Speaker verification with multiple classifier fusion using Bayes based confidence measure

Fernando Huenupán, Nestor Becerra Yoma, Carlos Molina and Claudio Garretón

Speech Processing and Transmission Laboratory, Department of Electrical Engineering
Universidad de Chile, Santiago, Chile
nbecerra@ing.uchile.cl

10.21437/Interspeech.2007-170

Abstract

A novel framework based on Bayes-based confidence measure (BBCM) for Multiple Classifier System (MCS) fusion is proposed. As shown here, BBCM based MCS combination scheme corresponds to the ordinary Bayes fusion weighted by the reliability of each individual classifier. BBCM provides a formal model for heuristic weighting functions employed elsewhere. When compared with the ordinary Bayesian fusion, the proposed method leads to reductions as high as 20% and 50% in EER and the area below the ROC curve, respectively, in speaker verification.
Index Terms: speaker verification, multi-classifier system, confidence measure.

1. Introduction

In pattern recognition, the use of multiple classifier systems (MCS) has been addressed in several fields [8-10]. In speaker verification (SV), neural networks [4][13] and linear combination [11] are among the most popular approaches for MCS. Nevertheless, most of these techniques do not provide the required mathematical analysis to achieve the relevant performance evaluations [7]. In MCS, classifiers are usually combined in parallel into two levels [11]: abstract level (Fig.1) and score level. In Fig.1: O is the observed input signal; CL_j is the classifier j, where $1 \le j \le J$ and J is the number of classifiers; $S_{CL_j}(O)$ is the score at the output of classifier j; d_{CL_j} is the local decision or classification due to classifier j; and, $D(O)$ is the final decision or classification that corresponds to O. $D(O)$ and d_{CL_j} indicate one of the M classes denoted by C_m, where $1 \le m \le M$ and M is the total number of classes .

In pattern recognition theory, the most straightforward formal strategy to fusion classifiers is the Bayes classification theory [9]

$$D(O) = \arg\max_m \left[Pr(C_m \mid S(O)) \right] = \arg\max_m \left\{ \frac{Pr\left[S(O) \mid C_m\right] Pr(C_m)}{\sum_{\hat{m}=1}^{M} Pr\left[S(O) \mid C_{\hat{m}}\right] Pr(C_{\hat{m}})} \right\} \quad (1)$$

where $S(O) = \left[S_{CL_1}(O), ..., S_{CL_j}(O), ..., S_{CL_J}(O) \right]$.

The a priori multivariable p.d.f.'s in (1), $Pr[S(O)|C_m]$, may require an unmanageable amount of training data to be reliable estimated [9]. The maximization of (1) can be simplified by a priori p.d.f.'s of individual classifiers. The classical techniques of simplification are [9-10]: Product Rule; Max Rule; Min rule; Mean Rule and, Majority Vote Rule (MVR). One of the most widely employed approximations is MVR [8], which combines local decision of individual classifiers as in Fig. 1. For this reason MVR is adopted in this paper as evaluation framework.

In [14], Bayes-based confidence measure (BBCM) was proposed to address the problem of assessing the accuracy in automatic speech recognition (ASR). BBCM is a probability itself and incorporates a priori information about the recognizer performance. Compared with standard confidence measures, BBCM improves the discrimination of misrecognized words. Surprisingly, the problem of assessing the accuracy of SV systems has not been exhaustively addressed in the specialized literature.

This paper proposes: a) a MCS strategy applied to SV by using simplifications of the Bayesian Fusion; b) a new classification criterion based on BBCM applicable to any pattern recognition problem; c) a new strategy for classifier fusion in MCS based on BBCM also applicable to any pattern recognition problem, and d) modeling the problem of confidence measure in SV by applying BBCM. As shown here, BBCM based MCS fusion scheme corresponds to the ordinary Bayes fusion weighted by the reliability of each individual classifier. Moreover, BBCM provides a formal model for heuristic weighting functions employed elsewhere. Finally, the approach and analysis presented in this paper has not been found in the specialized literature.

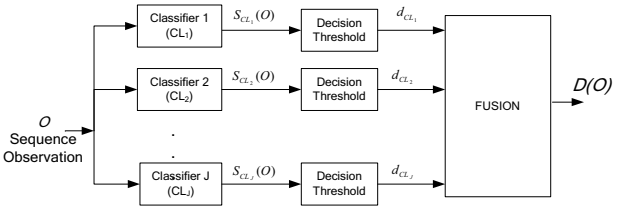

Figure 1: MCS fusion at abstract level

2. Bayes classifier fusion applied to SV

In SV, the task is to decide about the user identity that claims a given identity, and two classes are possible: client, C_1; and, impostor, C_2. From the comparison of the input speech signal with each classifier CL_j results a score $S_{CL_j}(O)$, local decision d_{CL_j}, $Pr\left[S_{CL_j}(O) \mid C_1\right]$ and $Pr\left[S_{CL_j}(O) \mid C_2\right]$. Consequently, the classifier array provides a set of scores $S(O)$, local decisions $d(O)$, $Pr[S(O)|C_1]$ and $Pr[S(O)|C_2]$. If the a priori probabilities of C_1 and C_2 are assumed uniformly distributed, then (1) can be written as:

$$D(O) = \arg\max_m \left\{ Pr\left[C_m \mid S(O)\right] \right\} = \arg\max_m \left\{ \frac{Pr\left[S(O) \mid C_m\right]}{\sum_{\hat{m}=1}^{M} Pr\left[S(O) \mid C_{\hat{m}}\right]} \right\} \quad (2)$$

As mentioned above, (2) can be approximated by MVR. In MCS, MVR corresponds to a straightforward scheme that

combines the outputs of classifiers. In this paper, a weighted version of MVR, WMVR-MCS, is employed. WMVR-MCS is defined as [2][9]:

$$D(O)=\sum_{j=1}^{J}\Delta_{CL_j}Pr\left[d_{CL_j}\mid S_{CL_j}(O)\right] \tag{3}$$

where $\Delta_{CL_j}=\begin{cases}1 & \text{if } \mathrm{d}_{\mathrm{CL_j}}(O)=C_1\\ -1 & \text{if } \mathrm{d}_{\mathrm{CL_j}}(O)=C_2\end{cases}$

3. BBCM in SV

BBCM was firstly proposed in the field of speech recognition. An ASR system receives a speech signal as input and delivers a set of recognized words (i.e. $w_1,w_2,...,w_i..,w_I$), where w_i denotes the i^{th} word in the string. If WF denotes a word feature, BBCM is defined in ASR as [14]:

$$\begin{aligned}&BBCM(WF_i)\\&=Pr(w_i \text{ is ok}|WF_i)=\frac{Pr(WF_i\mid w_i \text{ is OK})Pr(w_i \text{ is OK})}{Pr(WF_i)}\end{aligned} \tag{4}$$

where event "OK" corresponds to the fact that word w_i, which is contained at least in one of the N-best hypotheses, was properly recognized. Notice that $BBCM(WF_i)$ is a probability itself. Moreover, the distribution $Pr(WF_i \mid w_i \text{ is OK})$ and the probability $Pr(w_i \text{ is OK})$ provide information about the recognition engine performance.

In SV, the word feature employed in ASR can certainly be replaced with the score of a given classifier. As a consequence, applying the definition of BBCM to the SV problem leads to [6]:

$$\begin{aligned}&BBCM\left[S_{CL_j}(O)\right]=Pr\left[d_{CL_j}\text{ is OK}|\mathrm{S_{CL_j}}(O)\right]\\&=\frac{Pr\left[\mathrm{S_{CL_j}}(O)\mid d_{CL_j}\text{ is OK}\right]Pr(d_{CL_j}\text{ is OK})}{Pr\left[S_{CL_j}(O)\right]}\end{aligned} \tag{5}$$

where "d_{CL_j} is OK" corresponds to the decision of classifier CL_j is correct. By considering two classes as mentioned above, (5) can be expressed as [6]:

$$\begin{aligned}&BBCM\left[S_{CL_j}(O)\right]=Pr\left[d_{CL_j}\text{ is OK}|\mathrm{S_{CL_j}}(O)\right]\\&=Pr\left\{\left[O\in C_1\wedge d_{CL_j}(O)=C_1\right]\vee\left[O\in C_2\wedge d_{CL_j}(O)=C_2\right]|\mathrm{S_{CL_j}}(O)\right\}\end{aligned} \tag{6}$$

where $O\in C_m$ denotes that the input signal O belongs to class C_m. Notice that $Pr\left[d_{CL_j}\text{ is OK}|\mathrm{S_{CL_j}}(O)\right]$ in (6) can also be written as:

$$\begin{aligned}Pr\left[d_{CL_j}\text{ is OK}|\mathrm{S_{CL_j}}(O)\right]=&Pr\left[O\in C_1\wedge d_{CL_j}(O)=C_1|\mathrm{S_{CL_j}}(O)\right]+\\&Pr\left[O\in C_2\wedge d_{CL_j}(O)=C_2|\mathrm{S_{CL_j}}(O)\right]\end{aligned} \tag{7}$$

The generalization of this result is direct and the application of (7) to the M-class classification problem leads to:

$$\begin{aligned}&BBCM\left[\mathrm{S_{CL_j}}(O)\right]=Pr\left[d_{CL_j}\text{ is OK}|\mathrm{S_{CL_j}}(O)\right]\\&=\sum_{m=1}^{M}\Pr\left[O\in C_m\wedge d_{CL_j}(O)=C_m|\mathrm{S_{CL_j}}(O)\right]\end{aligned} \tag{8}$$

By using the Bayes theorem, (8) can be expressed as:

$$BBCM\left[\mathrm{S_{CL_j}}(O)\right]=\sum_{m=1}^{M}\left\{\begin{aligned}&\frac{Pr\left[S_{CL_j}(O)\mid O\in C_m\wedge d_{CL_j}(O)=C_m\right]}{\sum_{\tilde{m}=1}^{M}Pr\left[S_{CL_j}(O)\mid C_{\tilde{m}}\right]Pr(C_{\tilde{m}})}\times\\&Pr\left[O\in C_m\wedge d_{CL_j}(O)=C_m\right]\end{aligned}\right\} \tag{9}$$

In (8) the following definition is suggested:

$$BBCM\left[\mathrm{S_{CL_j}}(O)\ \wedge C_m\right]=Pr\left[O\in C_m\wedge d_{CL_j}(O)=C_m|\mathrm{S_{CL_j}}(O)\right] \tag{10}$$

where $BBCM[S_{CL}(O)\wedge C_m]$ denotes the probability of "$d_{CL_j}(O)$ is OK and equal to $\mathrm{C_m}$" given score $\mathrm{S_{CL_j}}(O)$. Finally, BBCM for classifier j is re-written as:

$$BBCM\left[S_{CL_j}(O)\right]=\sum_{m=1}^{M}BBCM\left[S_{CL_j}(O)\wedge C_m\right] \tag{11}$$

Figure 2 shows the $BBCM\left[S_{CL_j}(O)\right]$ curve estimated with a text-dependent SV system based on forced Viterbi alignment [5]. As can be seen, the minimum value of BBCM takes place when the classifier score is equal to TEER (Threshold of Equal Error Rate). This result is intuitive and means that the closer the score from the decision threshold, the lower the classification confidence. It is worth highlighting that BBCM provides a formal framework to model confidence as a probability, which in turn is not an heuristic result.

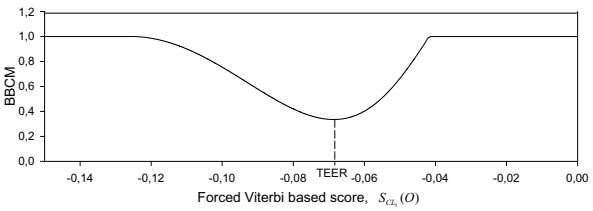

Figure 2: BBCM curve in SV with Forced Viterbi algorithm.

4. BBCM Based Classifier Fusion in SV

BBCM as defined in (10) could also be employed as the classification criterion that needs to be maximized to optimally decide about the recognized class.

4.1 BBCM as a classification criterion

According to(10), $BBCM\left[S_{CL_j}(O)\wedge C_m\right]$ can be written as:

$$\begin{aligned}&BBCM\left[S_{CL_j}(O)\wedge C_m\right]=Pr\left[O\in C_m\wedge d_{CL_j}(O)=C_m|\mathrm{S_{CL_j}}(O)\right]\\&=Pr\left[O\in C_m\mid S_{CL_j}(O)\right]\cdot Pr\left[d_{CL_j}(O)=C_m\mid S_{CL_j}(O),O\in C_m\right]\end{aligned} \tag{12}$$

where $Pr\left[O\in C_m\mid S_{CL_j}(O)\right]=Pr\left[C_m\mid S_{CL_j}(O)\right]$ is the a posteriori probability that appears in the maximization according to the Bayes classification rules in (2); and, $Pr\left[d_{CL_j}(O)=C_m\mid S_{CL_j}(O),O\in C_m\right]$ corresponds to an additional information incorporated by BBCM that is related to the reliability of an individual classifier. As a consequence, it sounds reasonable to classify by selecting the class that maximizes $BBCM\left[S_{CL_j}(O)\wedge C_m\right]$:

$$\begin{aligned}d_{CL_j}&=\arg\max_m\left\{BBCM\left[S_{CL_j}(O)\wedge C_m\right]\right\}\\&=\arg\max_m\left\{Pr\left[O\in C_m\mid S_{CL_j}(O)\right]\times\right.\\&\qquad\left.Pr\left[d_{CL_j}(O)=C_m\mid S_{CL_j}(O),O\in C_m\right]\right\}\end{aligned} \tag{13}$$

Observe that (13) maximizes a confidence measure instead of the conventional a posteriori probability $Pr\left[C_m\mid S_{CL_j}(O)\right]$. Nevertheless, BBCM is a probability itself and also incorporates $Pr\left[C_m\mid S_{CL_j}(O)\right]$, besides the information on classifier reliability. Finally, the result denoted by (13) should be applicable to any classification problem independently of the number of classes.

4.2 Fusion with BBCM

According to (13), the local decision of classifier *j* is the class that maximizes $BBCM\left[S_{CL_j}(O)\right]$. By considering the MCS environment in Fig. 1, the selected class corresponds to:

$$\begin{aligned} D(O) &= \arg\max_m \left\{ BBCM\left[S(O) \wedge C_m\right] \right\} \\ &= \arg\max_m \left\{ Pr\left[O \in C_m \mid S(O)\right] \cdot Pr\left[D(O) = C_m \mid S(O), O \in C_m\right] \right\} \end{aligned} \quad (14)$$

where $BBCM[S(O) \wedge C_m]$ is estimated as:

$$\begin{aligned} & BBCM\left[\mathrm{S}(O) \wedge C_m\right] \\ &= \frac{Pr\left[O \in C_m \wedge D(O) = C_m\right] \cdot Pr\left[S(O) \mid O \in C_m \wedge D(O) = C_m\right]}{\sum_{\tilde{m}=1}^{M} Pr\left[D(O) \mid C_{\tilde{m}}\right] Pr(C_{\tilde{m}})} \end{aligned} \quad (15)$$

In (15), $Pr[S(O)|O \in C_m \wedge D(O)=C_m]$ required to estimate $BBCM[S(O) \wedge C_m]$ also demands a high amount of training data. Consequently, the same strategy adopted for Bayesian fusion (WMVR) is also used here. So, in BBCM fusion with WMVR, BBCM-WMVR, $Pr\left[\mathrm{d}_{\mathrm{cl_j}} \mid S_{CL_j}(O)\right]$ is replaced with $BBCM\left[S_{CL_j}(O) \wedge d_{cl_j}(O)\right]$ in (3):

$$D(O) = \sum_{j=1}^{J} \Delta_{CL_j} \cdot BBCM\left[S_{CL_j}(O) \wedge d_{cl_j}(O)\right] \quad (16)$$

According to (12), $D(O)$ in (16) corresponds to the Bayes WMVR approximation weighted by the reliability of classifier *j* given class C_m.

4.3 BBCM versus classic Bayesian score

BBCM incorporates information about the classifier performance, confidence, and the a posteriori probability itself. In contrast, Bayes classification employs only the a posteriori probability. Moreover, BBCM score is between 0 and 1 (see Fig.2), but the sum over the classes is not necessarily equal to one. On the other hand, in Bayes rule a posteriori probabilities are defined in a range that is classifier dependent but provide a sum over all the classes equal to one. As a result, BBCM seems an interesting framework to address the problem of MCS fusion due to the fact that the concept of confidence is applicable to any classifier.

5. Tuned weighting function (TWF)

As can bee seen in (11), $BBCM\left[S_{CL_j}(O)\right]$ is the sum of $BBCM\left[S_{CL_j}(O) \wedge C_m\right]$ over all classes. Usually, when two classes are employed:

$$BBCM\left[S_{CL_j}(O) \wedge C_m = d_{CL}\right] >> BBCM\left[S_{CL_j}(O) \wedge C_m \neq d_{CL}\right]$$

Consequently, the following approximation is considered:

$$BBCM[S_{CL_j}(O)] \cong BBCM[S_{CL_j}(O) \wedge d_{CL_j}] \quad (17)$$

By replacing (17) in (16) BBCM-WMVR can be written as:

$$D(O) = \sum_{j=1}^{J} \Delta_{CL_j} \cdot BBCM\left[S_{CL_j}(O)\right] \quad (18)$$

$BBCM\left[S_{CL_j}(O)\right]$ requires p.d.f.'s estimated with likelihood maximization. However, these a priori p.d.f.'s do not necessarily minimize the error rate. As show in Fig.2, $BBCM\left[S_{CL_j}(O)\right]$ curve could be approximated by an inverted Gaussian curve. For this reason, an empirical tuned weighting function (TWF) is proposed and evaluated:

$$TWF\left[S_{CL_j}(O)\right] = 1 - K_{CL_j} \cdot N\left[S_{CL_j}(O), TH_{CL_j}, \sigma_{CL_j}\right] \quad (19)$$

where K_{CL_j} and σ_{CL_j} are estimated with a training database (Fig. 3), and TH_{CL_j} is the threshold of equal error rate (TEER) for classifier *j*. In (18) $BBCM\left[S_{CL_j}(O)\right]$ is replaced with $TWF\left[S_{CL_j}(O)\right]$:

$$D(O) = \sum_{j=1}^{J} \Delta_{CL_j} \cdot TWF\left[S_{CL_j}(O)\right] \quad (20)$$

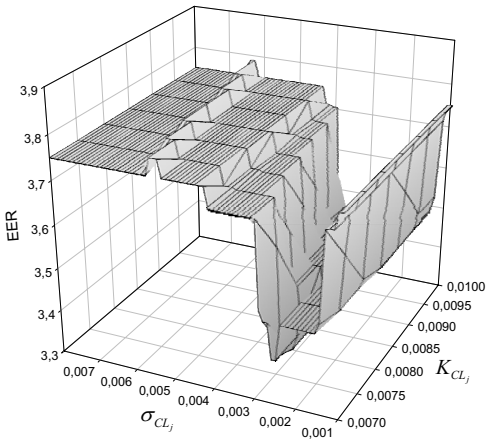

Figure 3: EER versus K_{CL_j} and σ_{CL_j} in TWF as defined in (19) with a forced Viterbi based SV system.

6. Experiments

The database is composed of 40 speakers (20 males and 20 females). The vocabulary corresponds to Spanish digits. Each speaker pronounced the 10-digit sequence "0-1-2-3-4-5-6-7-8-9" three times for enrolling. For verification, every speaker uttered the four-digit sequences "1-8-6-4", "4-5-2-0" and "9-5-7-3" three times each. The enrolling and verification speech signals were recorded on the same telephone. The database was divided in two groups, A and B. Database A, composed of 30 speakers (15 males and 15 females), is used for testing. Database B, composed of 10 speakers (5 males and 5 females), is employed to estimate the required a priori p.d.f.'s, for BBCM and Bayes curves, and the parameters K_{CL_j} and σ_{CL_j} in (19). The database employed in this paper is similar to those mentioned elsewhere [13]. Enrolling and verification utterances are decomposed as a sequence of triphones. Thirty-three cepstral coefficients are computed per frame: the frame energy plus ten static coefficients and their first and second time derivatives. The HMM's were trained with the Viterbi algorithm.

Three standard SV techniques are addressed: forced Viterbi based score (LL) [5]; Maximum Vote Rate for sequence of feature vectors, MVR-FV [12]; and, Support Vector Machines, SVM [2]. In LL system, the HMM's were trained with the Viterbi algorithm. Each triphone was modeled with a three-state left-to-right HMM topology without skip-state transition, with one multivariate Gaussian density per state in speaker dependent models, and eight multivariate Gaussian densities per state in the speaker

independent model. Both models employed diagonal covariance matrices. In MVR-FV, the size of window is equal to two frames and the threshold for local decisions is estimated with database B. In SVM, Gaussian kernel is employed [1] and K-means algorithm is used to group the training data with 256 and 512 codewords for client and impostor class, respectively. The optimal K_{CL_j} and σ_{CL_j} in (19) were estimated with the training database by minimizing EER: k= 0.08, 003, 0.002 and σ= 0.0035, 0.02 and 0.003 for LL, MVR-FV and SVM systems, respectively. FA/FR error rates are computed with database A. FR curves are estimated with 30 speakers x 9 verification signals per client = 270 signals; and, FA curves are computed by avoiding cross-gender impostor trials with (15 impostors) x 9 verification signals per impostor x 30 users = 3780 experiments. The baseline system in this paper is given by LL that is more accurate than MVR-FV and SVM. The baseline system gives an EER equal to 5.9%. The EER's provided by MVR-FV and SVM are 7.42% and 14.82%, respectively. Results are presented in Table 1 and Fig. 4.

	Individual Classifier			**MCS Fusion**			
	LL	**MVR FV**	**SVM**	**Bayes WMVR Eq.** (3)	**BBCM WMVR Eq.** (16)	**approx BBCM. WMVR Eq.** (18)	**WTF WMVR Eq.** (20)
EER	5,93	7,42	14,82	4,07	3,43	3,58	3,00
ROC	118	229	782	81	41	48	38

Table 1: EER and integral below the ROC curve with individual classifiers and MCS fusion approximation.

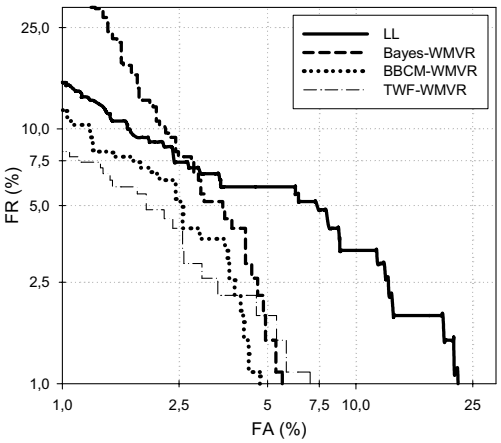

Figure 4: DET curves with LL, Bayes-WMVR, BBCM-WMVR and TWF-WMVR.

7. Discussion and conclusions

A new framework based on BBCM is presented for MCS fusion. The BBCM fusion scheme corresponds to ordinary Bayes combination weighted by the reliability of each classifier. Also, BBCM provides a formal model instead of the heuristic weighting functions employed elsewhere. As can be seen in Table 1 and Fig.4, the BBCM-WMVR fusion method defined in (16) outperformed the baseline system and the ordinary Bayes-WMVR fusion approximation. Compared with the baseline system, BBCM-WMVR leads to reductions in EER and in the integral below the ROC equal to 42% and 65%, respectively. When compared with the ordinary Bayes-WMVR fusion, BBCM-WMVR gives reductions of 16% and 41% in EER and in the integral below the ROC, respectively. The approximation of BBCM-WMVR according to (18) provide slightly lower improvements than BBCM-WMVR in (16). However, the BBCM based empirical weighting function, TWF-WMVR, leads to reductions as high as 26% and 53% in EER and the area below ROC, respectively, when compared with Bayes-WMVR.

Although tested with a SV task on the telephone line, the approaches presented here should easily be generalized to any classification problem independently on number of classes. Finally the applicability of confidence based MCS fusion to SV in mismatch conditions and to other classification problems is proposed as future work.

8. Acknowledgments

This work was funded by Conicyt-Chile under grants Fondef N° D05I-10243 and Fondecyt N° 1070382.

9. References

[1] Burges, C. "A tutorial on support vector machines for pattern recognition". Data Mining and Knowledge Discovery, 2(2), pp. 121–167, 1998.

[2] Campbell, W.M. et al. "High-level speaker verification with support vector machines". ICASSP, Montréal, Quebec, Canada, pp. 73-76, 2004.

[3] De Stefano, C. et al. "An adaptive weighted majority vote rule for combining multiple classifiers". IEEE ICPR, Vol. 2, pp. 192-195, 2002.

[4] Farrell, K.R. "Text-dependent speaker verification using data fusion". ICASSP, vol. 1, pp. 349-352. 1995.

[5] Furui, S. "Recent advances in speaker recognition". Pattern Recognition Letters. 18, pp. 859-872, 1997.

[6] Huenupán, F., Yoma N.B, Molina C. and Garreton C. "Confidence based multiple classifier fusion in speaker verification". Submitted to Pattern recognition letters, 2007.

[7] Kim, T., Ko, H. "Utterance verification under distributed detection and fusion framework". In Proc. of Eurospeech, Geneva, Switzerland, pp. 889-892. 2003.

[8] Kittler, J., Alkoot, F.M. "Sum versus vote fusion in multiple classifier systems". IEEE Trans. Pattern Anal. and Machine Intelligence. 25, pp. 110-115. 2003.

[9] Kittler, J. et al. "On combining classifiers". IEEE Trans. on Pattern Analysis and Machine Intelligence. 20, pp. 226-239. , 1998.

[10] Kuncheva, L.I. "A theoretical study on six classifier fusion strategies". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 281-286, 2002.

[11] Mak, M.W. et al. "Robust speaker verification from GSM-transcoded speech based on decision fusion and feature transformation". ICASSP, pp. 745 -748, 2003.

[12] Radova, V., Padrta, A. "Comparison of several speaker verification procedures based on GMM". In: Proc. Interspeech, pp. 1777-1780, 2004.

[13] Yegnanarayana, B. et al. "Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system". IEEE Trans. on Speech and Audio Process. 13(4), pp. 578-582, 2005.

[14] Yoma, N.B. et al. "Bayes-based confidence measure in speech recognition". IEEE Signal Processing Letters. 12, pp. 745 – 748, 2005.