# APPLICATION OF FUZZY-INTEGRATION-BASED MULTIPLE-INFORMATION AGGREGATION IN AUTOMATIC SPEECH RECOGNITION

Shuangyu Chang
*Berkeley Initiative in Soft Computing,*
*EECS Dept., University of California, Berkeley*
*Berkeley, CA 94720, USA*
*shawnc@eecs.berkeley.edu*

Steven Greenberg
*International Computer Science Institute*
*1947 Center St. Suite 600*
*Berkeley, CA 94704, USA*
*steveng@icsi.berkeley.edu*

**ABSTRACT**

Many real-world problems can be cast into a multiple-information aggregation framework where preliminary evaluations of separate information sources are combined to produce more accurate and reliable evaluation than would otherwise be the case. In this paper we describe a syllable-proximity evaluation problem in automatic speech recognition that fits well into this aggregation framework. A fuzzy-integration-based approach is adopted as the aggregation operator and a gradient-based algorithm is described for learning parameters automatically from training data. Experiments using spontaneous speech material demonstrate that the fuzzy-integration-based aggregation approach has many advantages over other techniques in terms of both performance and interpretability of the system.

## 1. INTRODUCTION

Many real-world problems such as pattern recognition and decision-making, involve input information from several different sources; the evidence from each source alone can only provide a partial account of all available information. A decision based on information from a single source may be sub-optimal or incorrect, and it is often the convergence of evidence from various sources that provides an accurate and reliable result. Thus, the appropriate aggregation of information from different sources contributes crucially to the success of a solution to such problems. For certain applications it may be possible to create a comprehensive model involving all sources of information with respect to the overall classification or decision. However, for many other tasks a comprehensive model can be difficult to develop; it is often more practical to first perform an evaluation based on each information stream and then combine the results into a single overall result. This latter approach is often referred to as multiple-information aggregation.

Besides requiring less complex models, the multiple-information aggregation approach has several additional advantages over the single, comprehensive-mode approach in practical applications. Performing evaluation individually on each information source minimizes the potentially harmful impact of background noise derived from various sources. Moreover, separating the initial evaluation from the aggregation process provides for flexibility in using a different computing strategy for each information source; certain methods, such as pooling of subjective evaluations, possess an inherent structure similar to the multiple-information aggregation framework and thus may benefit from such an approach. Because of requirements associated with computation and interpretation, as well as the inherent uncertainty, the general class of multi-criteria decision-making and multi-attribute optimization problems lends itself well to soft-computing based techniques. Besides the fuzzy-integration-based aggregation described in this paper, many other soft computing approaches have been successfully applied in aggregation and decision support systems (e.g., using a combination of fuzzy logic and genetic algorithms [13]).

In this paper we describe a syllable-proximity evaluation task in the context of an automatic speech recognition application (ASR) that is well suited to a multiple-information aggregation framework. The syllable-proximity evaluation is one component of a novel multi-tier model of speech recognition [2]. The objective is to evaluate a score for the degree of proximity between an input syllable and a reference syllable based on partial proximity scores obtained individually along each of several articulatory-feature (AF) dimensions. As described in the following section, multiple-information aggregation tasks, such as syllable-proximity evaluation, require appropriate aggregation operators that take into account the relative importance of various information sources, as well as the synergy and redundancy of information contained in subsets of these sources. In this paper, we adopt a fuzzy-integration-based aggregation operator [5][7] that has such a capability and whose parameters have a meaningful interpretation. We also describe how the parameters of the fuzzy integration can be automatically derived from training data by recasting the syllable-proximity evaluation into a classification problem.

The rest of the paper is organized as follows. Section 2 provides a brief background describing the multiple-information aggregation approach and the syllable-proximity evaluation problem. Section 3 introduces the fuzzy-integration-based aggregation method and Section 4 describes an algorithm for learning fuzzy measures from data. Section 5 and 6 describe an experiment pertaining to syllable-proximity evaluation using the fuzzy-integration-based aggregation. Section 7 provides a brief summary of the conclusions.

## 2. MULTIPLE-INFORMATION AGGREGATION AND SYLLABLE-PROXIMITY EVALUATION

Multiple-information aggregation refers to the combining of results from the processing of multiple information sources to form a single overall end-result. Many important real-world problems can be formulated within such an aggregation framework (e.g., multi-criteria decision-making and multi-attribute classification). A common property of such problems is that a preliminary, partial decision or classification score can be obtained separately for each of the information sources. It is often difficult (or at least undesirable) to model directly the relationship between the raw inputs associated with all of the information sources and the final, combined evaluation result. For example, each partial score may be obtained from an unknown procedure, as would be the case for fusion of multiple subjective evaluations. This procedure would also allow each partial score to be determined in a manner that is both simpler and more robust than could be obtained through direct computation of the final score.

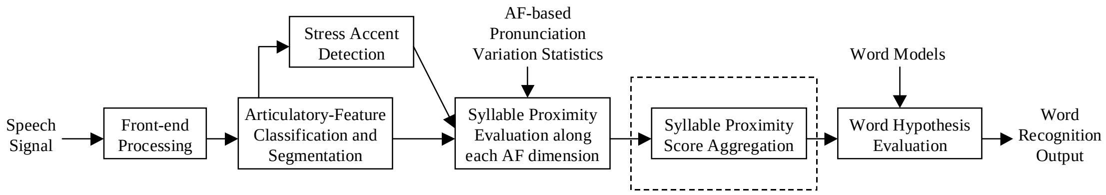


Figure 1. Major components of the syllable-based, multi-tier model of automatic speech recognition [2]. The present paper focuses on the step enclosed in the dashed box.

Conventional models of ASR (at least for English and many other Indo-European languages) assume that words are readily decomposable into constituent phonetic components ("phonemes"). A detailed evaluation of state-of-the-art speech recognition systems indicates that the conventional phonemic "beads-on-a-string" approach is of limited utility, particularly with respect to informal, conversational material [8]. In [2], a syllable-based, multi-tier model of ASR is introduced that explicitly relates articulatory features, syllable structure and stress accent to lexical representation. Articulatory features are the basic building blocks of the phonetic tier of speech; they are used to describe the abstract configuration of the vocal tract during speaking (e.g., the manner and place of articulation, etc). It was shown in [2] that the realization of articulatory features is systematically related to syllable position and stress accent level, and together, they provide a means to accurately and parsimoniously model pronunciation variation in spontaneous speech. Because the multi-tier model used in the recognition application (cf. Figure 1) entails many processing steps that lie outside of the scope of the present paper, we focus here on the computation of syllable proximity associated with a pair of input and reference syllable segments (cf. [2] for further details about the multi-tier model).

Between each pair of input and reference syllable segments a partial syllable-proximity score is separately computed along each of the AF dimensions (i.e., manner of articulation, place of articulation, voicing, lip-rounding, etc.). The computation of each partial syllable-proximity score associated with each AF dimension involves a series of neural-network-based AF classification and segmentation processes, as well as application of probabilistic models pertaining to pronunciation variation (cf. Figure 1 and [2]). Within the context of the multi-tier model, distributing this computation among various AF dimensions has several distinct advantages relative to conventional approaches, including (1) simplicity and robustness of computation, as well as (2) the capability of accurately encapsulating the sort of pronunciation variation that characterizes spontaneous speech. The partial syllable-proximity scores have to be combined appropriately in order to form a unitary syllable-proximity score capable of being used in the subsequent word-hypothesis evaluation component of the recognition process. For these reasons we can formulate the syllable-proximity evaluation task as a multiple-information aggregation problem. In other words, for each pair of input and reference syllable segments, given a set of partial syllable-proximity scores along $N$ articulatory-feature dimensions, $H = (h(x_1),\cdots,h(x_N))$, we are interested in deriving an appropriate combining operator that yields the desired combined score $y = f(H)$.

There exist a variety of aggregation operators that can be used for $f(\,)$, such as the arithmetic or geometric mean, linearly weighted average, ordered weighted average, neural networks, and fuzzy-logic-based systems [18]. Each operator possesses certain properties. Selecting an appropriate aggregation operator depends crucially on the characteristics of the particular application. With respect to the syllable-proximity evaluation, different AF dimensions contribute in varying measure to the computation of syllable proximity and therefore should be assigned different weights in the combining process. Moreover, the partial scores associated with various AF dimensions are not orthogonal, and there exists significant coupling among them. Thus, a simple linear combination of the partial scores is incapable of capturing the redundancy and synergy of the information associated with various subsets of the AF dimensions; a highly non-linear process is required in its place. Because the relationship among the partial scores associated with different AF dimensions can be quite complex, the aggregation operator should be flexible in its ability to extract information from heterogeneous sources without prior specification of their inter-relationship. On the other hand, good interpretability of the aggregation method is desirable, especially for diagnostic experiments where we would like to ascertain the relative importance and interaction of various AF dimensions to the combined decision. Interpretability of parameters would also be of help in the selection of the most appropriate AF dimensions to use. As described in the following section, a fuzzy-integration-based aggregation operator possesses a number of properties ideally suited to this task.

## 3. FUZZY INTEGRATION AS AN AGGREGATION OPERATOR

Over the past several years a fuzzy-integration-based approach has become increasingly popular for multiple-information aggregation [10]. There have been a number of successful applications of this approach in multi-criteria decision-making (e.g., [7][11]), multi-attribute classification and pattern recognition (e.g., [5][6][9][14]). This approach combines decision or classification scores from multiple information sources into a single composite score by applying a fuzzy integral with respect to a designated fuzzy measure, representing differential weighting of scores derived from a variety of information sources. In this section we briefly review the concepts of fuzzy measure and fuzzy integral, and then discuss the advantages of the approach for multiple-information aggregation with specific application to syllable-proximity-score aggregation.

The concept of fuzzy measure was originally introduced by Sugeno [17] in the early 1970's in order to extend the classical (probability) measure through relaxation of the additivity property. A formal definition of the fuzzy measure is as follows:

Definition 1. Fuzzy measure: Let $X$ be a non-empty finite set and $\Omega$ a Boolean algebra (i.e. a family of subsets of $X$ closed under union and complementation, including the empty set) defined on $X$. A fuzzy measure, $g$, is a set function $g{:}\Omega \rightarrow [0,1]$ defined on $\Omega$, which satisfies the following properties: (1) Boundary conditions: $g(\phi) = 0$, $g(X) = 1$. (2) Monotonicity: If $A \subseteq B$, then $g(A) \leq g(B)$. (3) Continuity: If $F_n \in \Omega$ for $1 \leq n < \infty$ and the sequence $\{F_n\}$ is monotonic (in the sense of inclusion), then $lim_{n\rightarrow\infty}\, g(F_n) = g(lim_{n\rightarrow\infty}\, F_n)$. And $(X, \Omega, g)$ is said to be a fuzzy measure space.

This definition of a fuzzy measure differs from that of a probability measure only in terms of the monotonicity property. Because additivity is a special case of monotonicity, the probability measure is, in effect, a special case of a fuzzy measure.

For the syllable-proximity evaluation let $X = \{x_1,\cdots,x_N\}$ represent the set of $N$ articulatory feature dimensions under consideration and $g$, a fuzzy measure, represent the contribution of each subset of $X$ (i.e. a set of some AF dimensions, including singleton sets) in evaluating the proximity between a reference syllable and the input. In many situations it is useful to ascertain the contribution of a particular AF dimension in the entire evaluation process. However, since each AF dimension is included in many subsets of $X$, the contribution of a particular AF dimension cannot be easily discerned from the fuzzy measure. A concept from cooperative game theory, the Shapley score [16][5], can be applied within this context to help in the interpretation.

Definition 2. Shapley score: Let *g* be a fuzzy measure on *X*. The Shapley score for every $x_i \in X$ is defined by

$$v_i \equiv \sum_{K \subset X \setminus \{x_i\}} \frac{(|X|-|K|-1)!\,|K|!}{|X|!}[g(K \cup \{i\}) - g(K)]$$

where $|X|$ and $|K|$ are the cardinality of *X* and *K*, respectively.

A Shapley score, $v_i$, can be interpreted as an average value of the contribution that an information source, $x_i$ alone, provides for all different combinations of information sources and it can be verified that Shapley scores sum (by necessity) to one. This concept has also been extended to computing the interaction of a pair of information sources [12], as well as interaction of any subset of *X* [4].

Definition 3. Two-way interaction index: Let *g* be a fuzzy measure on *X*. The two-way interaction index of elements $x_i, x_j \in X$ is defined by

$$I_{ij} \equiv \sum_{K \subset X \setminus \{x_i, x_j\}} \frac{(|X|-|K|-2)!\,|K|!}{(|X|!-1)}[g(K \cup \{x_i, x_j\}) - g(K \cup \{x_j\}) - g(K \cup \{x_i\}) + g(K)]$$

where $|X|$ and $|K|$ are the cardinality of *X* and *K*, respectively.

The interaction index, $I_{ij}$, provides an indication of the interaction between the pair of information sources $x_i$ and $x_j$. When $I_{ij} < 0$, there exists a negative interaction (redundancy) between information sources, $x_i$ and $x_j$, in that the value of the pair $x_i$ and $x_j$ is less than the sum of the values of $x_i$ alone and $x_j$ alone when they are incorporated into sets of information sources. On the other hand, if $I_{ij} > 0$, there exists a positive interaction (synergy) between $x_i$ and $x_j$. In cases where $I_{ij} = 0$, there is no interaction between the pair.

To combine scores obtained from various information sources with respect to a specific fuzzy measure a technique based on the concept of the fuzzy integral can be applied. There are actually several forms of fuzzy integral [5]; the one adopted here is the Choquet integral proposed by Murofushi and Sugeno [10].

Definition 4 (Choquet) Fuzzy integral: Let $(X, \Omega, g)$ be a fuzzy measure space, with $X = \{x_1, \cdots, x_N\}$. Let $h: X \to [0,1]$ be a measurable function. Assume without loss of generality that $0 \le h(x_1) \le \cdots \le h(x_N) \le 1$, and $A_i = \{x_i, x_{i+1}, \cdots, x_N\}$. The Choquet integral of *h* with respect to the fuzzy measure *g* is defined by:

$$\int_C h \circ g = \sum_{i=1}^{N} [h(x_i) - h(x_{i-1})]\, g(A_i) \text{ where } h(x_0) = 0.$$

An interesting property of the (Choquet) fuzzy integral is that if *g* is a probability measure, the fuzzy integral is equivalent to the classical Lebesgue integral and simply computes the expectation of *h* with respect to *g* in the usual probability framework. The fuzzy integral is a form of averaging operator in the sense that the value of a fuzzy integral is between the minimum and maximum values of the *h* function to be integrated. A number of commonly used aggregation operators are special cases of the fuzzy integral [6][7], e.g., the min and max operators, the weighted sum and the ordered weighted average. A distinct advantage of the fuzzy integral as a weighted operator is that, using an appropriate fuzzy measure, the weights represent not only the importance of individual information sources but also the interactions (redundancy and synergy) among any subset of the sources.

## 4. LEARNING FUZZY MEASURES

The review of fuzzy measure and fuzzy integral above describes how they can be used to combine scores associated with various AF dimensions into an aggregated proximity score. The intuitive interpretation of a fuzzy measure allows for the specification of the fuzzy measure based on expert knowledge. However, in practice it is more useful to be able to learn the fuzzy measure directly from data and to deduce the contribution patterns automatically. In our application, during the training phase the system has knowledge of which reference syllable in a pool of reference syllables best matches the input syllable. We may then recast the syllable-proximity evaluation into a classification problem and apply a supervised, discriminative training scheme for learning the fuzzy measure from the data. That is, for each input syllable, we would like to classify it as belonging to one of the reference syllable classes by selecting the reference syllable with the highest fuzzy integral score. Intuitively, the fuzzy-measure parameters that perform well on the classification problem are very likely to serve as the desired solution for syllable-proximity evaluation.

The classification problem can be formally described as follows. Suppose there are $M$ reference syllable classes, $N$ articulatory feature dimensions under consideration and $D$ data points (input syllables) in the training set. Let $X = \{ x_1, \cdots, x_N \}$ denote the set of $N$ articulatory-feature dimensions (each as a separate information source). For the $d$'th input syllable data point, let $H_d^m = \{ h_d^m(x_i), i = 1, \cdots, N \}$ be the set of the proximity scores provided by the AF dimensions for the $m$'th reference syllable, and $y_d^m = f^m(H_d^m)$ be the combined score for model syllable class $m$, where $f^m()$ is the overall combining function for reference syllable class $m$ (with its parameters). A well-known result in pattern recognition is that the minimum expected classification error is obtained if one always selects the winning class according to the maximum of the posterior probabilities of the classes given the input. In [15], Richard and Lippmann showed that a discriminant function trained with one-of-$M$ target outputs (i.e., one for the correct class and zeros for the others) approximates the posterior probabilities of the classes using either a minimum-squared-error (MSE) or a minimum-cross-entropy (MCE) error criterion. In the current implementation, the target outputs for the combining function adopt a one-of-$M$ scheme. The total errors for the MCE and MSE criteria are

$E^{MCE} = \sum_{d=1}^{D} E_d^{MCE} = \sum_{d=1}^{D} \sum_{m=1}^{M} t_d^m \log \frac{t_d^m}{y_d^m}$ and $E^{MSE} = \sum_{d=1}^{D} E_d^{MSE} = \sum_{d=1}^{D} \sum_{m=1}^{M} (y_d^m - t_d^m)$, respectively.

Furthermore, we let the overall combining function $f^m()$ take the following form:

$f^m(H_d^m) = \frac{\exp(\beta C^m(H_d^m))}{\sum_{j=1}^{M} \exp(\beta C^j(H_d^j))}$ where $C^j(H_d^j) = \int_C H_d^j \circ g^j$ is the (Choquet) fuzzy integral of the AF dimension proximity scores for model syllable $j$ with respect to the fuzzy measure $g^j$. The $\beta$ is a scaling factor and the exponential softmax form serves two functions: (1) to provide a normalization of the fuzzy integral outputs so that they are proper probability terms, (2) to adjust the sharpness of the estimated posterior probability distribution where a large $\beta$ represents a narrow distribution concentrating on the maximum class output and a small $\beta$ is associated with a broad distribution. In the current implementation system performance is relatively insensitive to the choice of the scaling factor, $\beta$, over a large numerical range. $\beta$ was set to 10 using trial-and-error methods for the experiments described in the following section. Note that the softmax transformation is used only for learning the fuzzy measures and it is actually the $C^j(H^j)$ terms that are used during recognition. In short, the learning problem essentially reduces to finding an optimal fuzzy measure, $g$, minimizing an error criterion $E$ given the input proximity scores $H$ and the desired target outputs $t$ (in a one-of-$M$ representation).

The optimization problem can be solved in several different ways, using linear and quadratic programming [6], gradient-based methods [3], as well as genetic-algorithm-based optimization methods. The current system adopts an iterative, gradient-based framework [3] for efficiency and scalability. The learning algorithm is similar in spirit to the one proposed by Grabisch and Nicholas [6][3], with some differences in the detailed update equations which reflect the different normalization and formulation of error criteria. Some major steps of the algorithm are given below, and details of the algorithm as well as the derivation of update equations can be found in [2]:

- Step 0 – initialization: each of the fuzzy measures $g^m$ for $m = 1, \cdots, M$ is initialized at the so-called equilibrium state [3], i.e., $g^m(\{x_i\}) = 1/N$ for all $i = 1, \cdots, N$ and $g^m$ is additive (i.e., $g^m(A \cup B) = g^m(A) + g^m(B)$ for all $A, B \subseteq X$ and $A \cap B = \Phi$).
- Step 1 – computing the fuzzy integral: for an input-output learning data pair ($H = \{ H^m, m = 1, \cdots, M \}, t = \{ t^m, m = 1, \cdots, M \}$), evaluate the fuzzy integrals and the softmax normalization to obtain $y^m = f^m(H^m)$ for $m = 1, \cdots, M$.
- Step 2 – updating fuzzy measures: for each $m = 1, \cdots, M$, do the following:
  - Step 2.1: compute the output error $\varepsilon^m = y^m - t^m$.
  - Step 2.2: let $g_{(0)}^m, g_{(1)}^m, \cdots, g_{(N)}^m$ denote the ordered fuzzy measures involved in the fuzzy integral evaluation of $y^m$. The order is determined by the input proximity scores such that

$h^m(x_{(N)}) \le h^m(x_{(N-1)}) \le \cdots \le h^m(x_{(1)})$. Then, for each $i = 1, \cdots, N-1$, compute $\delta_{(i)} = h^m(x_{(N+1-i)}) - h^m(x_{(N-i)})$ and update each $g^m_{(i)}$ by $g^{m^{new}}_{(i)} = g^{m^{old}}_{(i)} - \alpha\beta\varepsilon^m\delta_{(i)}$ for the MCE error criterion and by $g^{m^{new}}_{(i)} = g^{m^{old}}_{(i)} - 2\alpha\beta y^m(\varepsilon^m - \sum_{q=1}^{M}\varepsilon^q y^q)\delta_{(i)}$ for the MSE criterion. The $\alpha$ in both equations is a learning rate parameter.

- Step 2.3: verify the monotonicity to ensure that each $g^m$ is still a proper fuzzy measure.

For each training epoch, Steps 1 and 2 were repeated for all training data points. Several training epochs were performed and a separate cross-validation data set used to determine the total number of epochs to avoid over-training. In [3], an extra step is described that smoothes the fuzzy measures that have never been updated due to scarcity of training data. However, this additional step was unnecessary in the current application for two reasons: (1) our task requires the sharing of fuzzy measures among various model syllable classes, which significantly reduces the total number of parameters from $M \cdot 2^N$ to $2^N$, and (2) a sufficiently large amount of training data is available for the application.

## 5. EXPERIMENTS

Experiments have been carried out on speech material from the OGI Numbers95 corpus [1] consisting of digits and numbers extracted from spontaneous American English telephone interactions. Due to the limited scope of the current paper we limit the description here to the training and testing experiments for syllable-proximity score aggregation. The other components of the recognition system are described in detail in [2]. The training data consisted of 3,233 utterances (12,510 words, 15,306 syllables) from a training set and 357 utterances (1,349 words, 1,650 syllables) from a separate cross-validation set. Testing was performed on 1,206 utterances (4,669 words, 5,878 syllables). In these experiments we considered seven separate AF dimensions: manner of articulation, place of articulation, voicing, vocalic height, lip-rounding, spectral dynamics (mainly to distinguish between monophthongs and diphthongs), and vocalic tenseness (most closely related to the intrinsic length of a vowel segment). For each input syllable there were 25 reference syllables evaluated for degrees of proximity. As described in the previous section, we recast the training of the aggregation operator into a classification problem. That is, based on the proximity scores supplied by the seven AF dimensions, we sought to classify each input syllable as one of 25 reference syllable classes. A fuzzy-integration-based aggregation operator was trained using the gradient-based method described in Section 4 using the training data. Classification performance using either the MCE or MSE criterion is similar; we report here the result based on the MSE criterion. For comparison, we also performed the same classification experiments using several other aggregation operators: (1) “MAX” – the maximum operator; (2) “MIN” – the minimum; (3) “MEAN” – the arithmetic mean; (4) “WT-AVG” – linearly weighted average; (5) “MLP” – multi-layer-perceptron neural network.

## 6. RESULTS AND DISCUSSION

Performance of the aggregation methods is shown in Table 1 for the syllable-classification task. The results are given in terms of the percentage of correctly classified syllables. The weights for the linearly weighted average (WT-AVG) were trained using a ridge regression technique. The MLP contained a single hidden layer of 100 units and was trained using back-propagation. The results show that by taking into account the interaction among subsets of information sources, the fuzzy-integration-based technique amply outperforms simpler methods such as the arithmetic mean and linearly weighted average. On the other hand, the fuzzy-integration-based technique performed at a level close to the MLP-based classifier, which is quite capable of capturing complicated non-linear relationships among various inputs.

| Method | MAX | MIN | MEAN | WT-AVG | MLP | FUZZY-INT |
|---|---|---|---|---|---|---|
| Accuracy % | 33.21 | 74.77 | 83.70 | 84.59 | 90.56 | 90.17 |

Table 1. Syllable classification accuracy for several aggregation methods.

However, there is a singular advantage in using the fuzzy-integration-based method rather than MLPs, namely the interpretability of learned parameters. In many practical applications, it is quite useful to understand the significance of learned parameters associated with an aggregation operator. For example, knowing how each information source contributes to the combined result can facilitate feature selection in a principled fashion and helps in determining the required precision of input features. And in many cases the ability to interpret learned parameters generates new insights into the task domain such as structural patterns of information sources and heuristic rules governing score combination. Such insights can be very helpful in assessing the reliability of a classification system and adapting it to new situations. For the current task it is difficult to interpret the significance of parameters associated with the MLP classifier despite having provided good classification performance. On the other hand, the learned parameters of the fuzzy-integration-based aggregation operator can be easily interpreted, particularly with the aid of the Shapley score and interaction indexes. Figure 2 shows the mean Shapley scores derived from learned fuzzy measures (averaged over 15 random trials, along with the range of ±1 standard deviation). It is observed that both manner and place of articulation have above-average Shapley scores, while lip-rounding and voicing have well below-average Shapley scores. This pattern suggests that both manner and place dimensions contribute significantly to the aggregated result, while the lip-rounding and voicing dimensions contribute relatively little. This result is consistent with intuition derived from phonetic and linguistic knowledge, as the manner and place dimensions are expected *a priori* to provide the greatest amount of information pertaining to lexical discrimination.

In order to gain further insights into the interaction patterns associated with various AF dimensions in the recognition process the interaction indices (cf. Section 3) were computed for pairs of AF dimensions from the trained fuzzy measures, as shown in Figure 3. For each pair of AF dimensions the size of each square (below the minor diagonal) corresponds to the interaction index between the two AF dimensions indicated by the labels on the horizontal and vertical axes. A light-colored square indicates positive interaction (i.e., synergy) while dark-colored squares reflect negative interaction (i.e., redundancy). For example, the large positive interaction between voicing and manner suggests these two dimensions, considered together, contribute more to recognition than when considered separately. In tandem with the Shapley scores the interaction indices can be used to estimate the utility associated with each AF dimension. For example, the "lip-rounding" dimension has a relatively small Shapley score and mostly small or negative interactions with other AF dimensions, and thus may be removed from consideration without significant impact on classification performance. On the other hand, although the voicing dimension has a low Shapley score by itself, it may still be of significant utility because of the large positive interaction between voicing and manner (as well as place of articulation).

The insights that we have gained from the interpretation of learned fuzzy measures agree with our intuition that different AF dimensions assume different levels of importance with respect to speech recognition. Moreover, there exists a significant amount of interaction among different sets of AF dimensions. Understanding the nature of such interaction is very important for constructing appropriate models of speech recognition and for improving the performance of ASR systems. The fuzzy-integration-based multiple-information aggregation framework should also be helpful in many other applications that require effective and transparent combining of heterogeneous information sources. It would also be interesting to explore other approaches to learn and refine fuzzy measures such as using a genetic-algorithm or simulated-annealing-based global search.

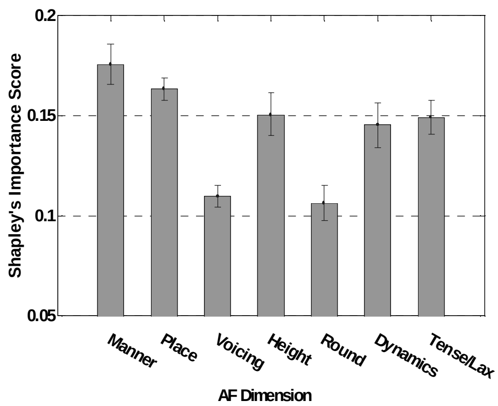


Figure 2. Shapley scores of seven AF dimensions derived from the trained fuzzy measure for syllable-proximity score aggregation. Error-bars indicate the range of ±1 standard deviation. Note that the Shapley scores sum to one.

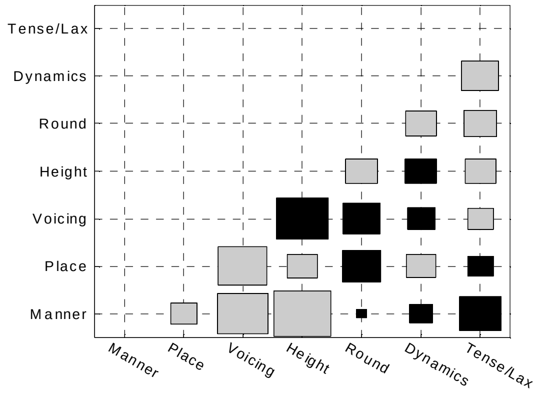


Figure 3. Two-way interaction indices between pairs of AF dimensions derived from the trained fuzzy measures for syllable-proximity score aggregation. Sizes of the squares below the minor diagonal indicate the magnitude and are normalized to the size of the grid. Dark-colored squares indicate a negative interaction and light-colored squares signify a positive interaction between a pair of AF dimensions.

## 7. SUMMARY AND CONCLUSIONS

This paper has described the general framework of a fuzzy-integration-based approach to multiple-information aggregation, a commonly encountered problem in a wide range of real-world applications. In particular, we have focused on the practical problem of syllable-proximity evaluation for automatic speech recognition and have illustrated how it fits into a multiple-information aggregation framework. We have shown via experiments and analyses that fuzzy measures and a fuzzy integral possess many advantages relative to other techniques for aggregating partial results from multiple information sources, in terms of both performance and interpretability. We have also shown how fuzzy-measure parameters can be learned automatically using a gradient-based algorithm by recasting the syllable-proximity evaluation into a classification problem. The learned fuzzy measures have provided useful insights into patterns of interaction among articulatory-acoustic feature dimensions.


## ACKNOWLEDGEMENTS

This work was conducted at the International Computer Science Institute, Berkeley. The authors would like to thank Dr. Lokendra Shastri and Dr. Masoud Nikravesh for helpful discussions. This research was supported in part by the U.S. Department of Defense and the National Science Foundation. The first author has also been supported by grants from British Telecom and the U.S. Office of Naval Research.


## REFERENCES


[1] Center for Spoken Language Understanding, Oregon Graduate Inst. (1995) *Numbers corpus, R.1.0.*

[2] S. Chang (2002) *A Syllable, Articulatory-Feature, and Stress-Accent Model of Speech Recognition*, Ph.D. Dissertation, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. Available at URL: *http://www.icsi.berkeley.edu/~shawnc*.

[3] M. Grabisch (1995) A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Proc. of 4th IEEE Int. Conf. on Fuzzy Systems*, pp.145-50, Yokohama, Japan.

[4] M. Grabisch (1996) K-order additive fuzzy measures. In *Proc. of 6th intl. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pp.1345-50, Spain.

[5] M. Grabisch (2000) Fuzzy integral for classification and feature extraction. In M. Grabisch, et al., eds., *Fuzzy Measures and Integrals:Theory and Applications*, pp.415-34. Physica-Verlag, NY.

[6] M. Grabisch and J. Nicolas (1994) Classification by fuzzy integral: Performance and tests. *Fuzzy Sets and Systems*, (65): 225-71.

[7] M. Grabisch and M. Roubens (2000) Application of the Choquet integral in multicriteria decision making. In M. Grabisch, et al., eds., *Fuzzy Measures and Integrals: Theory and Applications*, pp.348-73. Physica-Verlag, NY.

[8] S. Greenberg and S. Chang (2000) Linguistic dissection of switchboard-corpus automatic speech recognition systems, *Proc. of ISCA Workshop on ASR:Challenges for the New Millennium,* pp.195-202. Paris.

[9] J. Keller, P. Gader, and A. Hocaoglu (2000) Fuzzy integrals in image processing and recognition. In M. Grabisch, et al., eds., *Fuzzy Measures and Integrals: Theory and Applications*, pp.435-66. Physica-Verlag, NY.

[10] T. Murofushi and M. Sugeno (1989) An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy Sets and Systems*, (29): 202-27.

[11] T. Murofushi and M. Sugeno (2000) The Choquet integral in multiattribute decision making. In M. Grabisch, et al., eds., *Fuzzy Measures and Integrals: Theory and Applications*, pp.333-47. Physica-Verlag, NY.

[12] T. Murofushi and S. Soneda (1993) Techniques for reading fuzzy measures (III): interaction index. In *Proceedings of the 9th Fuzzy System Symposium*, pp.693-96, Sapporo, Japan.

[13] M. Nikravesh and B. Azvine (2002) Fuzzy queries, search, and decision support system. *Soft Computing*, (6): pp.273-99.

[14] T. Pham and H. Yan (1999) Combination of handwritten-numeral classifiers with fuzzy integral. In C. Leondes, ed., *Fuzzy Theory Systems:Techniques and Applications*, pp.1111-27. Academic.

[15] M. Richard and R. Lippmann (1991) Neural network classifiers estimate Bayesian a posteori probabilities. *Neural Computation*, (3): pp461-83.

[16] L. Shapley (1953) A value for n-person games. In H. Kuhn and A. Tucker, eds., *Contributions to the Theory of Games, no. 28 in Annals of Mathematics Studies,* pp.307-17. Princeton Univ. Press.

[17] M. Sugeno (1974) *Theory of fuzzy integrals and its applications*. Ph.D. Dissertation, Tokyo Institute of Technology.

[18] L. Zadeh (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man and Cybernetics*, (1): pp.28-44.