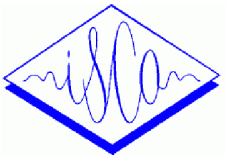
ISCA Archive
http://www.isca-speech.org/archive



SAPA-SCALE Conference
Portland, OR, USA
September 7-8, 2012

Cochlear Implant-like Processing of Speech Signal for Speaker Verification

Cong-Thanh Do[1] *and Claude Barras*[1,2]

[1]LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France
[2]Université Paris-Sud, 91405 Orsay Cedex, France
{cong-thanh.do, claude.barras}@limsi.fr

Abstract

In this paper, we investigate the cochlear implant-like processing of speech signal in speaker verification. This processing was applied on each speech utterance, in the temporal domain, to reduce spectral information in the original speech signal and synthesize a new one, called cochlear implant-like spectrally reduced speech (SRS), only from low-bandwidth subband temporal envelopes of the original speech. Spectral analyses, performed on voiced speech frames, showed that despite of the spectral and perceptual reduction induced by the cochlear implant-like signal processing, the global shape of the short-term spectral envelopes of the SRS signal is rather similar to that of the original speech signal.

Although the SRS is synthesized only from low-bandwidth subband temporal envelopes of original speech signal, its use in a baseline GMM-UBM speaker verification system, with cellular telephone conversational speech of the Switchboard corpus (used in NIST SRE 2002), did not alter substantially the minimal DCF (detection cost function) of the system. Furthermore, using appropriate SRS signals made it possible to reduce the minimal DCF (5.7% relative reduction) of the system. The linear combination at the score level, with equal weights, of the baseline and the SRS-based systems could also help in reducing the minimal DCF.

1. Introduction

In standard speaker verification system, the purpose of the feature extraction is to find a tradeoff between extracting relevant speaker specific information and to remove irrelevant variability from speech signal in order to achieve optimal recognition performance [1]. The feature extraction module first transforms the raw signal into feature vectors in which speaker specific properties are emphasized and reduce the number of coefficients really used by the system. Conventional speech acoustic feature, e.g. Mel frequency cepstral coefficients (MFCCs) [2] and perceptual linear predictive (PLP) coefficients [3] were introduced in early 1980s and 1990s, respectively, in speech and audio processing.

MFCCs and PLP coefficients are normally calculated from short-term windows of speech signal. These windows have typically 20-30 ms length. Indeed, the global shape of the discrete Fourier transform (DFT) magnitude spectrum, known as spectral envelope, contains information about the resonance properties of the vocal tract and has been found out to be the most informative part of the spectrum in speaker verification [4]. On the other hand, post-processing techniques of short-term feature vectors, which consider rather large time-spans of the speech signals, have been also incorporated in the feature extraction of speaker verification. In such techniques, e.g. dynamic (Δ and $\Delta\Delta$) features [5] or RASTA (relative spectra) filtering [6], larger time-spans of spectral features are processed in order to improve the noise robustness of speaker verification. These techniques operate in the spectral domains. Dynamic features have been adopted for using in conjunction with short-term speech features in speaker verification.

Cochlear implant-like spectrally reduced speech (SRS) is essentially the acoustic simulation of cochlear implant and is a spectrally reduced transform of original speech signal. In fact, cochlear implant-like SRS, henceforth abbreviated SRS, can be recognized by normal hearing listeners. The recognition scores then depend on the spectral resolution (or the number of frequency subbands of the SRS) [7]. Furthermore, human cochlear implant listeners relying on primarily temporal cues can achieve a high level of speech recognition in quiet environment [8]. The foregoing facts suggest that the SRS could contain sufficient information for human speech recognition, even though such a SRS is synthesized only from subband temporal envelopes of original speech signal [7]. On the other hand, certain speech analyses in conventional short-term acoustic features extraction, such as the Bark or Mel scale warping of the frequency axis or the spectral amplitude compression, derive from the model of human auditory system. Such auditory-like analyses, which mimic the speech processing performed by the human auditory system, are basically aimed at reducing speech signal variability and emphasizing the most relevant spectral information for recognition. As a result, the SRS should contain sufficient spectral information for speech processing systems using short-term acoustic features, e.g. MFCCs or PLP coefficients.

Indeed, it has been shown that high spectral resolution SRS contains sufficiently spectral information for hidden-Markov-model-(HMM)-based automatic speech recognition (ASR) based on MFCCs and PLP coefficients, [9, 10]. More specifically, when the ASR system was trained on original clean speech, testing speech consisting of 16-, 24-, or 32-subband SRS provided ASR performance which is comparable with that achieved with original clean speech.

Given that ASR and speaker verification, using conventional short-term acoustic features, exploit primarily information from the short-term spectral envelopes, the cochlear implant-like processing of speech signal should be also relevant for speaker verification. In this paper, we investigate the cochlear implant-like processing of speech signal for speaker verification using short-term PLP-like acoustic feature. This processing is applied on the speech signals (temporal domain). We use a standard GMM (Gaussian mixture model)-UBM (universal background model) speaker verification system for assessing the effect of the cochlear implant-like processing. These experiments are conducted on cellular telephone conversational speech from the Switchboard corpus which was used by NIST for the 2002 one-speaker detection task [11].

This paper is organized as follows. Section 2 introduces the

This work has been partially financed by OSEO, the French State Agency for Innovation, under the Quaero program and by the ANR project QCOMPERE.

cochlear implant-like processing algorithm as well as spectral analyses of cochlear implant-like SRS. Afterward, the application of cochlear implant-like processing of speech signal in speaker verification is presented in section 3. Finally, section 4 concludes the paper.

2. Cochlear implant-like processing of speech signal

In [7], Shannon et al. used the cochlear implant-like processing to synthesize acoustic simulation of cochlear implant, or cochlear implant-like spectrally reduced speech (SRS) [10], only from subband temporal envelopes of original speech signal. These signals, henceforth abbreviated SRS, are perceptually different compared to the original speech. However, it has been shown that normal hearing listeners could achieve a nearly perfect recognition score when listening to these signals [7]. In [9, 10], it has been shown that the SRS is also relevant for HMM-based ASR using MFCCs or PLP coefficients as acoustic features. The major difference between the SRS used in [7] and in [9, 10] lies in the type of carrier signals; subband temporal envelopes in [7] were used to modulate white noise whereas in [9, 10], they were used to modulated sinusoids. In the following section, we describe briefly the cochlear implant-like processing algorithm that we use to process speech signal. This algorithm is similar with that in [9, 10] and is inspired from the algorithm introduced in [7].

2.1. Cochlear implant-like processing algorithm

A speech signal $s(t)$ is first decomposed into N subband signals $s_i(t), i = 1, \ldots, N$ by using a perceptually-motivated analysis filterbank consisting of N bandpass filters. The filterbank consists of nonuniform bandwidth bandpass filters which are linearly spaced on the Bark scale in order to simulate the motion of the basilar membrane [12]. In this paper, each bandpass filter in the filterbank is a second-order elliptic bandpass filter having a minimum stopband attenuation of 50dB and a 2-dB peak-to-peak ripple in the passband. The lower, upper, and central frequencies of the bandpass filters are calculated as in [13]. Fig. 1 shows example of an analysis filterbank consisting of 16 bandpass filters.

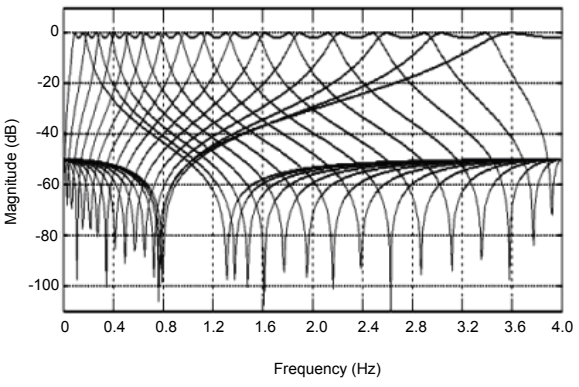


Figure 1: Frequency response of an analysis filterbank consisting of 16 second-order elliptic bandpass filters. The bandpass filters are linearly spaced on the Bark scale.

The subband temporal envelopes $m_i(t)$ of the subband signals $s_i(t), i = 1, \ldots, N$ are then extracted by, first, full-wave rectification of the outputs of the bandpass filters and, subsequently, lowpass filtering of the resulting signals. These envelopes have the same sampling rate (8 kHz) as that of the subband signal. In this work, the filter which was used to limit the bandwidth of the subband temporal envelopes is a fourth-order elliptic lowpass filter with 2-dB of peak-to-peak ripple and a minimum stop-band attenuation of 50-dB. The subband temporal envelope $m_i(t)$ is then used to modulate a sinusoid whose frequency f_{ci} equals the central frequency of the corresponding analysis bandpass filter of that subband. The subband modulated signal is then filtered again by the same bandpass filter used for the original analysis subband [7]. Finally, all the processed subband signals are summed to synthesize the SRS. The mathematical formula of the SRS $\widehat{s}(t)$ can be expressed as follows.

$$\widehat{s}(t) = \sum_{i=1}^{N} m_i(t) \cos(2\pi f_{ci} t) \quad (1)$$

As reported in [9, 10], the 16-, 24-, and 32-subband SRS provide comparable ASR performance compared to that obtained with original clean speech signals. Indeed, 16, 24 and 32 are the spectral resolution from which SRS signals contain sufficiently spectral information compared to the original speech signal. To this end, in this paper, we synthesize SRS signal from original speech signal with $N = \{16, 24, 32\}$ subbands. In addition, for the subband temporal envelopes extraction filter, we use two cut-off frequencies, 50 and 500 Hz, since these cut-off frequencies ensure a reasonable subband temporal envelope bandwidths, henceforth denoted as W, for human and machine speech recognition with short-term speech features [7, 9, 10].

2.2. Spectral analyses

In this section, we perform some spectral analyses in order to clarify the differences, in the spectral domain, between original and SRS speech signals. Fig. 2 shows examples of the spectrogram of an original speech signal (cellular telephone conversational speech) and those of the corresponding SRS signals which are synthesized from the original one. We can remark, from the spectrograms of SRS signals, that there are separable spectral subbands appearing horizontally at the frequency locations of the bandpass filters. This fact shows that the energy in the spectral domain of the SRS is concentrated around the central frequencies of the analysis filterbank.

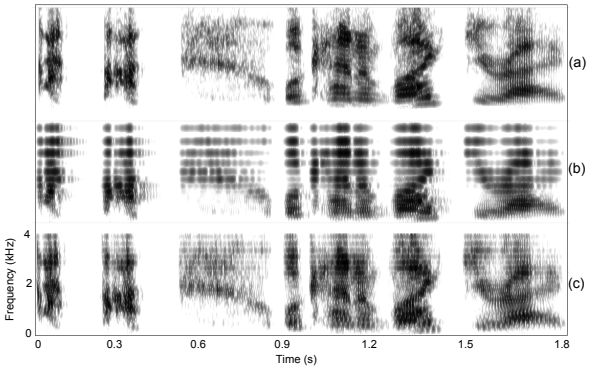


Figure 2: Spectrograms of an original speech signal (cellular telephone conversational speech) and those of the corresponding SRS signals which are synthesized from the original one. (a) Spectrogram of original speech. (b) Spectrogram of the corresponding SRS ($N = 16$, $W = 50$ Hz). (c) Spectrogram of the corresponding SRS ($N = 32$, $W = 500$ Hz).

Fig. 3 shows examples of short-term spectra and spectral envelopes analyses of original speech and corresponding SRS signals. The spectra and spectral envelopes were calculated from voiced speech frames of original speech and corresponding SRS signals. The voiced speech frames in Fig. 3 were extracted at the same instants in the original speech and the corresponding SRS signals which have the same overall lengths as the original one. The voiced speech frames were multiplied with Hamming windows and have 30ms lengths. Linear prediction (LP) [14] was used to estimate the short-term spectral envelopes.

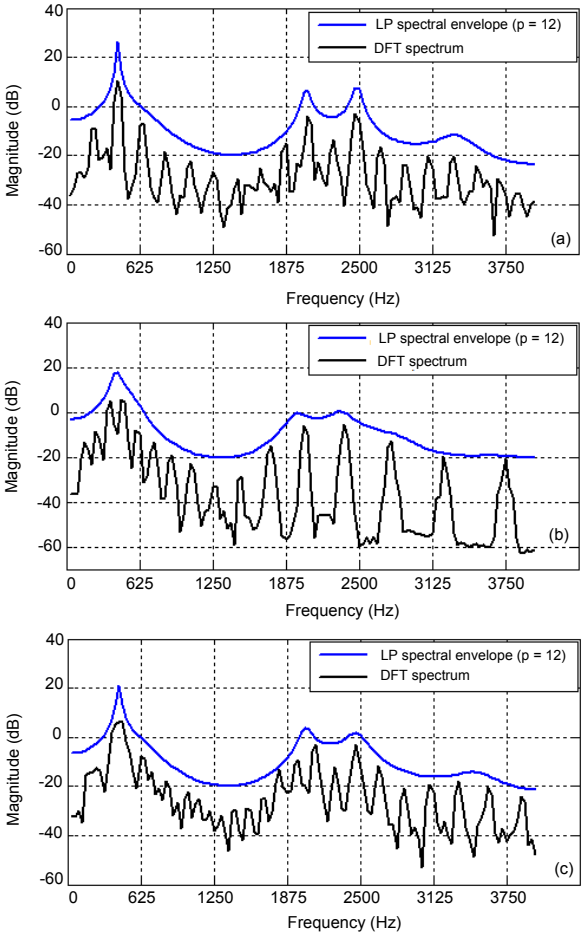


Figure 3: Discrete Fourier transform (DFT) spectra and linear prediction (LP) spectral envelopes of voiced speech frames of original speech and corresponding SRS signals. The voiced speech frames are extracted at the same instants in the original speech and the corresponding SRS signals which have the same overall lengths as the original one. (a) DFT spectrum and LP spectral envelope of original speech. (b) DFT spectrum and LP spectral envelope of SRS ($N = 16, W = 50$ Hz). (c) DFT spectrum and LP spectral envelope of SRS ($N = 32, W = 500$ Hz).

The basic idea of linear prediction is to model the short-term spectrum by an all-pole model. The order p of the all-pole model needs to be defined in advance. In this analysis, p was chosen to be equal 12 and the LP spectral envelopes were estimated by using the Levinson-Durbin recursion to solve the normal equations that arise from the least-squares formulation [14]. It can be observed from Fig. 3 that the global shapes of the LP spectral envelopes of the SRS voiced speech frames are rather similar with that of the LP spectral envelope of the voiced speech frame of original speech. In this example, the LP spectral envelope of the SRS ($N = 32, W = 500$ Hz) seems more similar to that of the original one compared to the LP spectral envelope of the SRS ($N = 16, W = 50$ Hz).

Table 1: Averages of the Pearson product-moment correlation coefficients (PPMCCs) calculated between the LP spectral envelopes of voices speech frames of original speech and cochlear implant-like spectrally reduced speech (SRS) signals, with different SRS synthesis parameters.

SRS synthesis parameters	Average PPMCC
SRS ($N = 16, W = 50$ Hz)	0.91
SRS ($N = 16, W = 500$ Hz)	0.94
SRS ($N = 24, W = 50$ Hz)	0.92
SRS ($N = 24, W = 500$ Hz)	0.96
SRS ($N = 32, W = 50$ Hz)	0.90
SRS ($N = 32, W = 500$ Hz)	0.95

We perform quantitative measure on the similarity between the global shapes of the LP spectral envelopes of SRS and that of the original speech signals. Assume that $\mathbf{x} = [x_1, x_2, \ldots, x_L]^T$ and $\mathbf{y} = [y_1, y_2, \ldots, y_L]^T$ are two LP spectral envelope vectors of two voiced speech frames, one of the original speech and another of the corresponding SRS signal, respectively. The similarity between the global shapes of $\mathbf{x}$ and $\mathbf{y}$ is measured by calculating the Pearson product-moment correlation coefficient (PPMCC) r between $\mathbf{x}$ and $\mathbf{y}$ as follows [15]

$$r = \frac{\sum_{i=1}^{L}(x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}})}{\sqrt{\sum_{i=1}^{L}(x_i - \bar{\mathbf{x}})^2}\sqrt{\sum_{i=1}^{L}(y_i - \bar{\mathbf{y}})^2}} \qquad (2)$$

where $\bar{\mathbf{x}} = \frac{1}{L}\sum_{i=1}^{L} x_i$ and $\bar{\mathbf{y}} = \frac{1}{L}\sum_{i=1}^{L} y_i$. In this analysis, the PPMCC r is calculated for all voiced speech frames of an original cellular telephone conversational speech utterance (and the corresponding SRS signals) which lasts 37 seconds. There are 1935 voiced speech frames in the utterance and the averages $\bar{r}$ of 1935 PPMCCs are shown in Table 1. It can be observed that the values of $\bar{r}$ are very close to 1. That is, the global shape of the LP spectral envelopes of voiced speech frames of SRS signals is very similar to that of the LP spectral envelopes of voiced speech frames of original speech. The values of $\bar{r}$ confirm also that the LP spectral envelope of the SRS ($N = 32, W = 500$ Hz) is more similar to that of the original one compared to the LP spectral envelope of the SRS ($N = 16, W = 50$ Hz), as shown in Fig. 3.

3. Speaker verification experiments

We made use of basic GMM-UBM speaker verification systems for comfortably evaluating the effect of the cochlear implant-like processing of speech signal on speaker verification. In this respect, we implemented two speaker verification systems, a baseline system, denoted as BL-SPKVR, using original speech signals and another system, denoted as SRS-SPKVR, using speech signals which were processed with the cochlear implant-like algorithm (see Fig. 4). The structures of the two systems were the same except the speech signals which were used in the two systems. In the SRS-SPKVR system, the cochlear implant-like processing algorithm was applied on both training and testing speech signals in order to reduce the mismatch between training and testing conditions [16]. In the following sections,

we describe the NIST one-speaker detection task, the data used to carry out the experiments and the specification of the speaker verification systems.

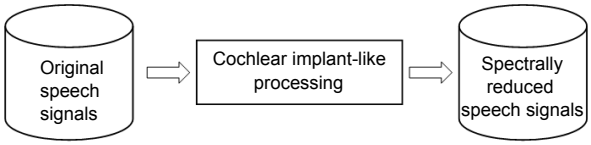


Figure 4: Original speech signals are processed with the cochlear implant-like processing algorithm (see section 2.1) to synthesize the spectrally reduced speech (SRS) signals. The SRS synthesis algorithm is applied on each original speech utterance to produce a corresponding SRS utterance.

3.1. One-speaker detection task on cellular data

The one-speaker detection task consists of determining whether a specified speaker is speaking during a given speech segment. In this work, we conduct experiments on cellular telephone conversational speech from the Switchboard corpus, following the framework defined for NIST SRE 2002 one-speaker detection task on cellular data [11]. For each of the 330 target speakers (139 males and 191 females), two minutes of concatenated segments of speech are provided for training the speaker model. The evaluation is performed on 3570 test segments (including 1442 males and 2128 females) with a mean duration of 30 seconds. Each test segment is scored against roughly 10 gender-matching imposters and against the true speaker. The gender of the target speaker is known. System performances are reported in terms of the minimal a posteriori detection cost function (DCF) defined by NIST[1] and in terms of equal error-rate (EER).

3.2. Speaker verification systems

The baseline GMM-UBM speaker verification system (BL-SPKVR) [17, 18] was implemented for the experiments using original speech signals. The front-end consists of 15 MEL-PLP coefficients along with their Δ, $\Delta\Delta$ coefficients, and the Δ and $\Delta\Delta$ energies for a total of 47 features. These features are extracted every 10ms using a 30ms window on the 0-3.8kHz bandwidth and are Gaussianized using feature warping [1] on a 3-second sliding window. Two gender-specific GMMs with 512 Gaussians and diagonal covariance matrices serve as the UBMs; they are trained on 2 hours of speech from 60 different speakers extracted from NIST SRE 2001 development data. For each target speaker, the Gaussian means of the gender-matching UBM are MAP-adapted to the speaker data. Then, for each trial, the log-likelihood of the test segment given the target model is scaled through T-norm [19] against a set of 174 impostor models from on NIST SRE 2001 training data.

The cochlear implant-like SRS signals were used in the implementation of another speaker verification system (SRS-SPKVR). As mentioned in section 2.1, we used three values of number of frequency subbands, $N = \{16, 24, 32\}$, and two values of subband temporal envelopes bandwidth, $W = \{50, 500\}$ Hz, in the cochlear implant-like processing of speech signal. Thus, we have in total six types of SRS signals. Correspondingly, six SRS-SPKVR systems were implemented. The technical specification (feature extraction and normalization, speaker modeling, scoring, etc.) of the SRS-SPKVR systems were the same as the baseline speaker verification system (BL-SPKVR).

[1] $C_{Norm} = P_{Miss} + 9.9 \times P_{FalseAlarm}$

Table 2: Speaker detection performance, in terms of min. DCF and EER, of the baseline (BL-SPKVR) and SRS-based (SRS-SPKVR) speaker verification systems. The last line contain min. DCF and EER of the linear combination at the score level, with equal weights, of the BL-SPKVR and the SRS-SPKVR ($N = 32, W = 500$ Hz) systems.

Speaker verification system	*min. DCF*	*EER (%)*
BL-SPKVR[1] (baseline)	0.3300	**8.34**
SRS-SPKVR ($N = 16, W = 50$)	0.3481	9.19
SRS-SPKVR ($N = 16, W = 500$)	0.3411	9.16
SRS-SPKVR ($N = 24, W = 50$)	0.3383	8.78
SRS-SPKVR ($N = 24, W = 500$)	0.3298	8.81
SRS-SPKVR ($N = 32, W = 50$)	0.3323	8.93
SRS-SPKVR[2] ($N = 32, W = 500$)	**0.3143**	8.69
BL-SPKVR[1] + SRS-SPKVR[2]	**0.3112**	**8.44**

3.3. Experimental results

Speaker detection results, performed by the baseline system and the SRS-based systems, are presented in Table 2. It can be observed that the minimal DCF of most SRS-SPKVR systems have not been substantially altered compared to the minimal DCF of the baseline system BL-SPKVR even though the cochlear implant-like processing has significantly modified the original speech signals. Especially, the SRS-SPKVR ($N = 32, W = 500$ Hz) system has substantial lower minimal DCF compared to that of the baseline system (5.7% relative reduction). The DET (detection error tradeoff) curves in Fig. 5 show that the performance of the SRS-SPKVR ($N = 32, W = 500$ Hz) system outperforms that of the baseline system for the low false alarm rate. The minimal DCFs in Table 2 are rather consistent with the PPMCCs which measure the similarity between the global shapes of LP spectral envelopes of voiced speech frames of original speech and SRS signals (see Table 1, section 2.2).

A linear combination (with equal weights) at the score level on the outputs of the baseline and the SRS-SPKVR ($N = 32, W = 500$ Hz) systems made it possible to achieve a lower minimal DCF compared to individual systems (see the last line of Table 2). In addition, the EER obtained with this combination approaches that of the baseline system. The EER of the baseline system is lower than that of the SRS-SPKVR systems.

4. Conclusion

In this paper, we have investigated the cochlear implant-like processing of speech signal in speaker verification. This processing has been previously investigated in speech recognition by humans [7] and machines [9, 10, 16]. The cochlear implant-like processing was inspired from speech signal processing algorithm in standard cochlear implant. This algorithm reduces the spectral information in the original speech signals and synthesizes new speech signals, called spectrally reduced speech (SRS), from low-bandwidth subband temporal envelopes of the original ones. This reduction made the SRS signals spectrally and perceptually different compared to the original speech signals. It has been shown, through spectral analyses of voiced speech frames, that even though the spectral information has been substantially reduced, the global shapes of the short-term linear prediction spectral envelopes have almost been retained in the SRS signals. This fact might be the critical factor which makes the SRS relevant for speaker verification using short-term speech features.

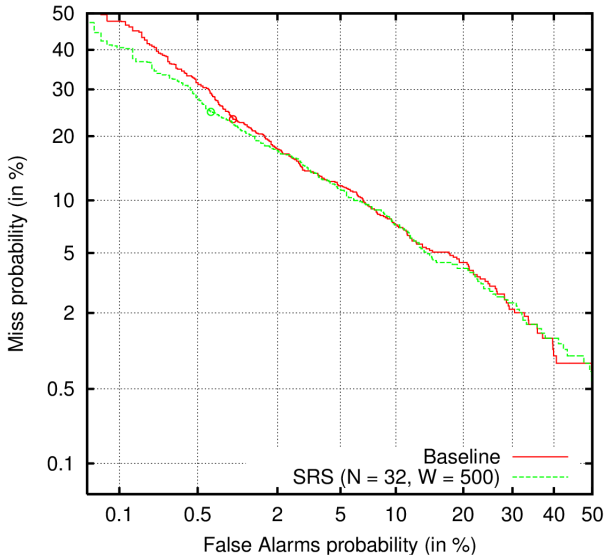


Figure 5: DET curves of the baseline (BL-SPKVR) and the SRS-SPKVR ($N = 32, W = 500$ Hz) systems. Circles are drawn at minimal DCF operating points.

Although the SRS is synthesized only from low-bandwidth subband temporal envelopes of original speech signal, its use in a standard GMM-UBM speaker verification system, with cellular telephone conversational speech of the Switchboard corpus, has not altered substantially the minimal DCF of the system. Furthermore, using appropriate SRS signals (SRS ($N = 32, W = 500$ Hz) in this case) has made it possible to reduce the minimal DCF (5.7% relative reduction) of the system. The linear combination at the score level, with equal weights, of the baseline and the SRS-based systems could also help in reducing the minimal DCF. The SRS is therefore relevant not only for ASR but also for speaker verification. This study might open potential research directions, e.g. speaker verification through low-bandwidth (or low bit-rate) telecommunication networks, since the SRS signal can be synthesized from low-bandwidth subband temporal envelopes of original speech signal.

5. References

[1] Pelecanos, J., and Sridharan, S., "Feature warping for robust speaker verification", in Proc. *Odyssey 2001: The Speaker Recognition Workshop*, pp. 213-218, Crete, Greece, Jun. 18-22. 2001.

[2] Davis, S.B., and Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuous spoken sentences", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357-366, 1980.

[3] Hermansky, H., "Perceptual linear predictive (PLP) analysis of speech", *J. Acoust. Soc. Am.*, vol. 87, no. 4, pp. 1738-1752, Apr. 1990.

[4] Kinnunen, T., and Li, H., "An overview of text-independent speaker recognition: from features to supervectors", *Speech Communication*, vol. 52, no. 1, pp. 12-40, Jan. 2010.

[5] Furui, S., "Speaker-independent isolated word recognition using dynamic features of speech spectrum", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 52-59, Feb. 1986.

[6] Hermansky, H., and Morgan, N., "RASTA processing of speech", *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 578-589, Oct. 1994.

[7] Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M., "Speech recognition with primarily temporal cues", *Science*, vol. 270, no. 5234, pp. 303-304, Oct. 1995.

[8] Zeng, F.-G., Nie, K., Stickney, G., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K., "Speech recognition with amplitude and frequency modulations", *Proceedings of National Academy of Sciences*, vol. 102, no. 7, pp. 2293-2298, Feb. 2005.

[9] Do, C.-T., Pastor, D. and Goalic, A., "On the recognition of cochlear implant-like spectrally reduced speech with MFCC and HMM-based ASR", *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 5, pp. 1065-1068, Jul. 2010.

[10] Do, C.-T., Pastor, D., Le Lan, G., and Goalic, A., "Recognizing cochlear implant-like spectrally reduced speech with HMM-based ASR: experiments with MFCCs and PLP coefficients", *INTERSPEECH*, pp. 2634-2637, Makuhari, Japan, September 26-30 2010.

[11] "The NIST year 2002 speaker recognition evaluation plan", http://www.itl.nist.gov/iad/mig/tests/spk/2002/index.html, 2002.

[12] Kubin, G., and Kleijn, B.W., "On speech coding in a perceptual domain", in Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 205-208, Arizona, USA, Mar. 15-19, 1999.

[13] Gunawan, T.S., and Ambikairajah, E., "Speech enhancement using temporal masking and fractional Bark gammatone filters", in Proc. *10th Australian Int. Conf. Speech Sci. Technol.*, pp. 420-425, Sydney, Australia, Dec. 08-10, 2004.

[14] Makhoul, J., "Linear prediction: a tutorial review", *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561-580, Apr. 1975.

[15] Rodgers, J.L., and Nicewander, W.A., "Thirteen ways to look at the correlation coefficient", *The American Statistician*, vol. 42, no. 1, pp. 59-66, Feb. 1988.

[16] Do, C.-T., Pastor, D. and Goalic, A., "A novel framework for noise robust ASR using cochlear implant-like spectrally reduced speech", *Speech Communication*, vol. 54, no. 1, pp. 119-133, Jan. 2012.

[17] Reynolds, D., Quatieri, T., and Dunn, R., "Speaker verification using adapted gaussian mixture models", *Digital Signal Processing*, vol. 10, pp. 19–41, 2000.

[18] Barras, C., and Gauvain, J.-L.,"Feature and score normalization for speaker verification of cellular data",*Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, pp. II-49–52, Hong Kong, avril 2003.

[19] Auckenthaler, R., Carey, M., and Lloyd-Thomas, H., "Score normalization for text-independent speaker verification systems", *Digital Signal Processing*, vol. 10, pp. 42–54, 2000.