# A STUDY ON IMPROVING DECISIONS IN CLOSED SET SPEAKER IDENTIFICATION

*Mübeccel Demirekler and Afşar Saranlı*

Department of Electrical and Electronics Engineering
Middle East Technical University, Ankara 06531, Türkiye.
e-mail: demirek@rorqual.cc.metu.edu.tr , afsars@rorqual.cc.metu.edu.tr

## ABSTRACT

In this study, closed-set, text-independent speaker identification is considered and the problem of improving the reliability of the decisions made by available algorithms is addressed. The work presented here is based on the idea of combining the *evidences* from different algorithms or decision strategies to improve the recognition performance and the reliability. For this purpose, the models generated by a single algorithm for 17 speakers from the SPIDRE database are considered and a matrix of speaker-to-model fitness values is processed by two different decision strategies. Ideas from the *Mathematical Theory of Evidence* are applied to combine the decisions produced by these two strategies to generate a better decision on the speaker identity. The combined decision show an improved degree of corectness hence suggesting a promising way of combining the decisions from partially successful algorithms.

## 1. INTRODUCTION

Closed-set text-independent speaker recognition deals with the identification of a speaker which is known to belog to a closed set of speakers by using an utterance of his/her speech. The speech uterance supplied for the recognition system is arbitrary and previously unknown to the system. A number of methods such as *Vector Quantization*(VQ) or *Second Order Statistics* exist in the literature to deal with this problem. However, all the algorithms are known to generate wrong identifications especially when training and test conditions(channel, microfone etc.) differ or the data is corrupted with noise.

In this work we will not be interested in the improvement of the methods themselves but rather with a new method of interpreting and combining the results given by one or more of the these identification methods. Our main objective is to use the *evidences* given by different methods to produce a single decision about the speaker identity with an improved degree of correctness and reliability. The ideas behind the method of combining the evidences given by different algorithms are taken from the theory developed mainly by Dempster[2][3] and Shafer[1], known as *Mathematical Theory of Evidence*. Some related work can be found in [4][5][6].

## 2. DESCRIPTION OF THE PROBLEM

Suppose there are $N$ persons in the set of speakers and a test speech utterance should be classified as belonging to one of them. Suppose also that by *some* method, models of these $N$ speakers are obtained. The same training data is then applied to all of these models leading to *fitness values* indicating the closeness of each training utterance to all of the available models. Table 1 illustrates a typical matrix composed of these values for 17 speakers taken from the SPIDRE database. The table values are in the range $[0, 100]$ and a higher correlation between a speaker and a model is characterized with a higher value. For some speakers, the models are *well fitted* with the training data so that a high fitness value is observed only between the utterance of the speaker and his own model while all the remaining fitness values are comparatively small. A typical example in Table 1 is speaker $Sp_1$, which has a fitness value of 81 with his own model while the next best fit is with the model $M_2$ of speaker $Sp_2$, with a fitness value of 38. This is approximately 48% of 81. Unfortunately, the situation is not so lucky for some other cases. Specifically, there are also some *bad models* like $M_{13}$, for which one obtains very small fitness values even for an utterance from its own speaker. For example, model $M_{13}$ *fits* to the training utterance from its own speaker with a fitness value of 4 while another model, ($M_{14}$) gives a fitness value of 39. An interesting point to note is that these type of *bad* models generally give small fitness values for training utterances *from all speakers* including their own. (E.g., Observe column $M_{13}$)

| Training | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | M13 | M14 | M15 | M16 | M17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sp1 | 81 | 38 | 2 | 13 | 0 | 9 | 0 | 6 | 5 | 15 | 17 | 11 | 1 | 15 | 13 | 31 | 28 |
| Sp2 | 48 | 49 | 7 | 14 | 2 | 11 | 0 | 5 | 9 | 28 | 18 | 10 | 0 | 38 | 11 | 29 | 29 |
| Sp3 | 6 | 2 | 59 | 9 | 11 | 4 | 3 | 25 | 10 | 16 | 32 | 33 | 0 | 26 | 45 | 10 | 4 |
| Sp4 | 42 | 13 | 7 | 54 | 4 | 12 | 0 | 8 | 17 | 11 | 8 | 10 | 0 | 25 | 36 | 29 | 17 |
| Sp5 | 19 | 6 | 12 | 18 | 24 | 12 | 1 | 37 | 11 | 8 | 17 | 41 | 0 | 13 | 39 | 7 | 10 |
| Sp6 | 13 | 5 | 16 | 12 | 2 | 22 | 4 | 17 | 4 | 3 | 47 | 44 | 0 | 21 | 45 | 11 | 11 |
| Sp7 | 3 | 2 | 34 | 2 | 8 | 3 | 32 | 45 | 3 | 7 | 60 | 65 | 0 | 16 | 48 | 4 | 3 |
| Sp8 | 9 | 4 | 25 | 12 | 8 | 9 | 2 | 53 | 3 | 6 | 21 | 52 | 0 | 18 | 53 | 8 | 11 |
| Sp9 | 18 | 9 | 25 | 22 | 5 | 1 | 0 | 1 | 51 | 20 | 16 | 5 | 2 | 62 | 21 | 48 | 2 |
| Sp10 | 32 | 22 | 22 | 13 | 5 | 5 | 0 | 4 | 11 | 55 | 22 | 9 | 0 | 43 | 14 | 31 | 17 |
| Sp11 | 7 | 7 | 20 | 2 | 1 | 2 | 2 | 18 | 2 | 5 | 80 | 49 | 0 | 24 | 32 | 8 | 6 |
| Sp12 | 4 | 4 | 22 | 1 | 2 | 2 | 18 | 53 | 2 | 2 | 68 | 76 | 0 | 19 | 56 | 9 | 4 |
| Sp13 | 32 | 20 | 12 | 11 | 2 | 5 | 0 | 6 | 6 | 19 | 27 | 14 | 4 | 39 | 19 | 38 | 8 |
| Sp14 | 20 | 18 | 16 | 12 | 4 | 3 | 1 | 6 | 11 | 20 | 30 | 18 | 1 | 60 | 22 | 30 | 9 |
| Sp15 | 17 | 3 | 19 | 20 | 9 | 12 | 1 | 28 | 11 | 5 | 17 | 34 | 0 | 19 | 74 | 25 | 7 |
| Sp16 | 19 | 7 | 13 | 22 | 1 | 2 | 0 | 1 | 31 | 13 | 12 | 3 | 0 | 68 | 37 | 82 | 2 |
| Sp17 | 33 | 23 | 9 | 10 | 1 | 16 | 0 | 16 | 3 | 14 | 28 | 26 | 0 | 31 | 16 | 17 | 47 |

Table 1. Model fitness values for the training data set

To propose our method of combining evidences, we consider the data in Table 1 which contains the fitness values generated by an "unknown" speaker modelling algorithm. The algorithm is not of much importance for the work presented here. However, two points should be mentioned: Firstly, we assume that the resulting table contains information which is sufficient for the discrimination of most of the speakers, although this discrimination may not be made by a trivial decision rule. Secondly, we observe that different decision strategies(algorithms) lead to the correct identification of different sets of speakers while mis-identifying others. I.e., a correctly identified speaker for one algorithm may be mis-identified by another algorithm.

Based on this observation two different decision strategies are considered here as illustrative examples. The novel method considered in this work is then applied on the decisions given by these two strategies to combine their *evidences* as a better decision rule with reduced mis-identification rate. Note that these two methods are not the only ones that can be proposed. Also, the algorithm generating Table 1 is not the only one that can be

| Test | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | M13 | M14 | M15 | M16 | M17 | Chk |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sp1 | 79 | 31 | 3 | 15 | 0 | 11 | 0 | 8 | 3 | 17 | 24 | 14 | 1 | 10 | 14 | 26 | 31 | √ |
| Sp2 | 48 | 41 | 8 | 15 | 1 | 4 | 0 | 2 | 10 | 20 | 27 | 6 | 3 | 42 | 20 | 46 | 13 | × |
| Sp3 | 3 | 1 | 62 | 2 | 4 | 2 | 8 | 30 | 4 | 6 | 53 | 46 | 0 | 33 | 56 | 16 | 2 | √ |
| Sp4 | 17 | 4 | 30 | 38 | 6 | 6 | 0 | 4 | 14 | 25 | 17 | 10 | 0 | 39 | 37 | 34 | 3 | × |
| Sp5 | 4 | 1 | 35 | 24 | 28 | 6 | 2 | 12 | 18 | 15 | 15 | 20 | 0 | 42 | 52 | 19 | 1 | × |
| Sp6 | 44 | 29 | 3 | 11 | 1 | 41 | 0 | 15 | 2 | 15 | 14 | 22 | 0 | 16 | 9 | 7 | 47 | × |
| Sp7 | 3 | 2 | 42 | 1 | 11 | 2 | 23 | 50 | 2 | 10 | 53 | 65 | 0 | 16 | 43 | 4 | 3 | × |
| Sp8 | 14 | 5 | 20 | 3 | 4 | 6 | 2 | 66 | 1 | 4 | 33 | 65 | 0 | 6 | 36 | 4 | 16 | √ |
| Sp9 | 16 | 4 | 35 | 15 | 3 | 1 | 0 | 1 | 43 | 20 | 24 | 6 | 2 | 59 | 24 | 47 | 0 | × |
| Sp10 | 26 | 11 | 29 | 9 | 9 | 2 | 0 | 1 | 19 | 64 | 20 | 3 | 1 | 45 | 13 | 36 | 6 | √ |
| Sp11 | 9 | 8 | 19 | 1 | 1 | 3 | 4 | 17 | 2 | 6 | 79 | 45 | 0 | 24 | 28 | 13 | 6 | √ |
| Sp12 | 27 | 17 | 7 | 4 | 3 | 14 | 1 | 43 | 3 | 9 | 21 | 49 | 0 | 11 | 20 | 7 | 28 | √ |
| Sp13 | 34 | 35 | 13 | 4 | 1 | 2 | 1 | 9 | 5 | 15 | 44 | 18 | 14 | 37 | 16 | 31 | 8 | × |
| Sp14 | 18 | 7 | 15 | 26 | 2 | 6 | 0 | 2 | 31 | 19 | 10 | 7 | 0 | 70 | 29 | 53 | 5 | √ |
| Sp15 | 4 | 1 | 31 | 9 | 2 | 4 | 3 | 33 | 10 | 1 | 48 | 57 | 0 | 25 | 88 | 15 | 1 | √ |
| Sp16 | 30 | 34 | 17 | 6 | 1 | 2 | 0 | 2 | 8 | 17 | 23 | 7 | 2 | 62 | 18 | 63 | 9 | √ |
| Sp17 | 28 | 25 | 12 | 6 | 1 | 10 | 0 | 9 | 2 | 22 | 36 | 18 | 0 | 41 | 17 | 26 | 37 | × |
| $w$ | 404 | 256 | 381 | 189 | 78 | 122 | 44 | 304 | 177 | 285 | 541 | 458 | 23 | 578 | 520 | 447 | 216 | |

Table 2. Fitness values for the test data and the normalization coefficients $w$.

| Normalized Test | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | M13 | M14 | M15 | M16 | M17 | Chk |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sp1 | 19.6 | 13.4 | 0.9 | 6.1 | 0.0 | 8.5 | 0.0 | 2.4 | 1.6 | 6.9 | 4.6 | 2.8 | 2.0 | 1.9 | 2.4 | 6.2 | 14.4 | √ |
| Sp2 | 11.9 | 17.7 | 2.5 | 6.1 | 1.1 | 3.1 | 0.0 | 0.6 | 5.3 | 8.1 | 5.2 | 1.2 | 6.0 | 7.8 | 3.4 | 11.0 | 6.0 | √ |
| Sp3 | 0.7 | 0.4 | 19.4 | 0.8 | 4.5 | 1.5 | 12.5 | 9.1 | 2.1 | 2.4 | 10.2 | 9.2 | 0.0 | 6.1 | 9.6 | 3.8 | 0.9 | √ |
| Sp4 | 4.2 | 1.7 | 9.4 | 15.4 | 6.7 | 4.6 | 0.0 | 1.2 | 7.4 | 10.1 | 3.3 | 2.0 | 0.0 | 7.3 | 6.4 | 8.2 | 1.4 | √ |
| Sp5 | 1.0 | 0.4 | 10.9 | 9.7 | 31.5 | 4.6 | 3.1 | 3.6 | 9.5 | 6.1 | 2.9 | 4.0 | 0.0 | 7.8 | 9.0 | 4.6 | 0.5 | √ |
| Sp6 | 10.9 | 12.5 | 0.9 | 4.5 | 1.1 | 31.5 | 0.0 | 4.6 | 1.1 | 6.1 | 2.7 | 4.4 | 0.0 | 3.0 | 1.5 | 1.7 | 21.9 | √ |
| Sp7 | 0.7 | 0.9 | 13.1 | 0.4 | 12.4 | 1.5 | 35.9 | 15.2 | 1.1 | 4.0 | 10.2 | 13.0 | 0.0 | 3.0 | 7.4 | 1.0 | 1.4 | √ |
| Sp8 | 3.5 | 2.2 | 6.3 | 1.2 | 4.5 | 4.6 | 3.1 | 20.1 | 0.5 | 1.6 | 6.3 | 13.0 | 0.0 | 1.1 | 6.2 | 1.0 | 7.4 | √ |
| Sp9 | 4.0 | 1.7 | 10.9 | 6.1 | 3.4 | 0.8 | 0.0 | 0.3 | 22.6 | 8.1 | 4.6 | 1.2 | 4.0 | 11.0 | 4.1 | 11.3 | 0.0 | √ |
| Sp10 | 6.5 | 4.7 | 9.1 | 3.6 | 10.1 | 1.5 | 0.0 | 0.3 | 10.0 | 25.9 | 3.8 | 0.6 | 2.0 | 8.4 | 2.2 | 8.6 | 2.8 | √ |
| Sp11 | 2.2 | 3.4 | 5.9 | 0.4 | 1.1 | 2.3 | 6.3 | 5.2 | 1.1 | 2.4 | 15.2 | 9.0 | 0.0 | 4.5 | 4.8 | 3.1 | 2.8 | √ |
| Sp12 | 6.7 | 7.3 | 2.2 | 1.6 | 3.4 | 10.8 | 1.6 | 13.1 | 1.6 | 3.6 | 4.0 | 9.8 | 0.0 | 2.0 | 3.4 | 1.7 | 13.0 | × |
| Sp13 | 8.4 | 15.1 | 4.1 | 1.6 | 1.1 | 1.5 | 1.6 | 2.7 | 2.6 | 6.1 | 8.5 | 3.6 | 28.0 | 6.9 | 2.8 | 7.4 | 3.7 | √ |
| Sp14 | 4.5 | 3.0 | 4.7 | 10.5 | 2.2 | 4.6 | 0.0 | 0.6 | 16.3 | 7.7 | 1.9 | 1.4 | 0.0 | 13.0 | 5.0 | 12.7 | 2.3 | × |
| Sp15 | 1.0 | 0.4 | 9.7 | 3.6 | 2.2 | 3.1 | 4.7 | 10.0 | 5.3 | 0.4 | 9.2 | 11.4 | 0.0 | 4.7 | 15.1 | 3.6 | 0.5 | √ |
| Sp16 | 7.4 | 14.7 | 5.3 | 2.4 | 1.1 | 1.5 | 0.0 | 0.6 | 4.2 | 6.9 | 4.4 | 1.4 | 4.0 | 11.5 | 3.1 | 15.1 | 4.2 | √ |
| Sp17 | 6.9 | 10.8 | 3.8 | 2.4 | 1.1 | 7.7 | 0.0 | 2.7 | 1.1 | 8.9 | 6.9 | 3.6 | 0.0 | 7.6 | 2.9 | 6.2 | 17.2 | √ |

Table 3. Normalized Fitness Values for the test data set.

applied to this particular data set. Therefore the methodology given here can be applied to an set of algorithms. The aim of this work is mainly to show on the example of the proposed two methods that decisions given by such different methods can be combined in a systematic way to result in better decisions.

The first decision strategy considered is based on the assumption that the models are good. Hence a high fitness value between an unknown voice and the model indicates the correct model, i.e., the correct speaker. Generally, this is the basic idea behind almost all speaker recognition systems. As an example, suppose that we are given *unknown* speaker utterances which are not used in the training set. Consider Table 2 which contains this time, the fitness values for this *test data* from the speakers of our closed set. Again, Row 1 of Table 2 gives the fitness of each model to speaker $Sp_1$. If the highest fitness value identifies the speaker as stated by the first decision strategy, then the correct model $M_1$ would be obtained. The last column in Table 2 shows the identification result(Correct/Wrong). Only 9 speakers are correctly identified with this strategy, while the remaining 8 speakers are associated with wrong models.

The second strategy considered is based on the idea of distinguishing the *good* models from the *bad* ones and give the decision after a normalization operation of the test data with respect to the *goodness* of the model. To exploit this idea, assume that this goodness is measured by the sum of the elements of each column interpreted as a *goodness measure* for the corresponding model. A large sum indicates that the model under test is well fitted to a large number of the test utterances whereas a small sum indicates that it is not well fitted to any utterance. These sums are computed as Row $w$ of Table 2. The normalization of any test data with this value $w$, is obtained by dividing each element of every column of Table 2 by the corresponding sum in Row $w$. The reasoning used in this approach can be observed by investigating the case of speaker $Sp_{13}$. Before the normalization the fitness of $Sp_{13}$ to model $M_{13}$ is only 14 while its fitness, for example to the model $M_{11}$ is 44, which is far above the former value. However when the fitness values generated by the model $M_{13}$ are examined for the training data, it is observed that the these values are very small for utterances from all the speakers, including the utterance from its own speaker. Indeed the model still generates its best fitness value for speaker $Sp_{13}$. This observation indicates that the normalization described above may lead to improved results especially for this kind of *bad models*. In agreement with these observations, after the normalization, Table 3 is obtained and Row 13 clearly indicates that $M_{13}$ gives the largest fitness. Although this new method seems to solve the problem of $Sp_{13}$ it may not solve the problems of other speakers and even some new problems may be introduced. The overall investigation of Table 3 shows that for this method, all speakers except $Sp_{12}$ and $Sp_{14}$ are correctly identified. One should, note that these two speakers were correctly identified for the former method.

Now method of combining the results obtained by using the above two decision strategies is presented. We note that these two methods are different interpretations of the data contained in the fitness table and hence are different identification algorithms. Many other different decision strategies may be applied which will in turn give different results. Still, the results of each of these methods carries a considerable amount of information about the correct decision. Therefore the procedure that will be used for the combination of the decisions is not only applicable to the results of the above mentioned decision strategies but it may also be used to combine the results of some other decision strategy. Indeed, it may also be used to combine decisions from entirely different algorithms applied to the same speaker set.

## 3. METHOD OF COMBINATION OF DECISIONS

In order to be able to apply the *rule of combination* described by Dempster[1], one has to define first, the so called *separable support functions* corresponding to each criteria. Then, these support functions will be combined. To define the support functions, the first step is the generation of the event sets for the given algorithm. This is followed by the assignment of a *weight of evidence* for each event set. There is no unique way of either generating the event sets or assigning weights of evidence to these sets. Different approaches are possible. However, a sensible way is to define the the event sets is by making use of some *simple intuitive rules* consistent with our problem. Hence, this part of the system can be considered as a rule based system with simple rules as discussed below. Clearly, to obtain better and more complicated rules, one must work with a large number of experiments. However, simple rules can be derived from the amount of data we consider for this work. The generation of such rules is a point to be studied and flexible methods like Neural Networks and Genetic Algorithms will should be used to generate better rules in an optimal way.

### 3.1. Rules for Generating Model Sets

Suppose that fitness values $f_1, \cdots, f_N$ are obtained for each speaker by applying its test utterance to all the models $M_1, \cdots, M_N$. In our case these fitness values are in the range $[0, 100]$. The rules for generating model sets for this speaker are as in Fig. 1. In this partitioning, the model set $T_1$ is the set of *highly probable members*.

### 3.2. Assignment of Weights of Evidence

The next step is the assignment of the *weight of evidence* values to the sets generated by the simple rules illustarted in Fig. 1. Each weight of evidence should be a non-negative number. For

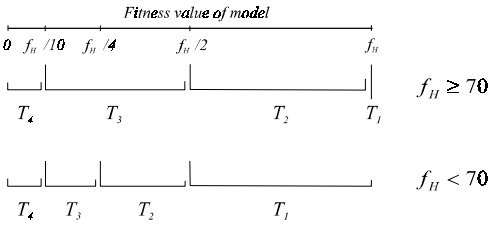


Figure 1. Rules for generating model sets.

our case, the weights of evidence are assigned as

$$w(T_i) = \log\left(\sum_{j\in T_i} \frac{f_j}{\#\text{ of elements in } T_i}\right), \quad i = 1,2,3. \tag{1}$$

for $T_i$ being non-empty. For empty $T_i$, this value is set to zero. The aim here is to use these weight of evidence values to obtain the *degree of support* given to each set $T_1, \cdots, T_4$. We will assign the degree of support given to set $T_4$ as zero since this is the set containing the models with the minimum fitness value. The word *support* here is used in the sense of [1], and the separable support functions mentioned are calculated according to the theory developed therein. These support functions represent the amount of support for the test utterance to belong to one of the models given in a certain set. As a result of the above procedure, a course division of all models will be obtained. However, even this coarse division may not be very reliable. Indeed together to all of the above mentioned sets, a positive support is also assigned to the universal set that contains all models. The value of this support function represents the *amount of ignorance* about the decision made[1]. I.e., the degree to which the experiment prefers not to make a decision at all. If this support has a value of 1 (which is the largest possible value) then this indicates that the experiment gives no information at all about the discrimination problem. Therefore, no model is favored.

The procedure described above is applied on the two selection strategies discussed, For the fist case, the fitness values of the test utterances from the 17 speakers which are given in Table 2 are used without any modification. For the second case, the table is first normalized to give the values in Table 3, which are then used for a decision.

Now, consider the first selection strategy. The decision sets are generated by making use of the rules we have defined. Note that an inherently present set is the universal set which is the union of all the decision sets. By making use of the formulation given in [1], first the weights of evidence are computed for these decision sets. The weight of evidence of the universal set is set to $\infty$ since it contains all possible sets and hence its truth is established. Then the theory enables us to compute the *degrees of support* associated with each decision set, starting from the weights of evidence.

The results obtained for some illustrative cases are given in Table 4. The example test speakers are $Sp_1, Sp_3, Sp_{13}, Sp_{17}$. From Table 1 and Table 2, it is clear that speaker 1 can be easily identified with high reliability. Therefore it constitutes one extreme case for this example group. Speaker $Sp_{13}$ is observed to be impossible to identify. This speaker constitutes the other extreme case given in the tables. The other two speakers represent the cases in between these two extremes.

One observes from Table 2 that $M_3$ has the highest fitness value ($f = 62$) to $Sp_3$ but this can not be claimed to be very reliable since $M_{11}$ (with $f = 53$) and $M_{15}$ (with $f = 56$) also fit to $Sp_3$ with comparable fitness values.

Speaker $Sp_{17}$ corresponds to such a case that $M_{17}$ does not have the highest fitness to $Sp_{17}$ but this fitness value is nevertheless quite close to highest fitness value $f_H = 41$ in the row.

| | Classification of Models | Weight of Evidence ($w$) | Degree of Support ($s$) |
|---|---|---|---|
| $Sp_1$ | $T_1 = \{M1\}$ | 4.37 | 0.81 |
| | $T_2 = \phi$ | 0 | 0 |
| | $T_3 = \{M_2, M_4, M_6, M_8, M_{10}, M_{11}, M_{12}, M_{14}, M_{15}, M_{16}, M_{17}\}$ | 2.91 | 0.18 |
| | $T_4 = \{M_3, M_5, M_7, M_9, M_{13}\}$ | 0 | 0 |
| | $T_5 = \theta$ | $\infty$ | 0.1 |
| $Sp_3$ | $T_1 = \{M_3, M_{11}, M_{12}, M_{14}, M_{15}\}$ | 3.9 | 0.62 |
| | $T_2 = \{M_8, M_{16}\}$ | 3.14 | 0.28 |
| | $T_3 = \{M_7\}$ | 2.08 | 0.09 |
| | $T_4 = \{M_1, M_2, M_4, M_5, M_6, M_9, M_{10}, M_{13}, M_{17}\}$ | 0 | 0 |
| | $T_5 = \theta$ | $\infty$ | 0.013 |
| $Sp_{13}$ | $T_1 = \{M_1, M_2, M_{11}, M_{14}, M_{16}\}$ | 3.56 | 0.62 |
| | $T_2 = \{M_3, M_{10}, M_{12}, M_{13}, M_{15}\}$ | 2.72 | 0.25 |
| | $T_3 = \{M_8, M_9, M_{17}\}$ | 1.99 | 0.11 |
| | $T_4 = \{M_4, M_5, M_6, M_7\}$ | 0 | 0 |
| | $T_5 = \theta$ | $\infty$ | 0.018 |
| $Sp_{17}$ | $T_1 = \{M_1, M_2, M_{10}, M_{11}, M_{14}, M_{16}, M_{17}\}$ | 3.42 | 0.56 |
| | $T_2 = \{M_3, M_{12}, M_{15}\}$ | 2.75 | 0.28 |
| | $T_3 = \{M_4, M_6, M_8\}$ | 2.12 | 0.14 |
| | $T_4 = \{M_5, M_7, M_9, M_{13}\}$ | 0 | 0 |
| | $T_5 = \theta$ | $\infty$ | 0.019 |

Table 4. Classification of models using the first decision strategy and the corresponding weights of evidence and degrees of support.

The following observations can be made on the example speakers in Table 4.

- $Sp_1$: As expected, $Sp_1$ is the only element in the set $T_1$ with a very high degree of support given by $s = 0.81$.
- $Sp_3$: This speaker is in the highest rank set $T_1$, However it is not the only element of this set and the degree of support given to the statement that "*one of the elements of* $T_1$ *will be the correct model*" is $s = 0.62$ (not very high).
- $Sp_{17}$: This speaker shows similarities to $Sp_3$, but the degree of support given to $T_1$ is only $s = 0.56$.
- $Sp_{13}$: The speaker $Sp_{13}$ is not even in the set $T_1$ but it appears in $T_2$. The degree of support given to $T_2$ by this experiment is only $s = 0.25$.

Now consider the same methodology applied to the second decision strategy by considering the normalized fitness values given in Table 3. Table 5 summarizes the results of this strategy in the same form as in Table 4.

Similar type of observations can be made about the results with the second decision strategy by observing Table 5. As expected, for speaker $Sp_{13}$, the set $T_1'$ contains only $M_{13}$ with a degree of support $s = 0.69$.

In the remaining part of this work, the results of the theory given in [1] are applied on the results of the two different decision strategies to *combine the evidences* supplied by these different approaches. This is realized by first generating new decision sets as the intersection of the existing sets $T_i$'s and $T_j$'s and then computing the combination of the degrees of support for each of the new intersection sets.

By using the theory presented in [1], the results given in Tables 6 and 7 are obtained. In these tables the set $T_{ij}$ is the intersection of the sets $T_i$ and $T_j'$ for $i, j = 1, \cdots, 5$. Note that we have $T_{i5} = T_i$ since it is the intersection of the set $T_i$ with the universal set. Similarly we have $T_{5j} = T_j'$. The degree of support for each of these intersection sets $T_{ij}$ is calculated as $s_{ij} = \frac{s_i \cdot s_j'}{\Delta}$, where we have $\Delta$ given by the expression $\Delta = \sum_{T_{ij}\neq\phi} s_i \cdot s_j$. Some promising observations can be made on these two final tables.

(i) At all cases considered, the set that has the highest degree of support contains the correct speaker.

(ii) Only speakers $Sp_1$ and $Sp_{13}$ are identified as unique elements of the sets with the highest degree of support.

| | Classification of Models | Weight of Evidence ($w$) | Degree of Support ($s$) |
|---|---|---|---|
| $Sp_1$ | $T_1' = \{M_1, M_2, M_{17}\}$ | 1.91 | 0.66 |
| | $T_2' = \{M_4, M_6, M_{10}, M_{16}\}$ | 1.0 | 0.20 |
| | $T_3' = \{M_8, M_{11}, M_{12}, M_{13}, M_{15}\}$ | 0.25 | 0.03 |
| | $T_4' = \{M_3, M_5, M_7, M_9, M_{14}\}$ | 0 | 0 |
| | $T_5' = \theta$ | $\infty$ | 0.114 |
| $Sp_3$ | $T_1' = \{M_3, M_7, M_8, M_{11}, M_{12}, M_{15}\}$ | 1.62 | 0.62 |
| | $T_2' = \{M_{14}\}$ | 0.80 | 0.19 |
| | $T_3' = \{M_5, M_9, M_{10}, M_{16}\}$ | 0.21 | 0.035 |
| | $T_4' = \{M_1, M_2, M_4, M_6, M_{13}, M_{17}\}$ | 0 | 0 |
| | $T_5' = \theta$ | $\infty$ | 0.154 |
| $Sp_{13}$ | $T_1' = \{M_{13}\}$ | 2.97 | 0.69 |
| | $T_2' = \{M_2\}$ | 1.83 | 0.20 |
| | $T_3' = \{M_1, M_{10}, M_{11}, M_{14}, M_{16}\}$ | 1.06 | 0.07 |
| | $T_4' = \{M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{12}, M_{15}, M_{17}\}$ | 0 | 0 |
| | $T_5' = \theta$ | $\infty$ | 0.0376 |
| $Sp_{17}$ | $T_1' = \{M_2, M_{17}\}$ | 1.77 | 0.61 |
| | $T_2' = \{M_1, M_6, M_{10}, M_{11}, M_{14}, M_{16}\}$ | 1.06 | 0.24 |
| | $T_3' = \{M_3, M_4, M_8, M_{12}, M_{15}\}$ | 0.22 | 0.03 |
| | $T_4' = \{M_5, M_7, M_9, M_{13}\}$ | 0 | 0 |
| | $T_5' = \theta$ | $\infty$ | 0.125 |

Table 5. Classification of models using the second decision strategy and the corresponding weights of evidence and degrees of support.

(iii) The test utterance of Speaker $Sp_3$ is identified to belong to one of the models $M_3, M_{11}, M_{12}$ and $M_{15}$, with a degree of support $s = 0.43$. However, the combined evidences do not differentiate the models within this set.

(iv) Similarly, the test utterance of Speaker $Sp_{17}$ is identified to belong either to model $M_{17}$ or $M_2$ with a degree of support of $s = 0.52$. Again, the present evidences of the two algorithms does not give any clue to make a decision between these two models.

(v) All the models which have zero support can be eliminated from the set of possible models for a given test utterance.

## 4. CONCLUSIONS

In this work, the mathematical theory of evidence is applied to the problem of closed set text-independent speaker identification, in order to combine the evidences given by different decision strategies on the output of a given algorithm. The theory is also applicable to the cases where the results of entirely different algorithms have to be combined to generate an improvement over their individual performances. Although this work is only at its very early stages, the results presented here are encouraging. New experiments (new identification algorithms) may give fresh information, i.e., evidence for the identification of an unknown utterance. These evidences can be combined with the previous results easily. The procedure may also be used to refine the course decisions previously made. For example, the theory and the decision strategies can be applied to the previously obtained set $\{M_3, M_{11}, M_{12}, M_{15}\}$ to identify Speaker $Sp_3$ uniquely. In the continuation of this work we will exploit all the ideas discussed here.

## REFERENCES


[1] G. Shafer, *A Mathematical Theory of Evidence,* Princeton Univ. Press., Princeton, 1976.

[2] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics,* Vol. 38, 1967, pp. 205-247.

[3] A. P. Dempster, "A generalization of Bayesian inference," *Journal of the Royal Statistical Society, Series B,* Vol. 30, 1968, pp. 205-247.


| Speaker $Sp_1$ | | Speaker $Sp_3$ | |
|---|---|---|---|
| Non-empty Subsets $T_{ij}$ | Support $s_{ij}$ | Non-empty Subsets $T_{ij}$ | Support $s_{ij}$ |
| $T_{11} = \{M_1\}$ | 0.8041 | $T_{11} = \{M_3, M_{11}, M_{12}, M_{15}\}$ | 0.4264 |
| $T_{15} = T_1$ | 0.1402 | $T_{12} = \{M_{14}\}$ | 0.1289 |
| $T_{33} = \{M_{13}\}$ | 0.0089 | $T_{15} = T_1$ | 0.1053 |
| $T_{34} = \{M_3, M_5, M_7, M_9\}$ | 0 | $T_{21} = \{M_8\}$ | 0.1915 |
| $T_{35} = T_3$ | 0.0310 | $T_{23} = \{M_{16}\}$ | 0.0108 |
| $T_{41} = \{M_2, M_{17}\}$ | 0 | $T_{25} = T_2$ | 0.0173 |
| $T_{42} = \{M_4, M_6, M_{10}, M_{16}\}$ | 0 | $T_{31} = \{M_7\}$ | 0.0609 |
| $T_{43} = \{M_8, M_{11}, M_{13}, M_{15}\}$ | 0 | $T_{35} = T_3$ | 0.015 |
| $T_{44} = \{M_{14}\}$ | 0 | $T_{43} = \{M_5, M_9, M_{10}\}$ | 0 |
| $T_{45} = T_4$ | 0 | $T_{44} = \{M_1, M_2, M_4, M_6, M_{13}, M_{17}\}$ | 0 |
| $T_{51} = T_1'$ | 0.0103 | $T_{45} = T_4$ | 0 |
| $T_{52} = T_2'$ | 0.0031 | $T_{51} = T_1'$ | 0.0087 |
| $T_{53} = T_3'$ | 0.0005 | $T_{52} = T_2'$ | 0.0026 |
| $T_{54} = T_4'$ | 0 | $T_{53} = T_3'$ | 0.0005 |
| $T_{55} = \theta$ | 0.0018 | $T_{54} = T_4'$ | 0 |
| | | $T_{55} = \theta$ | 0.0021 |

Table 6. Results of combining the results of two strategies by the Dempster's Rule of Combination, Part A: Speakers $Sp_1$ and $Sp_3$.

| Speaker $Sp_{13}$ | | Speaker $Sp_{17}$ | |
|---|---|---|---|
| Non-empty Subsets $T_{ij}$ | Support $s_{ij}$ | Non-empty Subsets $T_{ij}$ | Support $s_{ij}$ |
| $T_{12} = \{M_2\}$ | 0.2965 | $T_{11} = \{M_2, M_{17}\}$ | 0.5171 |
| $T_{13} = \{M_1, M_{11}, M_{14}, M_{16}\}$ | 0.1072 | $T_{12} = \{M_1, M_{11}, M_{14}, M_{16}\}$ | 0.2008 |
| $T_{15} = T_1$ | 0.0565 | $T_{15} = T_1$ | 0.1059 |
| $T_{21} = \{M_{13}\}$ | 0.4208 | $T_{23} = \{M_3, M_{12}, M_{15}\}$ | 0.0131 |
| $T_{23} = \{M_{10}\}$ | 0.0433 | $T_{25} = T_2$ | 0.0523 |
| $T_{24} = \{M_3, M_{12}, M_{15}\}$ | 0 | $T_{32} = \{M_6\}$ | 0.0496 |
| $T_{25} = T_2$ | 0.0228 | $T_{33} = \{M_4, M_8\}$ | 0.0065 |
| $T_{34} = \{M_8, M_9, M_{17}\}$ | 0 | $T_{35} = T_3$ | 0.0261 |
| $T_{35} = T_3$ | 0.0102 | $T_{44} = \{M_5, M_7, M_9, M_{13}\}$ | 0 |
| $T_{44} = \{M_4, M_5, M_6, M_7\}$ | 0 | $T_{51} = T_1'$ | 0.0174 |
| $T_{51} = T_1'$ | 0.0296 | $T_{52} = T_2'$ | 0.0068 |
| $T_{52} = T_2'$ | 0.0084 | $T_{53} = T_3'$ | 0.0009 |
| $T_{53} = T_3'$ | 0.0030 | $T_{54} = T_4'$ | 0 |
| $T_{54} = T_4'$ | 0 | $T_{55} = \theta$ | 0.0036 |
| $T_{55} = \theta$ | 0.0016 | | |

Table 7. Results of com