

# APPLICATION OF UNCERTAINTY-BASED METHODS TO FUSE LANGUAGE IDENTIFICATION EXPERT DECISIONS

Jorge Gutiérrez

Jean-Luc Rouas

Régine André-Obrecht

IRIT – Université Paul Sabatier  
118, Route de Narbonne  
F-31062 Toulouse Cedex 4 France

[gutierre@irit.fr](mailto:gutierre@irit.fr)

[rouas@irit.fr](mailto:rouas@irit.fr)

[obrecht@irit.fr](mailto:obrecht@irit.fr)

## Abstract

Relying on uncertainty theories, a formal methodology to fuse automatic language identification expert decisions is presented. Special attention is focused on representing and making use of *a priori* knowledge about the performance of experts: the Discriminant Factor Analysis method is applied to compute performance confidence indices at the class level. Experimentation results support the hypothesis that implementing some uncertainty-based inference techniques issued from recent research advances in Evidential or Possibility theories appears not only as a feasible fusion alternative to empirical weighted techniques but also as the one which best exploits the knowledge provided by such indices while delivering better language identification rates.

**Keywords:** Multicriteria Decision Making, Information Fusion, Performance Confidence Indices, Discriminant Factor Analysis, Automatic Language Identification.

## 1 Introduction

In the field of Automatic Language Identification (ALI), experts are primary systems, also known as sources of decision information, whose aim is to identify as soon as

possible the language in which an utterance has been pronounced. An ALI system can be composed of several experts whose architecture allows them to take advantage of language-discriminant specific features and characterises them as:

- Acoustic Expert: vocalic and consonant phones and their frequency of occurrence differ from language to language [12]; the acoustic information of each language is modelled by Gaussian Mixture Models (GMM) or Hidden Markov Models (HMM) [16];
- Phonotactic Expert: specific sequences of phonetic units that appear at different occurrence rate in each language [16]; bi-gram or tri-gram models translates the language phonotactic rules;
- Prosodic Expert: sound duration, fundamental frequency, intensity variation and rhythm are language discriminant lineaments [14]; this expert is mostly based on statistical moments computed on the rhythm and the fundamental frequency.

In taking into account the identification decisions issued from experts, an ALI system faces the problem of merging (fusing) them in a suitable way. Till now several merging techniques have been implemented and have evolved from the application of empirical operators (average, addition, multiplication, consensus, and so forth) still used not long ago,

to, nowadays estimations of confidence indicators [13] regarding the performance of experts; this is applied as heuristics-like *a priori* knowledge by weighting the expert decisions.

Both generation and application of confidence indices are carried out in an empirical iterative way by testing and adjusting values with no clear formal background: good performance is often obtained though [9]. So, great efforts have started to be deployed to try to formally justify such techniques [4] [10].

We propose an original method to fuse language identification expert decisions. It consists of developing a formal methodology to:

- represent and compute confidence indices by extracting language-discriminant information while processing a development corpus and using the Discriminant Factor Analysis (DFA) method in the decision score field. The DFA projection is used to obtain the confusion matrix and to provide expert and class performance confidence indices.

- model the language identification process by means of the concept of a Linguistic Variable, so that we can work on the scores in the domains of Possibility and Evidence Theories, where respectively:

- we implement a hierarchical searching inference mechanism based on the class confidence indices and apply an Adaptive Fusion [8] technique to compute the possibility degree of each language;
- we assign basic belief mass values to the language occurrence events [5], weight such event mass values [1] with the class confidence indices, apply the Dempster's orthogonal rule to fuse them and derive the pignistic probability [15] of each language.

Thus, in section 2 we present the methodology used to represent expert information such as the collected expert decisions and the computed performance confidence indices. In section 3 we describe the empirical weighted fusing techniques while the uncertainty-based fusion models and methodologies are elaborated and

depicted in section 4. Experiments are explained in section 5.

## 2 Information Representation

### 2.1 Expert Decisions are Scores

ALI experts accept a speech utterance called the observation, as input, and provide the class (or language) decision as output, after computing language-score values; mostly a statistical model is used and the language score is the language likelihood; so that the experts handle a vector of language-likelihood values. Given  $M$  languages to identify,  $L_i, 1 \leq i \leq M$ , and  $N$  experts,  $S_j, 1 \leq j \leq N$ , we obtain for each observation,  $N$  vectors of  $M$  values, each one ranging from 0 to 1; the higher the value, the more confident the expert is that the corresponding language is the right one.

This global observation is represented as a score matrix:  $\delta = [d_{ij}]_{1 \leq i \leq M, 1 \leq j \leq N}$  (Table 1), where  $L = \{L_1, L_2, \dots, L_i, \dots, L_M\}$  is the set of languages and  $S = \{S_1, S_2, \dots, S_j, \dots, S_N\}$  the set of experts.

Table 1: Matrix  $\delta = [d_{ij}]_{1 \leq i \leq M, 1 \leq j \leq N}$ , of scores obtained for each observation.

| $L \setminus S$ | $S_1$    | $S_2$    | ... | $S_j$    | ... | $S_N$    |
|-----------------|----------|----------|-----|----------|-----|----------|
| $L_1$           | $d_{11}$ | $d_{12}$ | ... | $d_{1j}$ | ... | $d_{1N}$ |
| $L_2$           | $d_{21}$ | $d_{22}$ | ... | $d_{2j}$ | ... | $d_{2N}$ |
| ...             | ...      | ...      | ... | ...      | ... | ...      |
| $L_i$           | $d_{i1}$ | $d_{i2}$ | ... | $d_{ij}$ | ... | $d_{iN}$ |
| ...             | ...      | ...      | ... | ...      | ... | ...      |
| $L_M$           | $d_{M1}$ | $d_{M2}$ | ... | $d_{Mj}$ | ... | $d_{MN}$ |

### 2.2 Computation of Confidence Indices

Estimation of expert performance, with a view to provide the language identification process with heuristic-like information, can be achieved beforehand by means of an evaluation phase where the expert is tested on a set of segments whose language is known.

We split a global speech corpus into three partitions: a learning corpus  $X = \{x_{learn}\}$ , a test corpus  $Y = \{y_{test}\}$  and a development corpus  $Z = \{z_{dev}\}$ . We use the last one to compute the

two families of indices: the performance expert and class indices. In order to explain our future fusion techniques, it is necessary to define not only such indices, but also the observation performance confidence indices which represent for each expert the confidence of the decision taken for the observation. The two first families of indices are independent of the current observation.

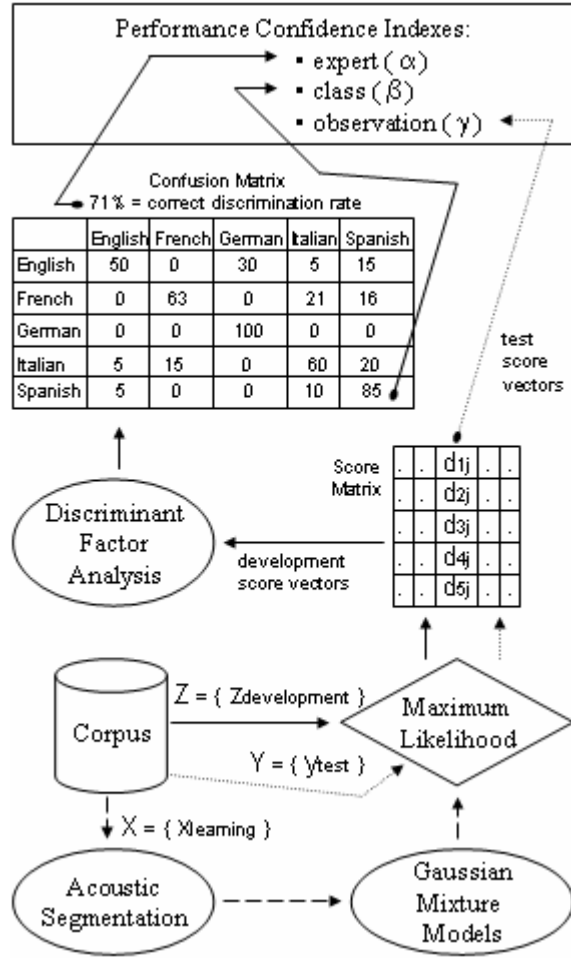


Figure 1: Computing expert ( $\alpha$ ), class ( $\beta$ ) and observation ( $\gamma$ ) confidence indices.

We collect the score matrices corresponding to the acoustic segments of the development corpus; each expert  $S_j$ ,  $1 \leq j \leq N$ , contributes with a score vector corresponding to an acoustic segment and is represented by column  $j$ , in the score matrix (Table 1). Then a matrix  $M_j$  (set of score vectors from expert  $S_j$ ) will correspond to several acoustic segments.

For each expert  $S_j$ , we apply the DFA statistical method to its matrix  $M_j$  in order to search for an

appropriate representation space for them and a way of obtaining performance confidence indices on a correct discrimination rate basis: we use the  $M-1$  factorial axis corresponding to the  $M-1$  eigen-values and project the set  $M_j$  of score vectors into this subspace. In building the corresponding confusion matrix (Figure 1), the class confidence indices ( $\beta_{ij, 1 \leq i \leq M}$ ) are directly mapped from the diagonal values of it while the expert confidence index must be computed as an averaged value:

$$\alpha_j = (1/M) \sum_{i \in [1, M]} \beta_{ij}.$$

Many solutions may be proposed to define the observation confidence indices. We retain two formulas to be applied to test-corpus matrices: given an identification expert  $S_j$  and  $\hat{i}$  the decision class,  $d_{ij} = \max_k (d_{kj})$ ,  $k \in [1, M]$ ,

- $g_j = d_{ij} - \max_{k \neq i} d_{kj}$ ;
- $g_j = d_{ij} - \frac{1}{M-1} \sum_{k \neq i} d_{kj}$ .

### 3 Empirical Fusion

The most current operations to empirically fuse decision scores are the so called linear and logarithmic ones that are respectively implemented by summing and multiplying score values. In addition, the estimated performance of each expert can be taken into account to weight its own scores in a heuristic-like way.

The concept of weighting by expert estimated performance [4] matches the one of weighting by the expert confidence index  $\alpha$  described above. Thus, a language is considered as the identified one if it corresponds to the greatest value computed with the following weighted rules:

- Sum  $L^* = \arg \max_{i \in [1, M]} [ \sum_{j \in [1, N]} \alpha_j d_{ij} ]$ ,
- Product  $L^* = \arg \max_{i \in [1, M]} [ \prod_{j \in [1, N]} d_{ij}^{\alpha_j} ]$ .

### 4 Modelling under Uncertainty

Let  $\Psi$  be a linguistic variable that is represented by a triplet (Figure 2) [2]:  $\Psi = (\delta, \mathbf{R}^{M \times N}, L)$ ;

- $\delta$  is a simple variable representing the score matrix, corresponding to an acoustic segment  $y$ , that is defined in the reference space  $\mathbf{R}^{M \times N}$ ;

•  $\mathbf{R}^{M \times N} = \{ x \mid x = [d_{ij}], 1 \leq i \leq M, 1 \leq j \leq N \}$  is the set of all score matrix values that  $\delta$  can take;

•  $L = \{ L_1, L_2, \dots, L_i, \dots, L_M \}$  is a finite set composed of fuzzy sets  $L_i$ , that is to say the set of different languages to be identified that characterise the variable  $\delta$  and define its value constraints in  $\mathbf{R}^{M \times N}$ .

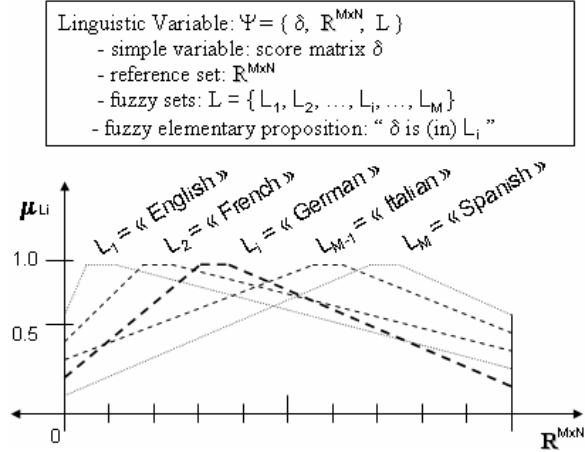


Figure 2: Automatic language identification is modelled as a linguistic variable concept.

$L_i$  is an infinite fuzzy set that is defined *a priori* by a membership function that associates to each element  $x \in \mathbf{R}^{M \times N}$  the degree  $m_i(x)$ , within the range  $[0,1]$ , with which  $x$  belongs to  $L_i$ :

$$m_i : \mathbf{R}^{M \times N} \rightarrow [0, 1];$$

and can be denoted either in ordered-pair notation:

$$L_i = \{ (m_i(x), x); x \in \mathbf{R}^{M \times N} \} = \{ (m_i([d_{ij}]), [d_{ij}]); [d_{ij}] \in \mathbf{R}^{M \times N}, 1 \leq i \leq M, 1 \leq j \leq N \},$$

or in additive continuous notation [2]:

$$L_i = \int_x m_i(x) / x = \int_{[d_{ij}]} m_i([d_{ij}]) / [d_{ij}].$$

Making an identification decision is figured out by means of fuzzy elementary propositions such as "(score matrix)  $d$  is (in language)  $L_i$ ". Such proposition is an *a posteriori* description that vaguely describes the language employed to pronounce an acoustic segment  $y$ , and it indicates the membership degree of the variable  $\delta$  to language  $L_i$ .

If for each language  $L_i$  we associate a possibility distribution to a fuzzy elementary proposition [2]:

$$\forall x \in \mathbf{R}^{M \times N}, \pi_{\delta, L_i}(x) = m_i(x);$$

then we will be able to make an identification decision after computing the possibility degree

$\pi_{\delta, L_i}(x)$  to which  $\delta$  belongs to each language  $L_i$ . This can be accomplished by means of directly applying uncertainty-based fusion techniques [8] on the matrix score values of the variable  $\delta$ .

Taking into account each score value  $d_{ij}$  in matrix  $\delta$  is a language-likelihood value, we can consider them as possibility values  $\pi(d_{ij})$  [5] [8] after normalising them:

$$\pi(d_{ij}) = d_{ij} / \max_{k \in [1, M]} d_{kj};$$

so that we can compute  $\pi_{\delta, L_i}(x)$  for each language  $L_i$  by means of fusing all possibility values in  $\delta$ .

#### 4.1 Possibility Theory

Before the fusion operation takes place, we exploit the *a priori* expert performance information provided by the class confidence indices  $(\beta_{ij}, 1 \leq i \leq M, 1 \leq j \leq N)$  to implement a hierarchical tree (Figure 3) of experts with a view to fuse their scores on a priority basis. The higher the performance of the experts at the class (language) level, the first they appear in the hierarchical tree. Each node of the tree comprises similarly-performing experts.

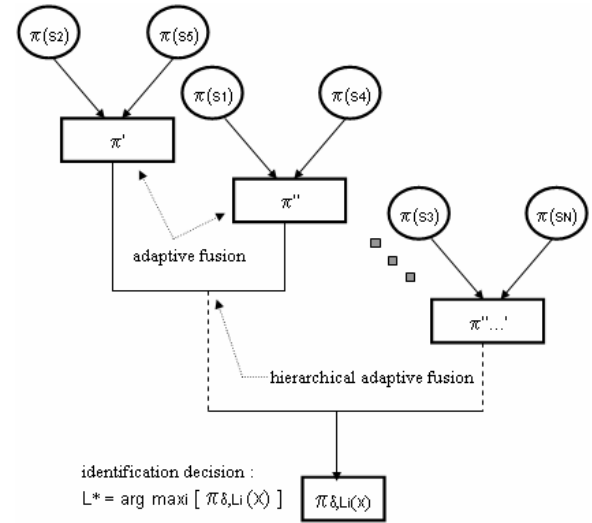


Figure 3: Hierarchical adaptive fusion of expert decisions

We apply the Adaptive Fusion [8] technique to fuse the score values issued from the experts that are inside each node of the tree. This technique implies computing the consistency index  $\gamma$  (a sort of observation confidence index) of the experts on a score matrix basis:

$$\gamma_{rk} = \sup_{L_i \in L} [ \min(\pi_{\delta, L_i}(S_r), \pi_{\delta, L_i}(S_k)) ],$$

so that conjunctive or disjunctive fusion can be done adaptively at the class level for each node:

$$\pi_{\delta, L_i}(X) = \max[ \pi_{\text{conj}, \delta, L_i}(S_r, S_k) / \gamma_{rk}, \min(1 - \gamma_{rk}, \pi_{\text{disj}, \delta, L_i}(S_r, S_k)) ].$$

The rules employed for conjunctive and disjunctive fusion are:

$$\pi_{\text{conj}, \delta, L_i}(S_r, S_k) = \min(\pi_{\delta, L_i}(S_r), \pi_{\delta, L_i}(S_k)) ;$$

$$\pi_{\text{disj}, \delta, L_i}(S_r, S_k) = \max(\pi_{\delta, L_i}(S_r), \pi_{\delta, L_i}(S_k)) .$$

Results from pairs of adjacent nodes are fused in an adaptive way as well. We start the fusion process from the upper node and end up with the lower node so that a global possibility value  $\pi_{\delta, L_i}(X)$  is obtained as result. We compute the consistency index  $\gamma$  and apply the adaptive rule the same way we explained above, but the conjunctive and disjunctive rules [8] between nodes are respectively the following:

$$\pi_{\text{conj}, \delta, L_i}(X)^{\prime\prime} = \min[ \pi_{\delta, L_i}(X)^{\prime}, \max(\pi_{\delta, L_i}(X)^{\prime\prime}, 1 - \gamma^{\prime\prime}) ] ;$$

$$\pi_{\text{disj}, \delta, L_i}(X)^{\prime\prime} = \max[ \pi_{\delta, L_i}(X)^{\prime}, \min(\pi_{\delta, L_i}(X)^{\prime\prime}, \gamma^{\prime\prime}) ] .$$

Having  $M$  languages  $L_i$ , we compute  $M$  global possibility distribution functions  $\pi_{\delta, L_i}(X)$  to make an identification decision by considering as the identified language the one that has been assigned to the score-matrix variable  $\delta$  with the maximum possibility degree:

$$L^* = \arg \max_i [ \pi_{\delta, L_i}(X) ] .$$

## 4.2 Theory of Evidence

Let  $L = \{ L_1, L_2, \dots, L_i, \dots, L_M \}$  denote the finite set of possible languages to be identified; this set  $L$  is composed of  $M$  exhaustive and exclusive hypotheses of the decision process and we assume every union of hypotheses may be a response of the decision process. The set  $\mathcal{L}$  of all possible events  $A$  based on  $L$  is the set of all subsets of  $L$ ,  $2^L = \{ A \mid A \subseteq L \}$ ,  $|2^L| = 2^M$ , that is to say:

$$2^L = \{ \emptyset, \{L_1\}, \{L_2\}, \dots, \{L_i\}, \dots, \{L_M\}, \{L_1, L_2\}, \dots, \{L_{M-1}, L_M\}, \dots, L \} .$$

For each unknown utterance, and for each expert  $S_j$ , we define a basic belief mass function  $m_L^{S_j}$ , which explains how the decision  $L^*$  belongs to the subset  $A$  of  $L$ :  $m_L^{S_j} : 2^L \rightarrow [0, 1]$  with the constraints:

$$\sum_{A \subseteq L} m_L^{S_j}(A) = 1 \text{ and } m_L^{S_j}(\emptyset) = 0 .$$

The basic belief mass function  $m_L^{S_j}$  is built from the score matrix values of the utterance; we assign basic belief mass values from the distances between their corresponding possibility values [1] [5] [7].

Let  $A_k$  represent an event  $A$  in position  $k$  when all the singleton events have been arranged in decreasing order taking into account its corresponding possibility value  $\pi_k$ . In the case of events that are different than singletons, the corresponding possibility value is the minimum value found among the several possibility values that correspond to the participating singletons [5].

If  $\pi_1 = 1 > \pi_2 > \dots > \pi_k > \dots > \pi_M > \pi_{M+1} = 0$ ; then for any non-empty set  $A$ :

$$m_L^{S_j}(A_k) = \pi_k - \pi_{k+1} ;$$

but:  $m_L^{S_j}(A_k) = 0$  if  $A_k$  represents  $\emptyset$ .

In order to verify the constraints above, we normalise all the belief values after computing a normalisation factor:

$$R_j = 1 / \sum_{A_k \subseteq L} m_L^{S_j}(A_k) ;$$

and we apply it as a multiplying factor:

$$m_L^{S_j}(A_r) = R_j m_L^{S_j}(A_k) ; \forall A_r \subseteq L .$$

Thus the set of focal elements includes all the subsets  $A$  such as its corresponding  $m_L^{S_j}(A_r) > 0$ .

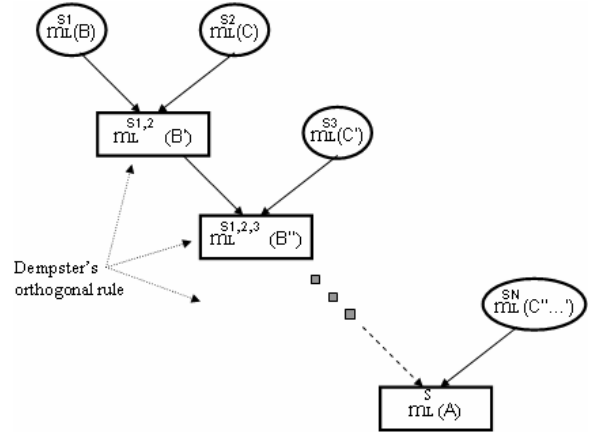


Figure 4: Cascade-like application of Dempster's orthogonal rule.

Let  $(S_k, S_r)$  represent any pair of the  $N$  experts, we may combine the belief mass values of the focal elements  $(B, C, \text{etc.})$  of these experts on a cascade-like pair basis (Figure 4) by applying the Dempster's orthogonal combination rule:

$$m_L^{S_r, k}(A) = K_L \cdot \sum_{B \cap C = A} m_L^{S_k}(B) \cdot m_L^{S_r}(C) ;$$

where  $\kappa_L = 1 / [1 - \sum_{B \cap C = \emptyset} m_L^{S_k}(B) \cdot m_L^{S_r}(C)]$  is a normalisation factor taking into account the case where the empty set results from conjoining focal elements (Figure 5). We obtain thus a global belief mass function, noted  $m_L^S(A)$ , for each event A.

We weight basic belief mass functions [1] of the events (B, C, etc.) by discounting the expert and class confidence indices (respectively  $\alpha$  and  $\beta$ ) before normalising to do the orthogonal operation:

$$\begin{aligned} m_L^{S_i, \beta_{ij}}(C) &= \beta^{ij} \cdot m_L^{S_i}(C), \forall C \neq L, |C| = 1; \\ m_L^{S_i, \alpha_j}(C) &= \alpha_j \cdot m_L^{S_i}(C), \forall C \neq L, |C| > 1; \\ m_L^{S_i, \alpha_j}(L) &= (1 - \alpha_j) + \alpha_j \cdot m_L^{S_i}(L). \end{aligned}$$

|                      |                      |                      |             |                |
|----------------------|----------------------|----------------------|-------------|----------------|
| $m_L^{S_i}(L)$       | $\gamma^i$           | $\gamma^i$           | $\gamma^i$  | $\gamma^i$     |
| .....                | $\emptyset$          | $\emptyset$          | $\gamma^i$  | $\gamma^i$     |
| $m_L^{S_i}(\{L_2\})$ | $\emptyset$          | $\gamma^i$           | $\emptyset$ | $\gamma^i$     |
| $m_L^{S_i}(\{L_1\})$ | $\gamma^i$           | $\emptyset$          | $\emptyset$ | $\gamma^i$     |
|                      | $m_L^{S_i}(\{L_1\})$ | $m_L^{S_i}(\{L_2\})$ | .....       | $m_L^{S_i}(L)$ |

Figure 5: Orthogonal combination of basic belief mass values of expert focal elements.

In order to make a language identification decision, we use the pignistic transformation [15] to derive a probability on L, from the belief mass values:

$$\text{BetP}(L_i) = \sum_{L \in A} m_L^S(A) / |A|.$$

Thus, the decision process can be carried out by maximum pignistic probability [6]:

$$L^* = \arg \max_i [ \text{BetP}(L_i) ].$$

## 5 Experimentation

### 5.1 Preliminaries

Acoustic data is provided by the MULTTEXT corpus [3] which comprises a set of 20 kHz 16-bit sampled records in 5 languages: English, French, German, Italian and Spanish. Data consists of read passages from the EUROM1 corpus pronounced by 50 different speakers (5 males and 5 females per language). The mean duration of each passage is 20.8 seconds.

The global corpus is split into three partitions for each language: the learning corpus, the development corpus and the test corpus (2 speakers: 1 male and 1 female who do not belong to the other corpora).

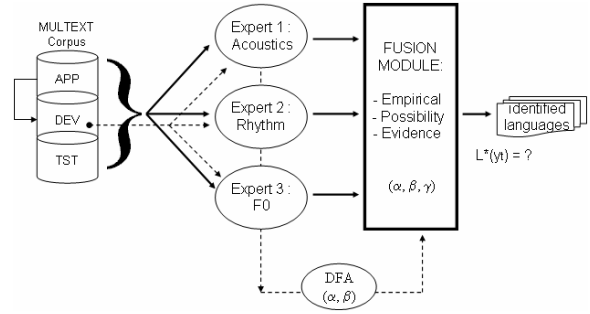


Figure 6: Architecture of the Fusion System.

The ALI system is based on three ALI experts and a fusion module (see Figure 6):

- **Acoustics Expert [12]:** After an automatic vowel detection, each vocalic segment is represented with a set of 8 Mel-Frequency Cepstral Coefficients and 8 delta-MFCC, augmented with the Energy and delta Energy of the segment. This parameter vector is extended with the duration of the underlying segment providing a 19-coefficient vector. A cepstral subtraction performs both blind removal of the channel effect and speaker normalisation. For each recording sentence, the average MFCC vector is computed and subtracted from each coefficient.

- **Rhythm Expert [14]:** Syllable may be a first-rate candidate for rhythm modelling. Nevertheless, segmenting speech in syllables is typically a language-specific mechanism and thus no language independent algorithm can be derived. For this reason, we have introduced the notion of pseudo-syllables derived from the most frequent syllable structure in the world, namely the CV structure. Using the vowel-non vowel segmentation, speech signal is parsed in patterns matching the structure:  $.C^nV$ . Each pseudo-syllable is then characterised by its consonants global duration, its vocalic duration, its complexity (the number of consonant segments), and its energy.

- **Fundamental Frequency Expert [14]:** The fundamental frequency outlines are used to compute statistics within the same pseudo-syllable frontiers (previously defined) in order to model intonation on each pseudo-syllable. The parameters used to characterise each pseudo-syllable intonation are a measurement of the accent location (maximum f0 location in regard to vocalic onset) and the normalised fundamental frequency bandwidth on each syllable.

For each expert, we applied the same learning-testing procedure: for each language, a Gaussian Mixture Model (GMM) is trained using EM algorithm with LBG initialisation [11]. The optimal number of components of the mixture is obtained from experiments on the learning part of the corpus. During the test, the decision relies on a Maximum Likelihood procedure.

The performance of these three experts is given in Table 2, and is considered as a reference to be compared with. We may observe the relatively bad performance of: the fundamental frequency-based expert in general and the three experts on the test set number two (see next section) in particular.

## 5.2 Tests

Three sets of the test corpus (2 speakers out of 10: 1 male and 1 female) are selected and tested on a round-robin basis with a view to analyse the fusion system behaviour over representative expert performance data of good (set 1) and rather-bad examples (sets 2 and 3).

The three techniques of fusion (empirical, possibility-based and evidential ones) are experimented to merge the decision scores (outputs of the three experts) as explained in the previous sections.

The development corpus is used to compute the class and expert performance confidence indices while the test corpus is used to compute the observation index. The information provided by these indices drive in a heuristic-like way the uncertainty-based inference.

The empirical fusion techniques are tested in their non-weighted and weighted versions. The expert confidence index is used for the weighted versions.

Minimum and maximum operations are selected and tested as conjunctive and disjunctive possibility-based aggregation techniques; we use them while applying the adaptive fusion technique explained above.

Regarding the evidential fusion techniques, two versions of focal element sets are tested depending on what events can participate to compose them: I) any event  $A \subseteq L$  is eligible; and II) any event  $A \subseteq L$  such that  $|A|=1$  and the event  $A=L$  are eligible.

Furthermore, 2-expert fusion is also tested to observe which combinations could provide better results and how efficient the fusion techniques were in obtaining the best identification rates when combining 3 experts at a time.

## 5.3 Results

Most important results in fusing the three experts are the following (see Table 2):

- The empirical fusion delivers better identification rates than those of any expert for sets 1 and 3 (up to 84%), but for set 2. Weighted versions work out better than non-weighted versions for set 1 (good-example data) only.
- The possibility fusion generally attains a good identification-rate delivery level: up to 85%. But it fails in set 2 (bad-example data).
- Excepting the evidential fusion (version II), all the others fail in set 2 (bad-example data). The performance of evidential fusion version II is better than version I for bad-example data (where the incoherence degree between experts is too high: from 0.5 to 0.9).
- The best identification rates are reached by the fusion system using the evidential method with data from either the good-example set or the bad-example set (version II only): up to 90%.
- Regarding the 2-expert fusion, we observe that two combinations barely deliver better identification rates than the 3-expert combination for the empirical (experts 2 and 3, set 1: 85%) and possibility (experts 1 and 2, set 2: 65%) fusion approaches. This scenario does not take place for the evidential fusion.

Table 2: Results of Fusion Strategies.

| Language Identification System<br>– Total Success Rate –<br>(%) |   |         |         |         |
|---|---|---------|---------|---------|
|   |   | 1st set | 2nd set | 3rd set |
| Reference Experts   | Expert 1 :<br>acoustics                       | 79      | 41      | 62      |
|   | Expert 2 :<br>rhythm                          | 71      | 63      | 60      |
|   | Expert 3 :<br>fundamental frequency           | 35      | 37      | 48      |
| Empirical Fusion Techniques                                     | Addition                                      | 83      | 60      | 69      |
|   | Product                                       | 67      | 49      | 68      |
|   | Weighted Addition                             | 84      | 58      | 67      |
|   | Weighted Product                              | 68      | 51      | 67      |
| Uncertainty-Based Fusion Methods                                | Possibility Theory                            | 85      | 52      | 69      |
|   | Theory of Evidence<br>version I<br>version II | 90      | 56      | 69      |
|   |   | 90      | 64      | 75      |

## 6 Conclusion

Uncertainty-based fusion methods can be applied properly to model the language identification expert process of interaction in the presence of robust confidence indices that reflect *a priori* knowledge on expert performance, like those computed by the Discriminant Factor Analysis method. This fusion methodology comes out as a formal strong alternative to empirical techniques.

Both Possibility and Evidence Theories provide us with inference techniques that can take advantage of weighting values in a more refined way: not only at the expert level but also at the class and observation levels, so that they will generally deliver better identification rates compared to empirical techniques.

Future works could include experimenting with: a) other conjunctive and disjunctive operations in the possibility/fuzzy domain: Lukasiewicz, Hamacher or Weber; and b) possibility-to-probability transformations [7] in search of a common risk-based function to make fused decisions in the probabilistic domain (note that the pignistic probability has already been computed from the evidential domain).

## References

- [1] Appriou A. Multisensor signal processing in the framework of the theory of evidence. In *NATO/RTO – Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission systems*, 1999.
- [2] Bouchon-Meunier B. Théorie des possibilités et variables linguistiques. In *La Logique Floue et ses Applications*. Addison-Wesley, Paris, 1995.
- [3] Campione E. and Véronis J. A multilingual prosodic database. In *Proceedings of the conference ICSLP'1998*, Sidney, Australia, 1998.
- [4] Cooke R.M. Experts in uncertainty. Oxford University Press, Oxford, United Kingdom, 1991.
- [5] Denoeux T. and Zouhal L.M. Handling possibilistic labels in pattern classification using evidential reasoning. In *Fuzzy Sets and Systems*, volume 122(3), pages 409-424, 2001.
- [6] Denoeux T. Pattern Recognition using belief function. In *Proceedings of the conference SFC'2002*, Toulouse, France, 2002.
- [7] Dubois D., Prade H. and Sandri S. On possibility-probability transformations. In *Fuzzy Logic*, Lowen R. and Roubens M., Kluwer Academic, pages 103-112, Dordrecht, Holland, 1993.
- [8] Dubois D. and Prade H. Possibility theory and data fusion in poorly informed environments. In *Control Engineering Practice*, volume 2(5), pages 811-823, 1994.
- [9] Hazen T.J., and Zue V.W. Segmented-based Automatic Language Identification. *Journal of the Acoustical Society of America*, 4(101), 1997.
- [10] Kittler J., Hojjatoleslami A.J. and Windeatt T. Weighting Factors in multiple expert fusion. In *Proceedings of the conference BMVC'97*, pages 41-50, Essex University, United Kingdom, 1997.
- [11] Linde Y., Buzo A. and Gray R.M. An algorithm for vector quantizer design. *IEEE Transaction on Communications*, volume 28, no. 1, pages 84-95, 1980.
- [12] Pellegrino F., André-Obrecht R. Automatic language identification: an alternative approach to phonetic modelling. In *Signal Processing*, Elsevier Science North Holland, volume 80, pages 1231-1244, 2000.
- [13] Rahman A. and Fairhurst M. A novel confidence-based framework for multiple expert decision fusion. In *Proceedings of the conference BMVC'98*, University of Southampton, United Kingdom, 1998.
- [14] Rouas J.L., Farinas J. and Pellegrino F. Automatic modelling of rhythm and intonation for language identification. In *15th International Congress of Phonetic Sciences (15th ICPhS)*, 2003, pages 567-570, Barcelona, Spain, 2003.
- [15] Smets P. Constructing the pignistic probability function in a context of uncertainty. In *Uncertainty in Artificial Intelligence 5*, Elsevier Science North-Holland, pages 29-39, 1990.
- [16] Zissman M. and Berkling K.M. Automatic language identification. In *Speech Communication*, volume 35, pages 115-124, 2001.