Fuzzy information transmission analysis for continuous speech features

Dirk J. J. Oosthuizen and Johan J. Hanekom[a)]
Department of Electrical, Electronic and Computer Engineering, University of Pretoria, University Road, Pretoria 0002, South Africa

(Received 19 September 2014; revised 18 February 2015; accepted 24 February 2015)

Feature information transmission analysis (FITA) estimates information transmitted by an acoustic feature by assigning tokens to categories according to the feature under investigation and comparing within-category to between-category confusions. FITA was initially developed for categorical features (e.g., voicing) for which the category assignments arise from the feature definition. When used with continuous features (e.g., formants), it may happen that pairs of tokens in different categories are more similar than pairs of tokens in the same category. The estimated transmitted information may be sensitive to category boundary location and the selected number of categories. This paper proposes a fuzzy approach to FITA that provides a smoother transition between categories and compares its sensitivity to grouping parameters with that of the traditional approach. The fuzzy FITA was found to be sufficiently robust to boundary location to allow automation of category boundary selection. Traditional and fuzzy FITA were found to be sensitive to the number of categories. This is inherent to the mechanism of isolating a feature by dividing tokens into categories, so that transmitted information values calculated using different numbers of categories should not be compared. Four categories are recommended for continuous features when twelve tokens are used. © *2015 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4916198]

[JL] Pages: 1983–1994

I. INTRODUCTION

Quantitative estimates of the amount of information transmitted by individual acoustic features are useful for the investigation of speech perception and the development of hearing prostheses. For example, with quantitative techniques, it is possible to estimate the degree to which information about important features are preserved by a hearing prosthesis. Shannon (1948) developed a mathematical model with which the information transmitted over a noisy communication channel can be calculated based on error probabilities. This model was employed by Miller and Nicely (1955) to develop a technique to estimate the information transmitted in a typical closed set listening experiment. This was achieved by collecting the results from such an experiment in a confusion matrix and estimating the pairwise error probabilities from this confusion matrix using the frequency approach to probability.

Miller and Nicely also developed a technique to isolate an individual acoustic feature and estimate the amount of information transmitted by that feature. This was achieved by assigning the tokens presented in an experiment to different categories according to the feature of interest. For example, to estimate the information transmitted by the place of articulation feature in a consonant identification experiment, three categories could be defined to contain the front, middle, and back consonants, respectively. In this way, the token confusion matrix is converted to a smaller category confusion matrix showing only within-category and between-category confusions. This category confusion matrix is then used to estimate the pairwise error probabilities for the computation of transmitted information. This technique is referred to as feature information transmission analysis (FITA). The present study proposes an expansion to the FITA technique.

Originally, the FITA technique was used to characterize categorical features of speech, e.g., voicing, nasality, affrication, and place of articulation. For these features, the assignment of tokens to categories is simple because it is usually dictated by the definition of the feature. For example, every consonant is either voiced or unvoiced.

As the FITA technique became a standard approach to the estimation of feature information, researchers started using it to characterize continuous features, such as formant frequencies (Blamey *et al.*, 1989; Van Wieringen and Wouters, 1999). Unlike a categorical feature, a continuous feature can assume an infinite number of different values within the typical range. If tokens are grouped according to a continuous feature, some tokens will be closer to the edge of the category and some will be closer to the center. Distances between tokens in the same category are, however, ignored by the FITA technique and distances between tokens close to, but on different sides of, a category boundary are exaggerated (more detail is provided in Sec. II A). These effects limit the resolution with which a FITA can estimate transmitted information and may cause the technique to be sensitive to the boundary locations.

No standard procedure or guidelines are available for the assignment of tokens to categories when performing a

[a)]Author to whom correspondence should be addressed. Electronic mail: johan.hanekom@up.ac.za

FITA and researchers differ in their approach to this task. This is seen clearly in Figs. 1(a)–1(c), where different studies employed the FITA technique using recorded vowels from the same study (Hillenbrand *et al*., 1995). Both the boundary locations and the number of categories per feature differ substantially between these three studies. Another study performed a FITA on the consonants recorded by Hillenbrand *et al*. but did not perform a FITA on the vowels because they could not determine reasonable ways to group the stimuli (Apoux and Healy, 2012).

From the category assignments found in literature, although seldom explicitly stated, it appears researchers consider three factors when assigning tokens to categories. First, some researchers seem to avoid splitting natural clusters formed by tokens (Donaldson *et al*., 2011). Second, the distribution of tokens is considered by assigning approximately the same number of tokens to each category (Van Wieringen and Wouters, 1999). Finally, some consider the theoretical origin of features, such as the relationship between formants and vowel height and backness (Yoon *et al*., 2012).

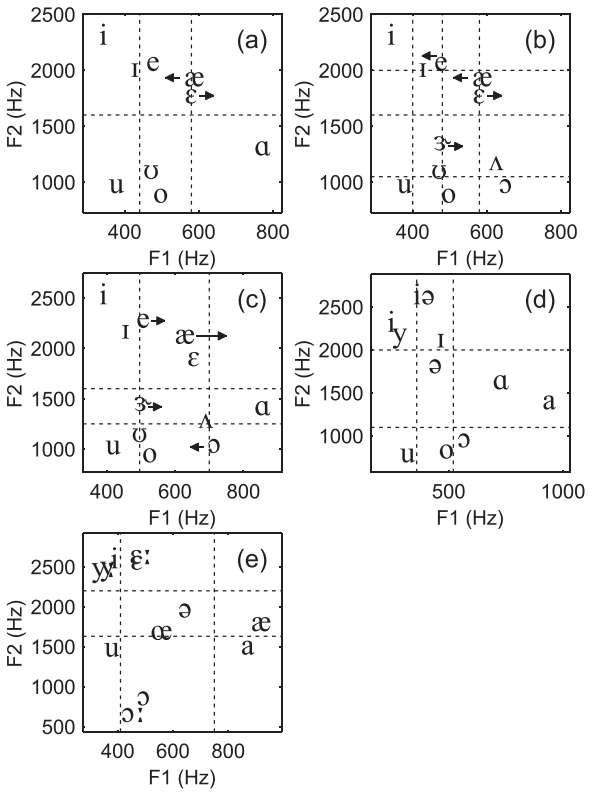

FIG. 1. Examples of category assignments from different studies. Panels (a) to (c) show how different researchers divided the English vowels recorded by Hillenbrand *et al*. (1995) into categories. Donaldson *et al*. (2011) [(a)] and Sheffield and Zeng (2012) [(b)] used only recordings from male speakers, whereas Yoon *et al*. (2012) [(c)] used both male and female speakers. These three studies do not report the actual boundary locations, but the locations were inferred from their reported category assignments and the vowel spaces reported by Hillenbrand *et al*. Arrows clarify the assignments of vowels near category boundaries. Where studies used only a subset of the vowels recorded by Hillenbrand *et al*., only the subset is shown in the relevant panel. Instances where a boundary appears to pass through a vowel location or where a vowel appears to have been assigned to a category that does not correspond to its location may be due to the fact that individual speakers were used in the studies the FITAs were performed in, whereas Hillenbrand *et al*. report vowel spaces as averaged across speakers of the same gender. Panel (d) shows the vowel space for the Dutch vowels recorded by Van Wieringen and Wouters (1999) and their own category assignments. Panel (e) shows the vowel space for the Afrikaans vowels recorded for the purposes of the present study and the manual category assignments used.

The factors that direct the choice of category boundaries may often be in conflict. For example, in Fig. 1(e), an approach based on natural clusters formed along the F2 feature dimension would result in the /ɑ/ and /u/ tokens being included in the middle category. However, interpreting the F2 frequency as a consequence of the place of articulation suggests the category assignments indicated on the figure. These categories also agree with other literature on Afrikaans vowels (Taylor and Uys, 1988; Van der Merwe *et al*., 1993; Botha, 1996).

The process of manual category assignment introduces a human effect the magnitude of which has not yet been determined. This human effect may enter either as systematic bias or as random noise, both of which are unwanted. For example, when comparing the information transmitted by different features, a bias may exist in favor of features that fall naturally into distinct categories if the effect is systematic. If the effect is random, it can be averaged out by collecting enough data, but this would require the collection of more data than what would have been necessary in the absence of the effect.

If the process of token grouping could be automated in a thoughtful manner, this automation may be advantageous. The human effect is removed, which should increase the comparability of FITA outputs between studies by different authors. Also this would relieve researchers of the time and effort involved in the selection of boundaries, especially in cases where no obvious divisions exist between tokens.

To achieve a method for automated grouping, the work presented here investigated the sensitivity of the FITA technique to the grouping parameters[1] and an adaptation to the technique, which reduces this sensitivity by incorporating the theory of fuzzy sets (Zadeh, 1965), is suggested.[2] The concept of fuzzy sets and how this applies to FITA should become clear in Sec. II B.

II. THEORY

A. The crisp FITA

The traditional FITA technique developed by Miller and Nicely (1955) will be referred to as the crisp FITA[3] in this paper to distinguish it from the fuzzy FITA. The difference between these two techniques is that the first has sharp boundaries between categories, whereas the second has gentle transitions between categories as will be explained in the next subsection. The mathematical representation of the FITA technique in the following text uses matrix notation to demonstrate the effect of the grouping parameters and to develop the fuzzy FITA as a general case of the crisp FITA.

The transmitted information T over a discrete memoryless channel can be expressed as (Shannon, 1948; Miller and Nicely, 1955)

$$T(x;y) = -\sum_{x,y} p_{xy} \log\left(\frac{p_x p_y}{p_{xy}}\right), \quad (1)$$

where p_x is the *a priori* probability at the source of producing a certain message x from a closed set of possible messages, p_y is the probability at the destination of identifying any received signal as a certain message y, and p_{xy} is the joint probability of producing message x and identifying it as message y. The transmitted information can be divided by the entropy $H(x)$ of the source to compute the relative information transmitted,

$$T_{rel}(x;y) = \frac{T(x;y)}{H(x)} = \frac{-\sum_{x,y} p_{xy} \log\left(\frac{p_x p_y}{p_{xy}}\right)}{-\sum_{x} p_x \log(p_x)}. \quad (2)$$

The relative information transmitted is a normalized measure of the covariance of the output with the input (Miller and Nicely, 1955). A value of zero is obtained if the source and the destination are statistically independent ($p_{xy} = p_x p_y$), indicating that no information about the source has reached the destination. A value of one is obtained if all messages are received correctly ($p_{xy} = p_x$ if $x = y$ and 0 otherwise), indicating that all information available at the source has reached the destination.

Equations (1) and (2) can be used in the context of a closed set phoneme identification experiment by setting p_x to the *a priori* probability of presenting token x, p_y to the probability of observing token y, and p_{xy} to the joint probability of observing token y when token x is being presented. These probabilities are unknown but can be estimated from the results of the experiment. Experiment results are typically represented in a confusion matrix C, such that each entry c_{ij} represents the number of times token i was presented and recognized as token j. Estimating the probabilities from the confusion matrix yields the following estimates for the transmitted information measures:

$$\hat{T}(x;y) = T_M(C), \quad (3)$$

$$\widehat{T_{rel}}(x;y) = \frac{T_M(C)}{H_M(C)}, \quad (4)$$

where T_M and H_M are functions that operate on matrices and return scalars. These are defined as

$$T_M(M) = -\sum_{i,j}\left[\left(\frac{m_{ij}}{N_M}\right)\log\left(\frac{m_i m_j}{N_M m_{ij}}\right)\right], \quad (5)$$

$$H_M(M) = -\sum_{i}\left[\left(\frac{m_i}{N_M}\right)\log\left(\frac{m_i}{N_M}\right)\right], \quad (6)$$

$$m_i = \sum_j m_{ij},$$

$$m_j = \sum_i m_{ij},$$

$$N_M = \sum_{i,j} m_{ij} = \sum_i m_i = \sum_j m_j.$$

m_i is the number of times token i was presented (the sum of the entries in row i in matrix M), m_j is the number of times token j was observed (the sum of the entries in column j in matrix M), and N_M is the total number of stimuli (the sum of all entries in matrix M). Normalization by N_M is used to estimate probabilities according to the frequency approach to probability.

The transmitted information measures in Eqs. (3) and (4) represent the combined information available in all features transmitted. If the probabilities are calculated for a scenario where only one feature is transmitted, $T(x;y)$ and $T_{rel}(x;y)$ represent the information transmitted about that particular feature. A feature is typically isolated by grouping the tokens according to the feature being evaluated, so that tokens that are similar in that feature fall into the same category, whereas tokens that differ substantially in that feature fall into different categories (see Fig. 1). A category confusion matrix A is then computed, so that each entry a_{ij} represents the number of times a token in category i was presented and recognized as a token in category j. The information transmitted about this feature is then estimated as follows:

$$\hat{T}(x;y) = T_M(A), \quad (7)$$

$$\widehat{T_{rel}}(x;y) = \frac{T_M(A)}{H_M(A)}. \quad (8)$$

We can define a category membership matrix G such that each entry G_{ij} is equal to one if token i is in category j and zero otherwise. The category confusion matrix A can then be calculated from the token confusion matrix C as follows:

$$A = G^T C G. \quad (9)$$

The transmitted information measures for a specific feature can now be estimated as a function of the token confusion matrix C and the category membership matrix G as follows:

$$\hat{T}(x;y) = T_M(G^T C G), \quad (10)$$

$$\widehat{T_{rel}}(x;y) = \frac{T_M(G^T C G)}{H_M(G^T C G)} = \frac{T_M(G^T C G)}{T_M(G^T G)}. \quad (11)$$

Equations (10) and (11) are mathematically equivalent to the FITA described by Miller and Nicely (1955). The source entropy $H_M(G^T C G)$ simplifies to the expression $T_M(G^T G)$ if all tokens were presented the same number of times (the row sums of C are equal). This indicates that the entropy (information content) of the source is characterized completely by the number of categories and the number of tokens in each category. This may be expected because the category membership matrix is the only variable that changes as different features are selected. Equation (11) can also be used to illustrate mathematically that in the case where all tokens are identified correctly, the relative information transmitted is equal to one (C becomes an identity matrix).

From a mathematical perspective, the crisp FITA presents three problems. First, discontinuities exist near category boundaries; this may cause the technique to be sensitive to the boundary locations. Consider what happens when one

boundary is modified while all feature values and all other boundaries remain constant. If the boundary is shifted in such a way that no tokens are reassigned to new categories because of the shift, the estimated information transmitted does not change. However, if a token is reassigned to a different category, this estimate changes. It may happen that a large shift in boundary position may cause no tokens to be reassigned, whereas a small shift may cause one or more tokens to be reassigned.

The second and third problems concern the resolution and accuracy with which token feature values are represented. In a similar scenario as in the preceding text, if all boundaries remain constant and the feature value of one token is changed, this change will only be detected by the FITA if the token transcends a boundary. Small changes would therefore not be detected; this equates to a low resolution representation. This low resolution has a direct effect on the accuracy of the representation of distances between tokens. The distance between two closely spaced tokens that are on opposite sides of a boundary is represented in the same way as the distance between any two tokens in different categories, whereas the distance between any two tokens in the same category is represented as zero. For example, if the boundaries in Fig. 1(c) were used to estimate the information transmitted by the F1 feature using a crisp FITA, confusions between /ɪ/ and /e/ would be treated as between-category confusions, whereas confusions between /e/ and /ɛ/ would be treated as within-category confusions in spite of the fact that the former pair is more similar in F1 than the latter pair.

The fuzzy FITA discussed in Sec. II B aims to address all three of these problems. Note that none of these problems exist for categorical features for which the FITA technique was originally invented (Miller and Nicely, 1955).

B. The fuzzy FITA

The theory of fuzzy sets provides a framework for addressing problems where imprecision arises due to the absence of sharply defined criteria for category membership (Zadeh, 1965). An often quoted example problem in the literature of fuzzy sets is the classification of humans by height into two categories: Short and tall. Because the boundary between short and tall is not clearly defined, some humans will be difficult to classify. Furthermore, if a large, representative sample is used, it is inevitable that two humans on opposite sides of, but near the boundary height, will be more similar in height than two humans near different extremes within one category. This is analogous to the problem of classifying speech tokens by their formant frequencies.

The discontinuities at token category boundaries may be removed by using fuzzy categories.[4] A fuzzy category has no sharp, clearly defined boundary and may contain elements with a partial degree of membership. This partial degree of membership is denoted by a real number that ranges from zero (meaning no membership) to one (meaning complete membership). A membership function is usually defined to map the input space of a variable to a degree of membership for a particular fuzzy category. For example, if tokens were grouped into fuzzy categories according to their first formant frequency values, each category will be characterized by a membership function that takes a frequency value as input and returns a degree of membership for that category. This way, each token can belong to more than one category with different degrees of membership for each category. Fuzzy categories are a generalization of normal (crisp) categories and reduce to the latter when the membership functions are rectangular functions.

For the proposed fuzzy FITA, adjacent categories like those in the crisp FITA are used, but the membership functions are triangular (Fig. 2). The membership function of a particular category ranges linearly from zero at the center of the previous category to one at the center of the category under consideration to zero at the center of the next category. Exceptions are made for the first and last categories, where the membership functions are set to unity over the sections of the input space that do not overlap with other categories. This has the effect that the membership functions add up to unity at every point in the input space. This is to ensure that a token's degrees of membership of different categories add up to one, and each token therefore has the same amount of influence on the resulting estimate. Unless otherwise specified, equal-sized categories are used.

Mathematically, this fuzzy FITA is a generalization of the FITA described in Sec. II A. Equations (10) and (11) remain the same, but the contents of the category membership matrix G changes. Instead of containing only ones and zeros, g_{ij} now contains the degree of membership of token i to category j. To ensure that each token has the same amount of influence on the resulting estimate, the rows of G are scaled to add to one.

With this new way of calculating G, the two expressions on the right hand side of Eq. (11) are no longer equivalent because the new source entropy $H_M(G^TCG)$ is no longer identical to the new optimal information transmitted $T_M(G^TG)$. This is because the result of Eq. (9) has off-diagonal elements even if matrix C is an identity matrix. To

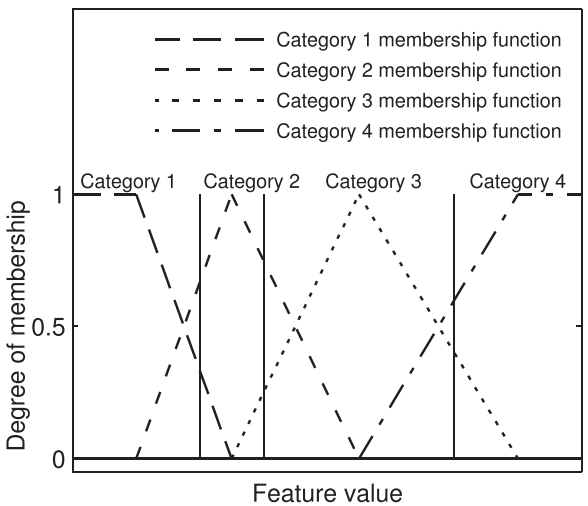

FIG. 2. Example of membership functions for a fuzzy FITA with four categories. Vertical lines indicate category boundaries.

use a fuzzy FITA, one of the preceding expressions (source entropy or optimal information transmitted) must therefore be selected as a norm for calculating the relative information transmitted. The optimal information transmitted is preferred because this yields a measure that ranges from zero to one and is therefore suitable for the comparison of different features. If the source entropy was used as a norm, the resulting estimate would never reach a value of one and would depend on grouping parameters even if the token confusion matrix is an identity matrix. The fuzzy FITA equations then become

$$\hat{T}(x;y) = T_M(G^T CG), \tag{12}$$

$$\widehat{T_{rel}}(x;y) = \frac{T_M(G^T CG)}{T_M(G^T G)}, \tag{13}$$

with G calculated as described in this section and the other parameters defined in Sec. II A.

Mathematically, the fuzzy FITA appears to address the three problems identified earlier. The discontinuities at category boundaries have been removed because if a boundary is shifted slightly to cross over a token, the token's degree of membership for the new category increases by a small amount and its degree of membership for its previous category decreases by a small amount (see Fig. 2). The resolution with which token feature values are represented is also improved because a slight change in one token feature value causes a slight change in its degrees of membership to two categories, causing a slight change in the FITA output. Finally, the accuracy with which inter-token distances are represented is improved, as tokens on opposite sides of the same boundary belong to both categories with similar degrees of membership and tokens that are far apart but in the same category differ considerably in their degrees of membership to that category and the bordering categories.

Although a mathematical analysis is effective at identifying problems in the internal structure of the FITA technique and in adapting the technique to address these problems, the effect of these problems in actual applications and the effect of the proposed adaptation are best investigated experimentally.

III. EXPERIMENT: METHOD

Two experiments were performed to evaluate the two FITA techniques. Because FITAs calculate information content at the perceptual level, their outputs are by definition estimates, and no true values are available to compare the estimates against. Therefore it is not possible to determine which of the two FITA techniques provide the most accurate estimate of the actual information transmitted. Instead the experiments focused on the ability of each technique to report consistent results if the grouping parameters are varied. The two experiments investigated the effects of two grouping parameters. In the first experiment, the number of categories was held constant, and the location of a boundary was varied. In the second experiment, equal-sized and randomly generated categories were used, and the number of categories was varied.

The two FITA techniques were evaluated using data taken from literature as well as data recorded in our lab. The latter consisted of confusion matrices that were measured in a closed set vowel identification test and acoustic features that were extracted from these recordings. The experiments in our lab were performed at four signal-to-noise ratios (SNRs) to investigate the effects of grouping parameters and the ability of the fuzzy FITA to compensate for these effects over a range of listener performance levels. The subsections in the following text provide detail on the collection of data, the implementation of the FITA, and the analysis of results.

A. Data from literature

A crisp FITA requires two inputs, namely, a confusion matrix and a table of category assignments. The category assignments can be calculated from extracted feature values and category boundary locations if both of these are available. A fuzzy FITA requires three inputs, namely, a confusion matrix, extracted feature values, and category boundary locations. Because the boundary locations were manipulated in these experiments, the minimum data required were confusion matrices and extracted feature values. Most studies report only confusion matrices and category assignments (Donaldson *et al.*, 2011; Sheffield and Zeng, 2012; Yoon *et al.*, 2012). However, two studies that report both measured confusion matrices and extracted feature values are Hillenbrand *et al.* (1995) and Van Wieringen and Wouters (1999). In addition, the latter also report category assignments and boundary locations.

Data from Van Wieringen and Wouters were used both in the first experiment, where the boundary location was modified, and the second experiment, where the number of categories was modified. Data from Hillenbrand and coworkers were used in the second experiment only.

B. Stimuli, equipment, and recording procedure

In addition to the data taken from literature, a closed-set vowel identification test was performed. Twelve vowel tokens (/ɑ/, /æ/, /ɛ/, /ɛ:/, /ɔ:/, /y:/, /i/, /ə/, /u/, /ɔ/, /œ/ and /y/) were recorded in a "pVt" context from an Afrikaans-speaking female speaker. The tokens were selected to be representative of the entire set of Afrikaans vowels in terms of the distribution of their F1, F2, and duration values. Tokens were recorded in an Acoustic Systems RE242 double-walled sound booth using a Sennheiser ME62 microphone. The microphone output was sampled at 44.1 kHz by an M-Audio FastTrack Pro external sound card and stored on a computer in .wav format.

The speaker was instructed to speak clearly and produce each utterance at the same pitch. From 45 recorded utterances of each token, the best 15 were selected. Utterances containing glottal fry, aspiration of the start consonant, where the speaker did not speak clearly, or where other sounds (e.g., due to movement of feet) were audible were rejected. Sound files were normalized so that the average power of the vowel part of the waveform was the same for all tokens.

Speech-weighted noise was generated based on the average speech spectrum of the speaker. This was added to

tokens to achieve SNRs from −10 to −13 dB in 1 dB increments. This SNR range was chosen, because pilot experiments revealed that it corresponded to a 30%–70% success rate for vowel identification. This range ensured that enough confusions occurred (needed to gain information about the utilization of acoustic features), while limiting the number of confusions due to random guessing (i.e., not based on acoustic features).

Vowel stimuli were padded at the start and end of each vowel so that each stimulus contained three intervals: A 700 ms noise-only interval followed by a noise-and-vowel interval followed by another 700 ms noise-only interval. This was done so that the listener could adapt to the noise before the vowel was presented. The 700 ms interval was chosen because pilot experiments revealed that listeners concentrated more easily and performed more consistently with a 700 ms interval than with a 400 ms interval. Concentration is important because confusions made due to a lapse in concentration do not characterize the utilization of acoustic features and therefore add noise to the measured confusion matrices.

C. Listeners and listening procedure

Twelve Afrikaans-speaking listeners between the ages of 19 and 23 participated in the listening tests. All listeners had normal hearing [pure tone thresholds smaller than or equal to 20 dB hearing level (HL) at octave frequencies ranging from 250 Hz to 8 kHz].

Listening tests were performed in the same sound booth used for recordings. The same sound card was used to present stimuli through a single KEF Q30 loudspeaker at 65 dB sound pressure level (SPL). Listeners were seated facing the loudspeaker. A user interface, developed in MATLAB, was used to present stimuli, facilitate identification by selecting buttons and store results.

Separate training protocols were developed for initial training, retraining after a short break and retraining after a long break. These training protocols were designed to ensure consistent performance and were tested in pilot experiments. During training, feedback was always given for tokens presented in quiet but never for tokens presented in noise. This was done to ensure that a listener is familiar with the token set and the task but did not use accidental features when listening in noise.

Tests were divided into four sessions, completed on two separate days. During each session, one set was completed at each of the four SNRs. A set consisted of 12 repetitions of the 12 vowels, resulting in a total of 144 presentations. SNRs were counter-balanced between sessions. In total, each listener spent approximately 8 h and listened to 12 presentations of each of 12 vowels at each of four SNRs for each of four sessions, a total of 2304 presentations.

D. Extraction of acoustic features

The transmitted information of three features, namely, the first two formant frequencies (F1 and F2) and vowel segment duration (D), was used to evaluate the two FITA techniques because these three features were all available in the data obtained from literature. For the vowels recorded in Sec. III B, feature values were extracted as follows. The starting and ending times of the vowel segments of each utterance were identified manually based on a visual inspection of the waveform and auditory inspection of different time segments. Vowel duration was extracted from these. Formant frequencies were extracted using the PRAAT software package (Boersma and Weenink, 2001), which uses an LPC-based algorithm. Default algorithm parameters for female speakers were used, including an LPC order of 10, a maximum formant frequency of 5500 Hz and reference frequencies of 550 and 1650 Hz for the first two formants, respectively.

E. Assignment of tokens to categories and FITA calculation

The category definitions for the data collected in our lab are summarized in Fig. 3. For the first experiment, only two categories were used, and the boundary between the two categories was varied in 5% increments between the two most extreme tokens for each feature (arrows on the left hand side of Fig. 3). Using only two categories simplified the experiment by restricting the number of parameters and also allowed a greater range for varying the boundary location. The second experiment was divided into two parts. For the first part, the range between the two most extreme tokens for each feature was divided into equal-sized categories, and the number of categories was increased in unit steps from 2 to 12 (arrows on the right hand side of Fig. 3). For the second part, the number of categories was varied, and for each

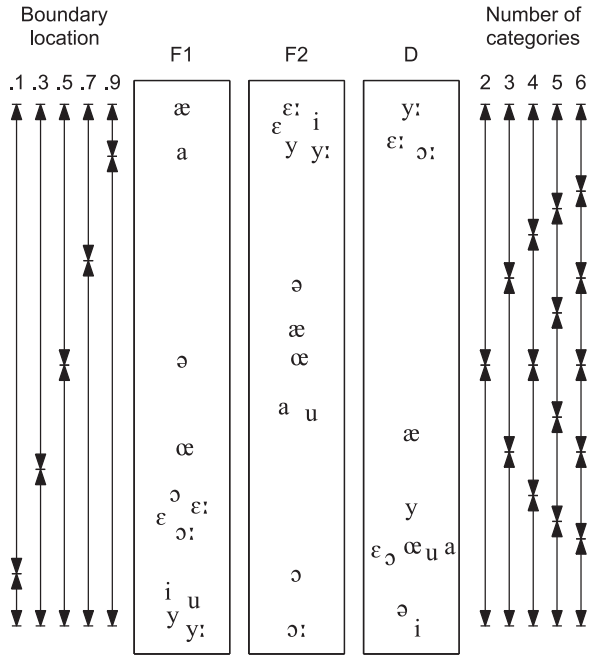

FIG. 3. Category assignments for the two experiments. The arrows on the left hand side indicate the category assignments for the first experiment, where the boundary location was varied. The arrows on the right indicate the category assignments for the first part of the second experiment, where the number of categories was varied. Due to practical limitations, only subsets of the category assignments for each experiment are displayed on the figure. Refer to the text for the complete sets.

number of categories, boundary locations were randomly generated.

Data from Van Wieringen and Wouters (1999) were included in the first the experiment. Here the number of categories was fixed to that used in their study (three categories for formants and two categories for duration). For the duration feature, the boundary location was varied in the same manner as indicated in Fig. 3. For the formant features, one boundary at a time was modified, while the other boundary was fixed at the value used by Van Wieringen and Wouters. The range over which each boundary was varied was the center 80% of the range containing that boundary and was bounded on either side by either a neighboring boundary line or an extreme token (the token that had the highest or lowest value among all tokens for the feature of interest).

Data from both Van Wieringen and Wouters (1999) and Hillenbrand *et al.* (1995) were included in both parts of the second experiment. In the first part, categories were equal-spaced and the number of categories was varied in the same manner as in Fig. 3. In the second part, boundary locations were randomly generated.

The relative information transmitted was calculated for each data set, feature, listener (where applicable), SNR (where applicable), and grouping method. This relative measure was selected because this measure is normalized and therefore more suitable for comparison across different conditions. Equations (11) and (13) were used for the crisp and fuzzy FITA, respectively. Note that these equations are identical except that the category membership matrix G is calculated differently (see Sec. II B). As an example, the category membership matrices for a selected feature and grouping method are shown in Fig. 4.

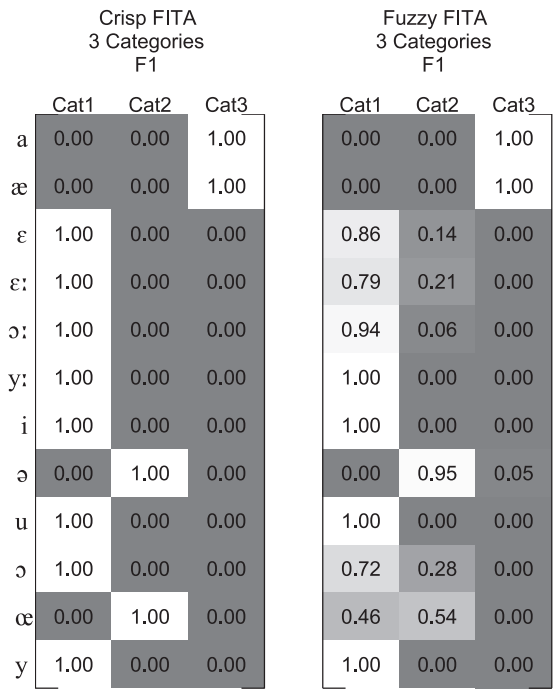

FIG. 4. Category membership matrices for the F1 feature for category assignments that divide the token space into three equal-sized categories. The matrices were calculated as discussed in Sec. II A for the crisp FITA and as discussed in Sec. II B for the fuzzy FITA.

The techniques were then tested for their sensitivity to the two grouping parameters (boundary location and number of categories). The crisp and fuzzy FITA outputs were plotted as a function of each grouping parameter to examine the effect of discontinuities at the crisp FITA boundaries (see Sec. II A) and visually evaluate the sensitivity of the two techniques to the two parameters. Linear regression analyses were performed on the data obtained in our lab with the transmitted information reported by the FITA as output variable and the listener, SNR, and grouping parameter (either boundary location or number of categories) as predictor variables. The regression analyses were used to calculate the portions of the variance in the FITA outputs accounted for by the grouping parameter, the SNR, and the listener, respectively. All predictors were treated as categorical for this purpose. The regression analyses were repeated for each feature and each FITA technique separately.

Finally, a Monte Carlo simulation was performed to measure the variance of each technique when using random boundary locations. The simulation was repeated for each FITA technique, feature, SNR, listener, and number of categories separately. In each case, the required number of boundary locations was generated from a uniform distribution ranging between the two extreme tokens in the feature value of interest. Such a set of boundary locations was accepted if no categories were empty or narrower than one-fifth of the average category width (10% of the feature range for two categories or 2.5% for eight categories). If a set of boundary locations was rejected, all boundaries in the set were regenerated from the same distribution. The process was repeated until 1000 FITAs were performed for each combination of FITA technique, feature, SNR, listener, and number of categories after which the mean and standard deviation of the outputs were computed for each combination separately. This process was also repeated with the data from Van Wieringen and Wouters (1999) and Hillenbrand *et al.* (1995), where the SNR and listener were not varied.

IV. EXPERIMENT: RESULTS

A. The effect of boundary location

The relative information transmitted as reported by the two FITA techniques is shown as a function of the boundary location in Fig. 5.

Discontinuities are evident in the crisp FITA outputs in Fig. 5, as predicted by the theory presented in Sec. II A. The fuzzy FITA has effectively removed these discontinuities as discussed in Sec. II B. In addition to the discontinuities, the crisp FITA outputs appear more sensitive to the boundary location than the fuzzy FITA outputs as judged by the extents of the y values of the traces in Fig. 5.

The crisp estimation of the information transmitted by the duration feature seems to fail close to the left extreme for the data recorded in our lab and close to the right extreme for the data from Van Wieringen and Wouters. This can happen when either short or long vowels are clustered and the boundary line enters the cluster, causing the FITA to interpret confusions between tokens within the cluster as a loss of information. This is not a problem because researchers would

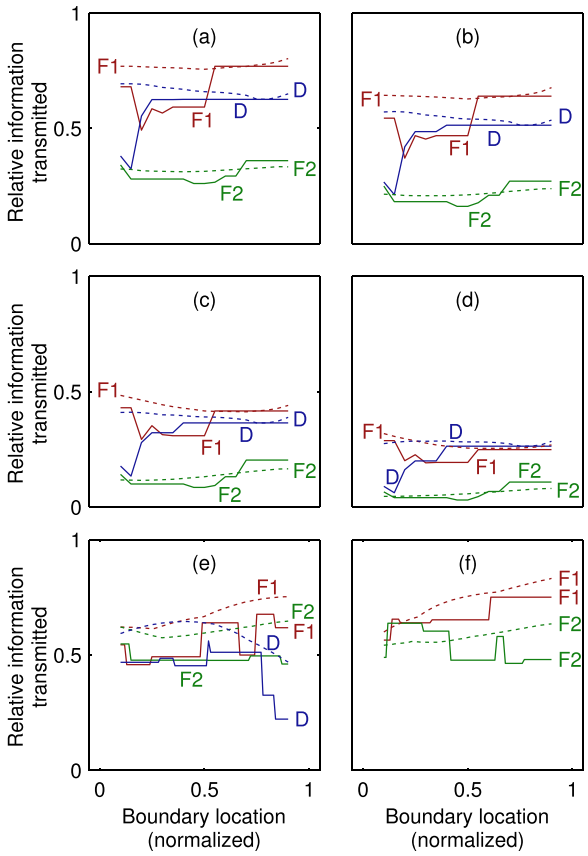

FIG. 5. (Color online) Relative information transmitted as reported by the crisp (solid lines) and fuzzy (dotted lines) FITA techniques as a function of boundary location. Panels (a) to (d) show FITAs performed on data recorded in our lab at four SNRs, decreasing linearly from −10 dB in (a) to −13 dB in (d). Here two categories were used and the boundary location variable was normalized between the extreme values of each feature. Panels (e) and (f) show FITAs performed on the data recorded by van Wieringen and Wouters (1999). Here three categories were used for formants and two categories for duration as was done in their study. The first and second boundaries were varied in (e) and (f), respectively. The duration feature is absent from (f) because it has only one boundary. Each boundary location variable was normalized between either the extreme feature value or the neighboring boundary on either side. Where data from multiple listeners were available, these data were averaged across listeners in the final step before drawing to the graph.

not typically choose such extreme values for the duration feature. It is notable, however, that the fuzzy FITA is shielded from this effect.

The crisp estimates of the information transmitted by the F1 feature for the data recorded in our lab have a large discontinuity near the middle of the range. A crisp FITA was performed with manually selected boundaries to determine which side of this discontinuity is erroneous. Boundaries were selected to reflect the typical distribution of Afrikaans vowels (Taylor and Uys, 1988; Van der Merwe *et al.*, 1993; Botha, 1996) (see Fig. 1). Results agreed with the estimates to the right of the discontinuity, suggesting an error to the left of the discontinuity. Again the fuzzy FITA is shielded from this effect, agreeing with the estimates obtained with the manually selected boundaries.

For the data from the study by Van Wieringen and Wouters, the fuzzy FITA consistently reports higher values for the relative information transmitted than the crisp FITA. Because we do not have access to "true" values, it is not possible to determine which of the techniques are correct here. It is of interest, however, that this difference is consistent across features and boundary locations and therefore does not influence relative comparisons between features. It should also be noted that the fuzzy FITA traces are smoother and have smaller extents of their y values than the crisp FITA traces.

Both FITA techniques seem to report the effect of SNR consistently at any particular choice of boundary location. For the purpose of comparison between features, the ratios between the relative information transmitted as reported for different features are more consistent for the fuzzy FITA than for the crisp FITA. The fuzzy FITA traces resemble the ideal trace (a constant function with a zero slope) more closely than their crisp counterparts.

The portions of the variance in the FITA outputs explained by the listener, SNR, and boundary location as obtained from the linear regression analysis (see Sec. III E) are shown in Table I. Boundary locations below 0.2 and above 0.8 were discarded because even when choosing carelessly, researchers would typically not choose a boundary location outside this range.

Table I shows that this linear regression model with these three categorical predictors is a good model of the relative information transmitted as reported by the two FITA techniques for the three features considered. In all cases, approximately 90% of the total variance was accounted for by the combination of the three predictors. The SNR accounts for the largest portion of the variance; this is reasonable as less information can be transmitted at lower SNRs. A large portion of the variance is accounted for by the listener; this indicates that the listeners chosen for this experiment differ in their ability to use the information transmitted by the three features considered. The boundary location accounts for less than 1% of the variance in the FITA outputs when a fuzzy FITA is used. If this was true in general, a researcher would not need to invest effort in selecting boundary locations. For the crisp FITA, however, the boundary location accounts for approximately 10% of the variance in the FITA outputs for two features. Because this parameter plays a larger role here, more effort is required to choose it correctly. When using a FITA to compare the information transmitted by different features, a fuzzy FITA may also be preferred because the risk of

TABLE I. Variance accounted for by the listener, SNR and boundary location in a linear regression model of the relative information transmitted as estimated by two FITA techniques for three features. The linear regression model was fit to the data presented in the Figs. 5(a)–5(d).

	Crisp FITA			Fuzzy FITA		
Factor	F1	F2	D	F1	F2	D
Listener	23.38	13.75	33.04	24.84	19.62	35.19
SNR	57.95	63.04	59.70	68.52	69.71	58.66
Boundary location	9.54	11.06	1.86	0.13	0.72	0.56
Total	90.67	87.85	94.61	93.49	90.04	94.41

introducing bias due to the effect of the boundary location is smaller for a fuzzy FITA.

B. The effect of the number of categories

The relative information transmitted as reported by the two FITA techniques is shown as a function of the number of categories in Fig. 6.

The difference between the two FITA techniques is less pronounced for this parameter than for the boundary location parameter. In most cases, the traces resemble the shape of a funnel with the open end on the left, so that differences in the transmitted information between features are emphasized (possibly even exaggerated) when using a small number of categories and masked when using a large number of categories. This behavior is probably inherent to the mechanism of isolating a feature by grouping similar tokens according to that feature. When the number of categories is small, the categories themselves are large, and tokens that are large distances apart could end up in the same category. When this happens, some of the information transmitted by the feature under evaluation is not reflected in the FITA outputs. When the number of categories is large, some tokens that are not actually distinguished by the feature under evaluation end up in different categories and some information transmitted by other features is included in the FITA outputs.

This funnel shape is especially prominent in the F2 traces for the data measured in our lab and the fuzzy F1 and D traces for the data measured by Van Wieringen and Wouters (1999). The effect suggests that researchers should refrain from using different numbers of categories for different features when comparing features with one another.

Small oscillations were observed in many of the crisp FITA traces. This may be due to the sensitivity of the crisp FITA to the boundary location parameter and the fact that some boundaries exhibit oscillatory behavior if equal-sized categories are used and the number of categories is varied. For example, a boundary exists in the center of the token space for every second number of categories and a boundary exists at a normalized location of 0.33 for every third number of categories.

The data from Hillenbrand *et al.* (1995) do not provide much information about the transmission of individual features; this is understandable, because this was not part of the aim of their study. The overall percentage of correct responses in their study was 95.4%, which suggests that all features were transmitted effectively. When investigating the transmission of individual features, it is important to control the difficulty of the listening task.

The portions of the variance in the FITA outputs explained by the listener, SNR, and number of categories are shown in Table II. The number of categories was varied from two to eight for this analysis. Considering more categories would not make sense because of the inclusion of unwanted features in the FITA estimates as discussed in the preceding text (see Fig. 6).

Again this linear regression model is a good model of the relative information transmitted as reported by the FITAs, accounting for more than 90% of the variance in the FITA outputs for all combinations of acoustic feature and FITA technique. As before, the SNR accounts for the largest portion of the variance and the listener for another large portion. For both FITA techniques, the variance accounted for by the number of categories is small for the F1 and duration features but larger for the F2 feature. This result reinforces the interpretation that the number of categories has an actual effect on the output for both FITA techniques. Note that the funnel shape in Fig. 6 is most prominent for the F2 feature for the data measured in our lab. If this analysis was repeated for the data measured by Van Wieringen and Wouters

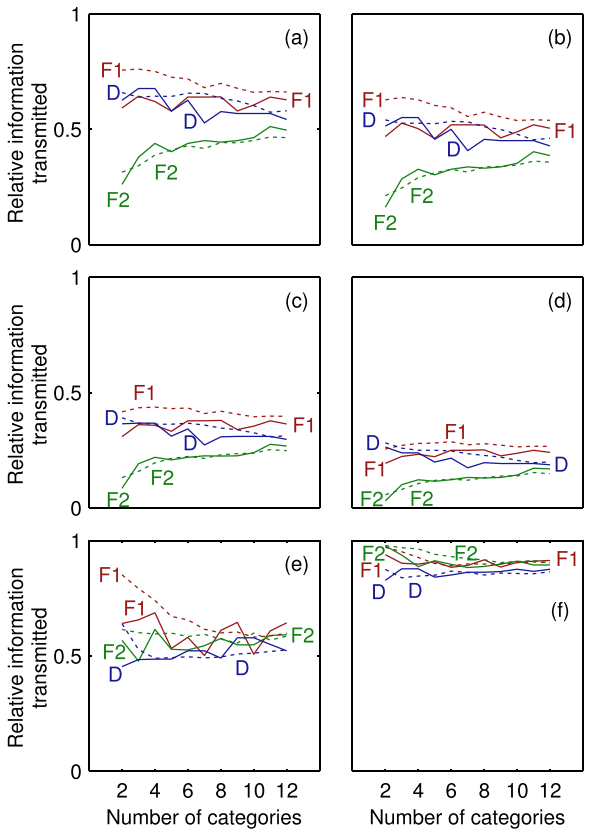

FIG. 6. (Color online) Relative information transmitted as reported by the crisp (solid lines) and fuzzy (dotted lines) FITA techniques as a function of the number of categories. Panels (a) to (d) show FITAs performed on data recorded in our lab at four SNRs, decreasing linearly from −10 dB in (a) to −13 dB in (d). Panels (e) and (f) show FITAs performed on the data recorded by Van Wieringen and Wouters (1999) and Hillenbrand *et al.* (1995), respectively. Where data from multiple listeners were available, these data were averaged across listeners in the final step before drawing to the graph.

TABLE II. Variance accounted for by the listener, SNR, and number of categories in a linear regression model of the information transmitted as estimated by two FITA techniques for three features. The linear regression model was fit to the data presented in Figs. 6(a)–6(d).

	Crisp FITA			Fuzzy FITA		
Factor	F1	F2	D	F1	F2	D
Listener	27.56	17.55	29.19	28.01	15.74	32.59
SNR	65.47	63.55	61.91	65.63	68.62	62.78
Number of categories	1.48	12.98	4.39	0.52	7.96	0.30
Total	94.52	94.07	95.49	94.16	92.32	95.67

(1999), more variance might have been accounted for by the number of categories for the F1 feature.

The means and standard deviations of the relative information transmitted as estimated in the Monte Carlo simulation are shown in Figs. 7 and 8, respectively. As in Fig. 6, the traces in Fig. 7 resemble the shape of a funnel, reinforcing the interpretation that the effect of the number of categories on the FITA output is inherent in the mechanism of feature isolation by means of token grouping.

An important observation from Fig. 8 is that the standard deviation across boundary locations is smaller for the fuzzy FITA than for the crisp FITA in almost all cases, highlighting the effectiveness of the fuzzy grouping strategy in reducing the sensitivity of the FITA technique to manually selected parameters.

V. DISCUSSION

A. General observations and conclusions

The fuzzy FITA technique was designed to reduce the effect of grouping parameters on the transmitted information reported by a FITA. It appears to have succeeded for the first grouping parameter, namely, the boundary location. Although the linear regression model in Sec. IV A explained

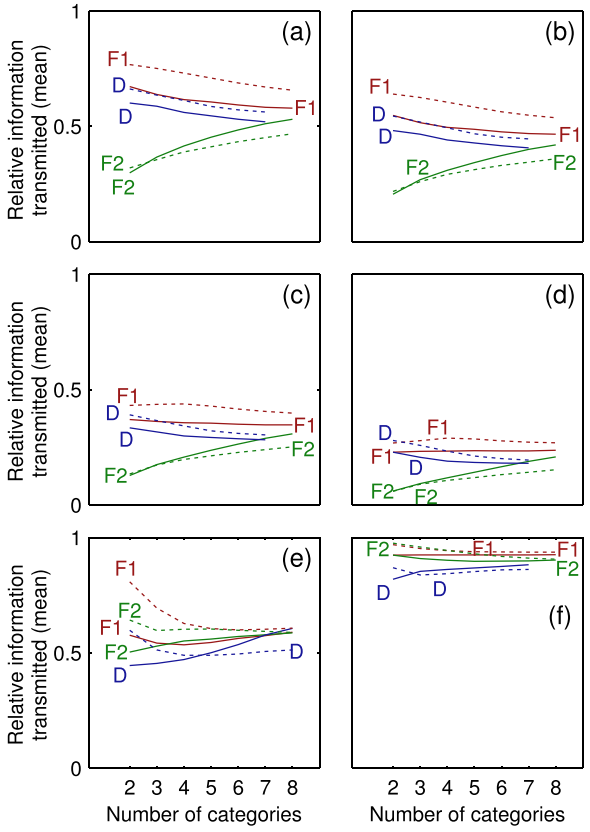

FIG. 7. (Color online) Mean relative information transmitted as reported by the crisp (solid lines) and fuzzy (dotted lines) FITA techniques for 1000 randomly selected sets of boundary locations. Panels (a) to (d) show FITAs performed on data recorded in our lab at four SNRs, decreasing linearly from −10 dB in (a) to −13 dB in (d). Panels (e) and (f) show FITAs performed on the data recorded by Van Wieringen and Wouters (1999) and Hillenbrand *et al.* (1995), respectively. Where data from multiple listeners were available, these data were averaged across listeners in the final step before drawing to the graph.

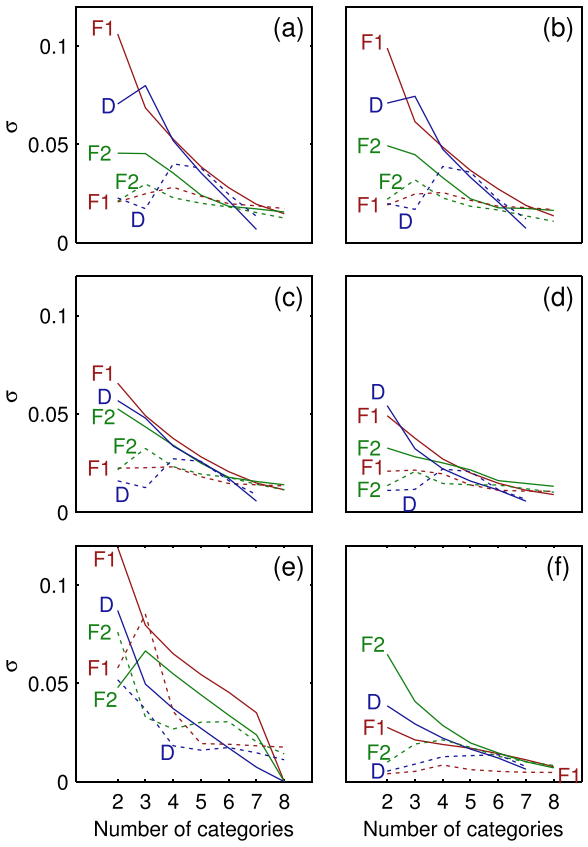

FIG. 8. (Color online) Standard deviation (σ) of the relative information transmitted as reported by the crisp (solid lines) and fuzzy (dotted lines) FITA techniques for 1000 randomly selected sets of boundary locations. Panels (a) to (d) show FITAs performed on data recorded in our lab at four SNRs, decreasing linearly from −10 dB in (a) to −13 dB in (d). Panels (e) and (f) show FITAs performed on the data recorded by Van Wieringen and Wouters (1999) and Hillenbrand *et al.* (1995), respectively. Where data from multiple listeners were available, these data were averaged across listeners in the final step before drawing to the graph.

almost all of the variance in the data for all of the features considered, less than 1% of this variance was explained by the boundary location when using a fuzzy FITA as opposed to approximately 10% for two features when using a crisp FITA. When random boundary locations were selected in the Monte Carlo simulation, the standard deviation in reported relative information transmitted was less than 0.04 for all combinations of variables and close to 0.02 for most combinations when a fuzzy FITA with four of five categories were used. This result confirms the ability of the fuzzy FITA to reduce the effect of the boundary location across a range of listener performance levels.

The effect of the second grouping parameter, namely, the number of categories, appears inherent to the mechanism used to isolate individual features, namely, by grouping tokens according to the feature of interest. When using a small number of categories, not all information transmitted by a feature is characterized, and differences between features are often exaggerated. When using a large number of categories, information transmitted by other features enters the equation, and differences between features are attenuated. No technique based on this mechanism of feature isolation is exempt from this effect.

Because no independent reference is available, determining which number of categories represents the most accurate estimation of transmitted information is not straightforward. For a fuzzy FITA, the standard deviation is close to its minimum when four or five categories are used (see Fig. 8). The effect of boundary location is therefore small for four or five categories, and in the opinion of the authors, four categories would normally be a good choice for 12 tokens. A number of categories slightly larger than the square root of the number of tokens would be a good initial choice in the opinion of the authors, but this warrants further investigation for cases where the number of tokens differs substantially from the 12 used in the present study.

B. Applicability of the fuzzy FITA

The fuzzy FITA is applicable to a broad range of feature-based classification problems, including classification tasks performed by humans and computational classification algorithms. For human classifiers, the fuzzy FITA may be used as an exploratory tool to determine the degree to which different features are being used as elucidated by the experiments in the present study. For classification algorithms, the fuzzy FITA may be used as a confirmatory tool to verify that the intended features have been used effectively. It may be used to analyze the effect of noise on different features for both types of classifiers. The fuzzy FITA may also be used to identify the features used by humans as a starting point toward developing an automatic classification algorithm.

The applicability of the fuzzy FITA to a particular problem depends on the type of features used, the number of classes and the typical performance of the classifier. Features should be scalar values defined over a continuous interval. Examples of non-scalar features include formant frequency contours (Neel, 2004) and spectral shape templates (Hillenbrand and Houde, 2003). These may still be used in a fuzzy FITA if they are sampled at a few discrete points. For example, a formant frequency contour may be represented by the frequencies at the start, middle, and end of the time segment (Neel, 2004). These samples are then interpreted as three separate scalar features when used in a fuzzy FITA. Alternatively, scalar features may be defined to describe non-scalar features in terms of their mean, variance, maximum slope, or other descriptive parameter (Chen and Maher, 2006). The requirement that features should be defined over a continuous interval excludes categorical features, which are handled effectively by the existing crisp FITA. Umapathy and co-workers (2007) list mel-frequency cepstral coefficients, timbral texture, band periodicity, linear prediction coefficient derived cepstral coefficients, zero-crossing rate, and MPEG-7 descriptors as typical features used in audio classification, all of which satisfy the two requirements in the preceding text.

For the fuzzy FITA to apply to a classification problem, the number of classes should be finite, but not too small. If the number of classes is infinite, a confusion matrix cannot be constructed. However, if the number is smaller than around six, it is unlikely that the confusion matrix will contain meaningful information about the use of specific features. Similarly, the performance of the classifier must be above chance but not too good. It is impossible to determine whether any features were used if performance is below chance. If performance is almost perfect, however, the fuzzy FITA does not have access to the information required to analyze feature transmission, which resides in the observed confusions. For classifiers that perform too well in quiet conditions, the fuzzy FITA becomes useful when the classification task is performed in noise.

Examples of acoustic classification problems apart from vowel recognition that satisfy the requirements for the applicability of the fuzzy FITA include automatic emotion classification (Ooi *et al.*, 2014), animal vocalization classification (Clemins *et al.*, 2005; Chen and Maher, 2006; Binder and Hines, 2014), aircraft classification (Sánchez Fernández *et al.*, 2013), and musical genre classification (Tzanetakis and Cook, 2002). Where some of these studies do not exactly satisfy all the applicability requirements, they can be adjusted to fit the requirements by adding background noise or increasing the number of classes.

One shortcoming of the fuzzy FITA is that it does not consider how consistently a feature is produced. For example, if the same speaker repeats the same token and features are extracted from every utterance, features that have a low standard deviation presumably contain more information than features that have a high standard deviation. This factor is not accounted for by either the crisp or the fuzzy FITA because neither of these techniques accepts repeated utterances as inputs. The problem of accounting for the repeatability of features therefore remains to be solved.

For problems that satisfy the preceding applicability requirements and that are not severely affected by the preceding shortcoming, the fuzzy FITA is appropriate for use with automatically selected boundaries. Even if a flawless way of defining boundaries was available, the fuzzy FITA is still preferred over the crisp FITA for continuous features because it constitutes a more accurate representation of distances between tokens (see Sec. II B). In addition, because the fuzzy FITA is less sensitive to human parameters, its use promotes comparability between studies by different researchers, and it reduces the risk of bias or noise added due to human error.

C. Recommendations

Several recommendations can be made based on results from the present study. The first is to use a fuzzy FITA when working with continuous features and a crisp FITA when working with categorical features, because of the advantages outlined in the preceding text and because this would correspond to the context in which both techniques were designed. Second, the same number of categories should be used for each feature when comparing different features because the number of categories has a measureable effect on the output. The third recommendation is to use a number of categories slightly larger than the square root of the number of tokens when working with continuous features (unless a good reason exists to do otherwise, such as well-defined

clusters in the token space). For categorical features, the number of categories should be guided by the nature of the categories. A final recommendation is to use automatically selected, equal-spaced categories because this will promote comparability between studies if similar speech material is used. Furthermore automating the grouping process will enable researchers to process larger datasets, for example, more listeners, more SNRs, and a wider variety of features.

[1]The term "grouping parameter" is used in this document as a collective noun describing all manually adjustable parameters that affect the assignment of tokens to categories, including the two that are investigated in the present study, namely, the number of categories and the boundary locations.

[2]This new technique, hereinafter referred to as the fuzzy FITA, was implemented using MATLAB, a product of The Mathworks (www.mathworks.com). The implementation is available online on the website of the research group (www.up.ac.za/bioengineering).

[3]The adjective "crisp" in the term "crisp FITA" is borrowed from fuzzy set theory, where ordinary sets are referred to as "crisp sets" to distinguish them from fuzzy sets.

[4]The term "categories" is retained for uniformity. In fuzzy set theory, this entity is known as a "fuzzy set."

Apoux, F., and Healy, E. W. (**2012**). "Use of a compound approach to derive auditory-filter-wide frequency-importance functions for vowels and consonants," J. Acoust. Soc. Am. **132**, 1078–1087.

Binder, C. M., and Hines, P. C. (**2014**). "Automated aural classification used for inter-species discrimination of cetaceans," J. Acoust. Soc. Am. **135**, 2113–2125.

Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., and Clark, G. M. (**1989**). "Speech perception using combinations of auditory, visual, and tactile information," J. Rehabil. Res. Dev. **26**, 15–24.

Boersma, P., and Weenink, D. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**, 341–345.

Botha, L. (**1996**). "Towards modelling acoustic differences between L1 and L2 speech: The short vowels of Afrikaans and South-African English," in *Proceedings of the Institute of Phonetic Sciences* (University of Amsterdam), pp. 65–80.

Chen, Z., and Maher, R. C. (**2006**). "Semi-automatic classification of bird vocalizations using spectral peak tracks," J. Acoust. Soc. Am. **120**, 2974–2984.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," J. Acoust. Soc. Am. **117**, 956–963.

Donaldson, G. S., Dawson, P. K., and Borden, L. Z. (**2011**). "Within-subjects comparison of the HiRes and Fidelity120 speech processing strategies: Speech perception and its relation to place-pitch sensitivity," Ear Hear. **32**, 238–250.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Houde, R. A. (**2003**). "A narrow band pattern-matching model of vowel perception," J. Acoust. Soc. Am. **113**, 1044–1055.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Neel, A. T. (**2004**). "Formant detail needed for vowel identification," Acoust. Res. Lett. Online **5**, 125–131.

Ooi, C. S., Seng, K. P., Ang, L. M., and Chew, L. W. (**2014**). "A new approach of audio emotion recognition," Expert Syst. Appl. **41**, 5858–5869.

Sánchez Fernández, L. P., Sánchez Pérez, L. A., Carbajal Hernández, J. J., and Rojo Ruiz, A. (**2013**). "Aircraft classification and acoustic impact estimation based on real-time take-off noise measurements," Neural Process. Lett. **38**, 239–259.

Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.

Sheffield, B. M., and Zeng, F. G. (**2012**). "The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception," J. Acoust. Soc. Am. **131**, 518–530.

Taylor, J. R., and Uys, J. Z. (**1988**). "Notes on the Afrikaans vowel system," Leuvense Bijdragen **77**, 129–149.

Tzanetakis, G., and Cook, P. (**2002**). "Musical genre classification of audio signals," IEEE Trans. Speech Audio Process. **10**, 293–302.

Umapathy, K., Krishnan, S., and Rao, R. K. (**2007**). "Audio signal feature extraction and classification using local discriminant bases," IEEE Trans. Audio Speech Lang. Process. **15**, 1236–1246.

Van der Merwe, A., Groenewald, E., Van Aardt, D., Tesner, H. E. C., and Grimbeek, R. J. (**1993**). "Die formantpatrone van Afrikaanse vokale soos geproduseer deur manlike sprekers" ("Formant patterns of Afrikaans vowels as produced by male speakers"), Suid Afrikaanse Tydskrif vir Taalkunde **11**, 71–79.

Van Wieringen, A., and Wouters, J. (**1999**). "Natural vowel and consonant recognition by Laura cochlear implantees," Ear Hear. **20**, 89–103.

Yoon, Y. S., Li, Y., and Fu, Q. J. (**2012**). "Speech recognition and acoustic features in combined electric and acoustic stimulation," J. Speech. Lang. Hear. Res. **55**, 105–124.

Zadeh, L. A. (**1965**). "Fuzzy sets," Inf. Control **8**, 338–353.