 4D show the results for pink noise, white noise, 4-speaker babble noise and 24-speaker babble noise. 200 randomly selected utterances from the all tested utterances are plotted. The highest estimation accuracy is observed for pink noise, white noise and 24-speaker babble noise. The lowest estimation accuracy can be seen for 4-speaker babble noise.

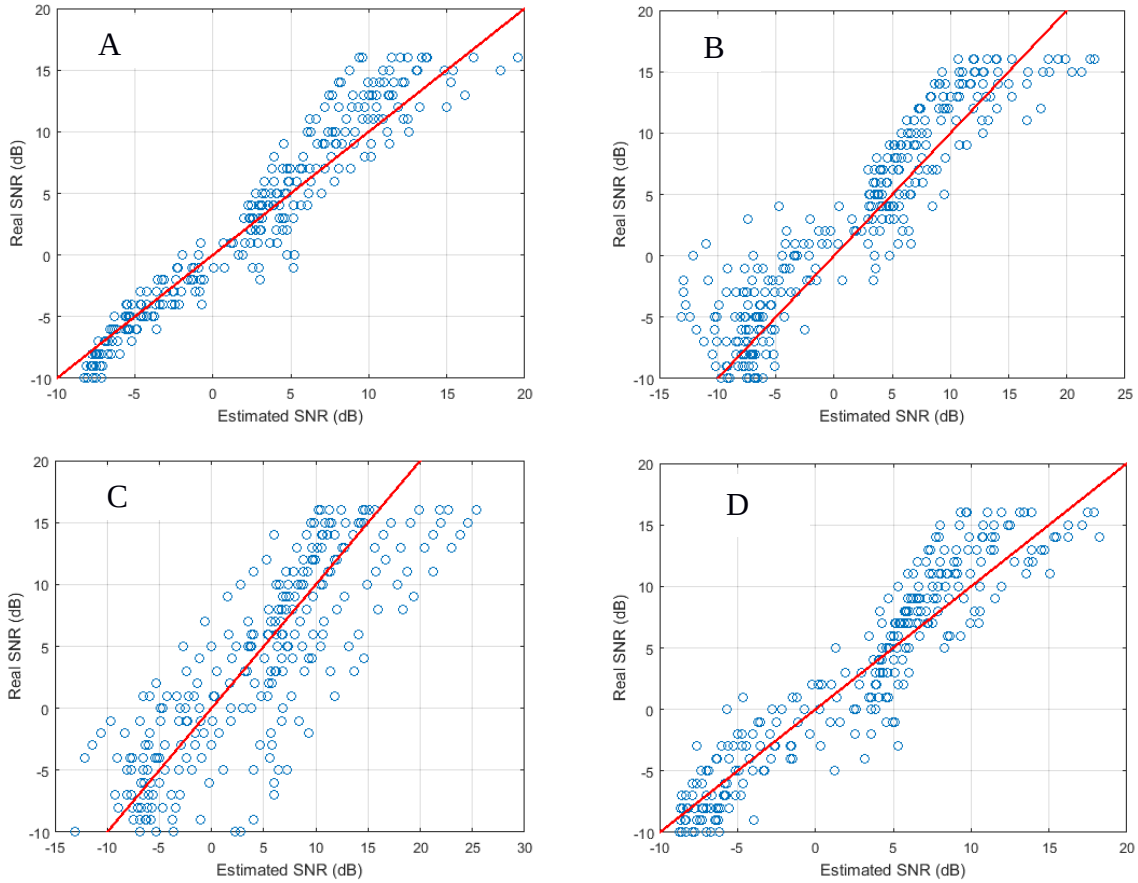

Figure 4: Estimated SNR along with the true SNR estimation line (in red) in A: pink noise; B: white noise; C: 4-speaker babble noise; and D: 24-speaker babble noise.

Table 1 shows the mean absolute SNR estimation errors (MAEE) for VSE, WADA and NIST methods when tested with different noises. The MAEE is the mean value of the absolute difference between the estimated SNR and the real SNR of testing utterances in all SNRs. The MAEE of the spectral entropy method is compared with that of published studies using the WADA and NIST methods [4], [3]. It can be seen that the VSE based method has higher accuracy than the NIST method in all types of tested noise. The reason for this may be that the NIST method relies on the energy of the noisy speech, but the energy change at the lower SNRs is very small and difficult to characterise. When compared with WADA, the VSE based method shows higher accuracy in low-number of speakers babble noise and pink noise. The amplitude of the babble noise may not follow a Gaussian distribution, but the measure of spectral entropy between speech and babble noise may still be very different. The WADA method shows slightly higher estimation accuracy than the VSE method in white noise and higher-number of speakers babble noise, but the slightly higher accuracy may be gained at the expense of computational efficiency. Also the precise number of speakers in the babble noise used in their study [3] is unknown so a direct comparison cannot be made. Overall, it appears

that spectral entropy using the VSE measure shows the potential to provide a more reliable performance in different types of background noise compared to the WADA and NIST methods.

Table 1: Mean absolute SNR estimation errors (MAEEs) with different noises

Type of noise	VSE	WADA	NIST
low-number of speakers babble noise	5.53 dB (4-speaker babble)	10.12 dB ○ (1-speaker)	15.34 dB ○ (1-speaker)
High-numbers of speakers babble noise	3.19 dB (24-speaker babble)	2.45 dB ◊	5.56 dB ◊
Pink noise	2.45 dB	2.77 dB ◊	5.31 dB ◊
White noise	4.00 dB	2.47 dB ◊	5.3 dB ◊
○: Data obtained from Figure 4a in paper [4]. ◊: Data obtained from Figure 2 in paper [3].			

4. Discussion and conclusion

In this paper a more computationally efficient method of estimating the SNR of noisy speech using a spectral entropy measure was introduced; the variance of spectral entropy (VSE). Since the spectral entropy of speech and noise are often very different, this method estimates the average SNR among frames by measuring the variance of the spectral entropy of the speech-noise mixture. The relationship functions of the variance vs. SNR in different noise types were estimated. The VSE method shows higher SNR estimation accuracy than the NIST method in all four types of noise (white noise, pink noise, low-number of speakers babble and high-number of speakers babble) and higher estimation accuracy than the WADA method in pink noise and low-number of speakers babble.

Any SNR estimation method implemented for use in hearing aids should focus particularly on performance robustness (in terms of SNR estimation) in different types of noise environments and also have low computation delays. The spectral entropy method using the VSE measure appears to provide an improved estimate of SNR in a variety of background noises when compared to the NIST or WADA methods. In addition, the VSE method has fewer computational steps (thereby reducing system delays) than the NIST and WADA methods, and may be more appropriate for implementation in hearing aids where system delays should be reduced as much as possible.

REFERENCES

1 P. Papadopoulos and A. Tsiartas, “A supervised signal-to-noise ratio estimation of speech signals,” *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 8287–8291, 2014.

2 S. Lee, C. Lim, and J. H. Chang, “A new a priori SNR estimator based on multiple linear regression technique for speech enhancement,” *Digit. Signal Process. A Rev. J.*, vol. 30, pp. 154–164, 2014.

3 P. Papadopoulos, S. Member, and A. Tsiartas, “Long-Term SNR Estimation of Speech Signals in Known and Unknown Channel Conditions,” vol. 24, no. 12, pp. 2495–2506, 2016.

4 R. M. Stern, “Robust signal-to-noise ratio estimation based on waveform amplitude dstribution analysis,” *Interspeech*, pp. 2598–2601, 2008.

5 S. Suhadi, C. Last, and T. Fingscheidt, “A data-driven approach to a priori SNR estimation,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 1, pp. 186–195, 2011.

6 M. Vondrášek and P. Pollák, “Methods for speech SNR estimation: Evaluation tool and analysis of VAD dependency,” *Radioengineering*, vol. 14, no. 1, pp. 6–11, 2005.

7 Y. Rao, I. S. Member, Y. Hao, I. M. S. Panahi, I. S. Member, and I. Fellow, “Smartphone-based Real-time Speech Enhancement for Improving Hearing Aids Speech Perception,” pp.

5885–5888, 2016.
8 I. Panahi, I. S. Member, N. Kehtarnavaz, I. Fellow, and L. Thibodeau, “Smartphone-Based Noise Adaptive Speech Enhancement for Hearing Aid Applications,” pp. 85–88, 2016.
9 T. Neher and K. C. Wagener, “Investigating Differences in Preferred Noise Reduction Strength Among Hearing Aid Users,” *Trends Hear.*, vol. 20, no. 0, pp. 1–14, 2016.
10 T. O. N. Neher, K. I. C. W. Agener, and M. A. M. Eis, “Relating hearing aid users preferred noise reduction setting to different measures of noise tolerance and distortion sensitivity,” no. August, 2015.
11 R. Yao, Z. Zeng, and P. Zhu, “A priori SNR estimation and noise estimation for speech enhancement,” *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, p. 101, 2016.
12 “The NIST Speech SNR Measurements.” [Online]. Available: https://www.nist.gov/information-technology-laboratory/iad/mig/nist-speech-signal-noise-ratio-measurements.
13 M. A. Stone and B. C. J. Moore, “Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay.,” *Ear Hear.*, vol. 24, no. 2, pp. 175–83, 2003.
14 J. Shen, J. Hung, and L. Lee, “Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments,” *5th Int. Conf. ICSLP ’98 Sydney, Aust.*, no. 1, p. 4, 1998.
15 B. F. Wu and K. C. Wang, “Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments,” *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 762–774, 2005.
16 J. H. and W. Holmes, *Speech Synthesis and Recognition, Second Edition*. 2003.
17 C. E. Shannon, “A Mathematical Theory of Communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
18 Y. Ma and A. Nishihara, “Efficient voice activity detection algorithm using long-term spectral flatness measure,” *EURASIP J. Audio, Speech, Music Process.*, vol. 2013, no. 1, pp. 1–18, 2013.
19 M. Gales and S. Young, “The Application of Hidden Markov Models in Speech Recognition,” *Found. Trends® Signal Process.*, vol. 1, no. 3, pp. 195–304, 2007.
20 N. Krishnamurthy and J. H. L. Hansen, “Babble Noise: Modeling, Analysis, and Applications,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 17, no. 7, pp. 1394–1407, 2009.
21 H. Hirsch and D. Pearce, “The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions,” *ASR2000- Autom. Speech Recognit. Challenges new Millenium*, p. 8, 2000.
22 E. W. Yund and K. M. Buckles, “Multichannel compression hearing aids: effect of number of channels on speech discrimination in noise.,” *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1206–1223, 1995.
23 Yong Lian, Ying Wei, Y. Lian, S. Member, and Y. Wei, “A computationally efficient nonuniform FIR digital filter bank for hearing aids,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 52, no. 12, pp. 2754–2762, 2005.
24 T. Jürgens, N. R. Clark, W. Lecluyse, and R. Meddis, “Exploration of a physiologically-inspired hearing-aid algorithm using a computer model mimicking impaired hearing,” *Int. J. Audiol.*, vol. 55, no. 6, pp. 346–357, 2016.