New approaches to feature information transmission analysis (FITA)

Dirk J J Oosthuizen & Johan J Hanekom
Bioengineering, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

FITA: A measurement tool

The acoustic features that underlie speech perception have often been studied to advance understanding of human speech perception and to aid in the development of hearing prostheses. These features are typically investigated using closed set phoneme identification experiments (e.g. Miller & Nicely, 1955; Van Wieringen & Wouters, 1999). Results from such experiments are processed by a technique known as feature information transmission analysis (FITA) (Miller and Nicely, 1955) to obtain quantitative estimates of the amounts of information transmitted by individual features.

FITA was originally used with categorical features (e.g. voicing), for which the technique was originally developed (Miller & Nicely, 1955). In time, application expanded to continuous features (e.g. formant frequencies (Blamey *et al*., 1989; Van Wieringen & Wouters, 1999). The FITA technique treats continuous features as categorical, which gives rise to several problems, as described below. Two alternative techniques, namely the fuzzy FITA and continuous FITA, have been developed to address these problems, as discussed in subsequent sections. First, a model of the communication process and different information metrics will be discussed.

Information flow model

The main factors that influence the flow of information in a closed set phoneme identification experiment may by summarized by the model in Figure 1.

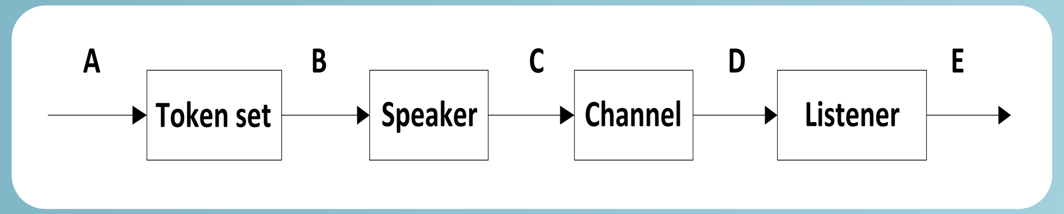


Figure 1. A model of information flow in a closed set phoneme identification experiment.

In the context of listening experiments, the "channel" may include any physical or simulated channel and any signal processing applied to the stimuli. Examples include an additive noise channel, the deliberate masking of selected acoustic features (Swanepoel *et al*., 2012) and the distortion of spectral content (Shannon *et al*., 1998).

If the use of a particular feature is not evident from experimental results, this may indicate one of four causes, namely that the feature was absent from the selected token set, that the speaker was unable to produce the feature, that the feature was not transmitted effectively by the channel or that the listener was unable to perceive the feature.

Three information metrics are reported by a FITA. The absolute transmitted feature information (T_{abs}) is the amount of information transmitted from point B to point E in Figure 1. The feature entropy (H_f) is the amount of information present at point B. The relative transmitted feature information (T_{rel}) is the ratio T_{abs}/H_f. T_{rel} is an important metric when evaluating the channel, whereas H_f and T_{abs} are important when evaluating the features.

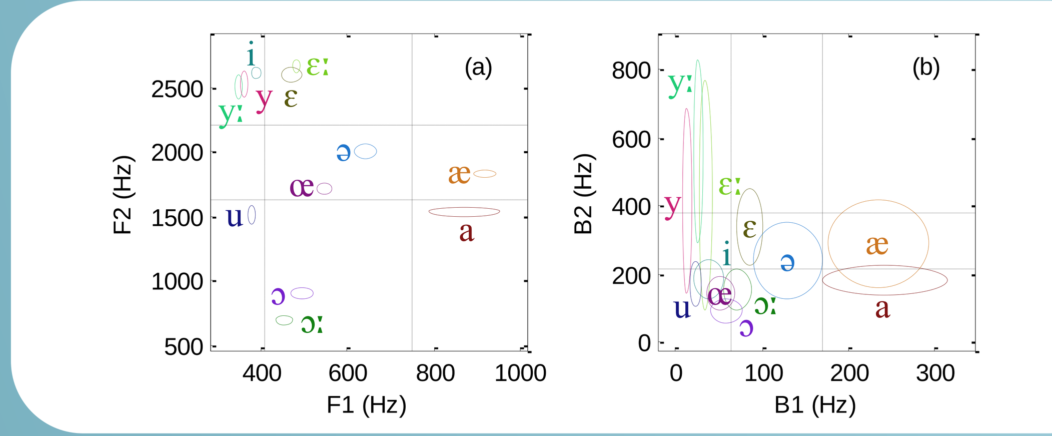


Figure 3. Token spaces for four features. Ellipse centres represent mean feature values and ellipse widths and heights represent twice the standard deviations in the corresponding dimensions after a single speaker produced each token multiple times. Boundary lines indicate divisions of tokens into categories along the four feature dimensions.

The traditional FITA

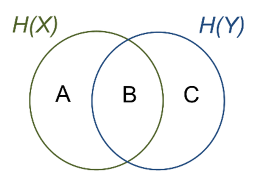
The traditional FITA (Miller & Nicely, 1955) is based on the Shannon's mathematical theory of communication (Shannon, 1948). Shannon used random variables *X* and *Y* to denote the messages produced at the source and received at the destination, respectively. He defined the source entropy, *H(X)* and the destination entropy, *H(Y)* as measures of the amount of uncertainty about the produced and perceived messages. The conditional entropy *H(Y|X)* was defined as the amount of uncertainty about the perceived message that is resolved by observing the source message. The mutual information *I(X,Y)* between the source and destination was defined as the difference between *H(Y)* and *H(Y|X)*. These relationships may be visualized on the Venn diagram in Figure 2.

All the above entropies and the mutual information can be calculated from the probabilities $P(x_i)$ of each message being produced and the conditional probabilities $P(y_j|x_i)$ of each message being perceived given that each message is produced.

Miller & Nicely (1955) recorded listening experiment results in a token confusion matrix (e.g. Figure 2). Tokens were assigned to categories according to the feature under investigation (e.g. Figure 3) and a category confusion matrix was compiled (e.g. Figure 2). $P(x_i)$ and $P(y_j|x_i)$ were derived directly from the category confusion matrix by normalization and the relevant entropies and mutual information were computed. H_f was set to $H(X)$, T_{abs} to $I(X,Y)$ and T_{rel} to $I(X,Y)/H(X)$.

	a	æ	ɛ	ɛː	ɔː	yː	i	ə	u	ɔ	œ	y
a	363	36	12	1	9	10	20	17	15	31	41	21
æ	72	452	3	6	6	0	6	6	8	5	6	6
ɛ	3	2	282	16	12	10	39	41	27	58	56	30
ɛː	0	3	34	224	241	20	4	4	12	20	5	9
ɔː	3	2	15	118	326	27	1	5	8	53	14	4
yː	0	0	2	4	7	411	20	0	21	1	0	110
i	1	0	33	2	2	3	346	8	99	2	6	74
ə	4	5	19	5	6	2	1	384	0	19	130	1
u	2	2	57	10	12	19	66	13	250	76	24	45
ɔ	9	4	83	12	35	2	1	65	18	238	107	2
œ	3	4	74	8	12	1	4	86	5	106	271	2
y	0	0	25	6	0	39	105	6	75	6	8	306

	Cat1	Cat2	Cat3
Cat1	1989	310	5
Cat2	230	3184	42
Cat3	86	143	923



$H(X)$ = A+B
$H(Y)$ = B+C
$H(X|Y)$ = A
$H(Y|X)$ = C
$I(X,Y)$ = B

Figure 2. Token confusion matrix (left), category confusion matrix (top right) and Venn diagram describing the relation ship between entropy, conditional entropy and mutual information (bottom right). Each entry in a confusion matrix represents the number of times the token or category in the corresponding row was identified as the token of category in the corresponding column.

Problems with continuous features

The use of the traditional FITA with continuous features poses three problems. Firstly, the assignment of tokens to groups is subjective and no official guidelines for this process exist. This requires effort on the part of the researcher and the estimated information metrics were found to be sensitive to the selected number of categories and category boundary locations.

Secondly, the precision with which a feature is produced is ignored by the traditional FITA. Features that are produced with higher precision have smaller standard deviations (see Figure 3) and contain more information about token identity.

Finally, the traditional FITA places an inaccurate upper bound on the maximum measurable information (see Figures 7 and 8). Theoretically, an ideal feature should be able to distinguish between all tokens in the set. The maximum information that can be measured by the traditional FITA is substantially lower than the information required to do this. Because the traditional FITA treats within-category confusions and correct responses the same, the information that enables a feature to distinguish between tokens in the same category is not measured.

The fuzzy FITA

The fuzzy FITA was developed to address the first problem mentioned above. Feature categories were implemented as fuzzy sets (Zadeh, 1965) so that each token can belong to multiple categories with different degrees of membership for each category. The conversion from the token confusion matrix to the category confusion matrix (Figure 2) may be implemented using matrix multiplications: the category confusion matrix is post-multiplied by a category assignment matrix and pre-multiplied by its transpose. In the traditional FITA, this category assignment matrix contains only ones and zeroes. In the fuzzy FITA, the category assignment matrix contains the degrees of membership with which each token belongs to each category.

Category membership functions were defined for each category to specify for any point on the feature value axis, the degree of membership of the token at that point to the category identified by the function. An example of a set of category membership functions is shown in Figure 4.

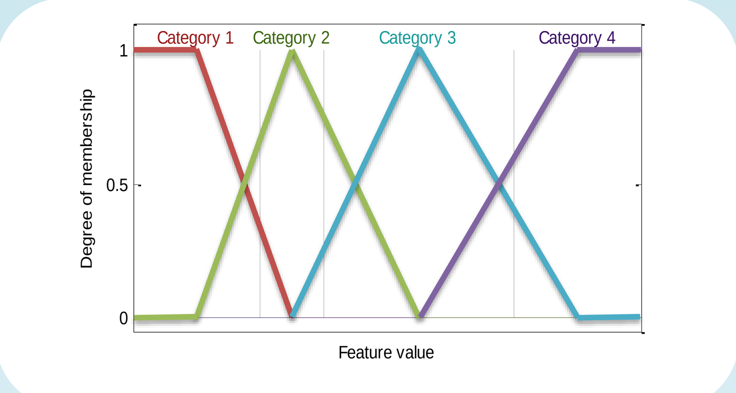


Figure 4. Example of category membership functions for a fuzzy FITA with four categories. Vertical lines indicate category boundaries.

The sensitivities of the traditional and fuzzy FITAs to the number of categories and category boundary locations were measured experimentally for three features (the first two formant frequencies and vowel segment duration) using data from two different studies (Van Wieringen & Wouters, 1999; Hillenbrand *et al*., 1995) as well as data from our lab. Results are presented in Figure 5. The sensitivity of the traditional FITA to the boundary location was deemed problematic (a linear regression analysis also showed that 10 % of the variance in the estimated information values were accounted for by the boundary location for two features when using the traditional FITA, compared to less than 1 % when using the fuzzy FITA). The fuzzy FITA was found to be sufficiently robust to boundary location to allow automation of boundary selection. However, the fuzzy FITA could not sufficiently reduce the effect of the number of categories. This effect appears to be inherent to the mechanism of feature isolation by category assignment and may be understood as a result of the third problem listed earlier.

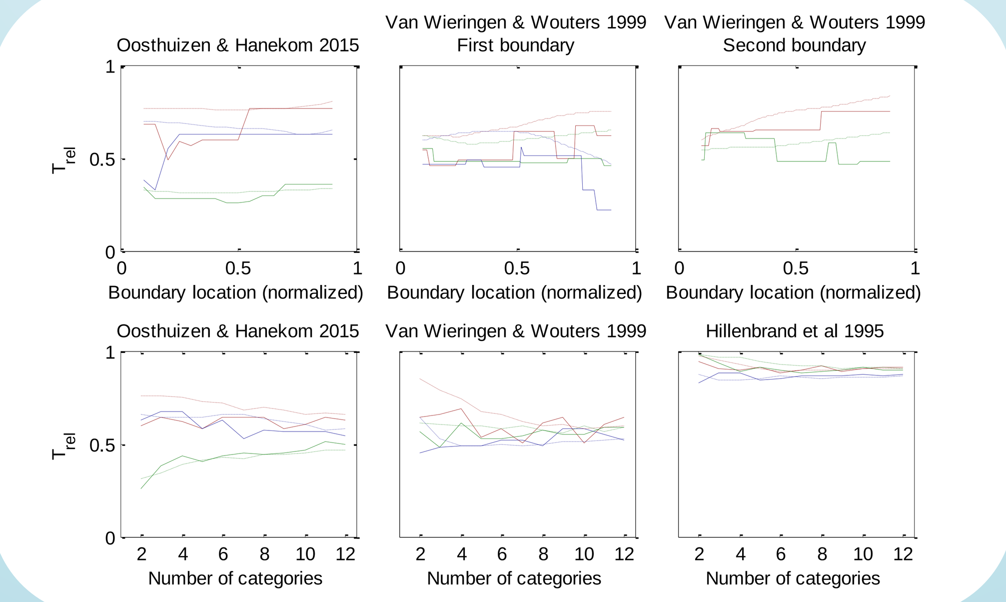
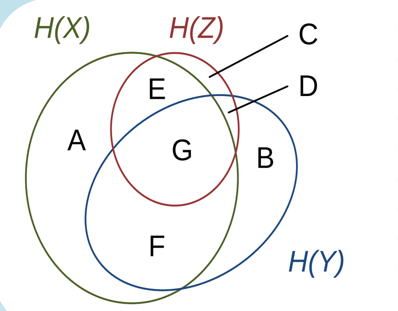


Figure 5. T_{rel} estimated by the traditional (solid lines) and fuzzy (dotted lines) FITAs for different boundary locations (top row) and different numbers of categories (bottom row) using data from three different studies. Red, green and blue traces represent the features F1, F2 and duration, respectively.

The continuous FITA

The continuous FITA was developed to address all three problems mentioned earlier by removing the concept of category assignments altogether. Instead, feature isolation is achieved by introducing a third random variable, *Z*, to denote the feature value. Unlike *X* and *Y*, *Z* is a continuous variable and is assumed to be normally distributed with its mean and standard deviation determined by extracting the feature under consideration from repeated recordings of the same token by the same speaker.

H_f is estimated as the mutual information $I(X,Z)$, T_{abs} is estimated as the multivariate mutual information $I(X,Y,Z)$ (Timme *et al.* 2014) and T_{rel} is estimated as the ratio $I(X,Y,Z)/I(X,Z)$. The relevant entropies and mutual information quantities can be visualized on Figure 6.



$H(X)$ = A+E+F+G
$H(Y)$ = B+D+F+G
$H(Z)$ = C+D+E+G
$H(X|Y)$ = A+E
$H(Y|X)$ = B+D
$H(X|Z)$ = A+F
$H(Z|X)$ = C+D
$H(Y|Z)$ = B+F
$H(Z|Y)$ = C+E

$I(X,Y)$ = F+G
$I(X,Z)$ = E+G
$I(Y,Z)$ = D+G
$I(X,Y|Z)$ = F
$I(X,Z|Y)$ = E
$I(Y,Z|X)$ = D
$I(X,Y,Z)$ = G

H_f = E+G
T_{abs} = G
T_{rel} = G/(E+G)

Figure 6. Venn diagram of the entropies of random variables used in the continuous FITA calculation. The area of each ellipse is proportional to the entropy of the represented variable (*X*, *Y* or *Z*) and the area of each overlapping region is proportional to the mutual information between the variables represented by the overlapping ellipses. The shapes of the ellipses have no special significance.

H_f, T_{abs} and T_{rel} were measured for five features (two formant frequencies, two formant bandwidths and vowel segment duration) using data recorded in our lab. T_{abs} and T_{rel} were measured at different signal-to-noise ratios (SNRs). H_f was also estimated by presenting one feature at a time to an ideal observer, constructing a confusion matrix from its responses and estimating T_{abs} for the ideal observer (T_{abs} of the ideal observer was taken as H_f of the feature presented). Results are shown in Figures 7 and 8.

The upper bound effect is clear from Figure 7 for the traditional and fuzzy FITAs. The effect is even more severe for the duration feature, for which two categories were used, than for the other features, for which three categories were used. The continuous FITA appears not to suffer from this effect. For the strong features, H_f is near 70% of its maximum value. The traditional and fuzzy FITAs do not clearly distinguish between high-precision (F1 and F2) and low-precision (B1 and B2) features. The continuous FITA achieves this distinction, as shown in Figures 7 and 8. The H_f estimates by the continuous FITA are almost identical to the amounts of information transmitted to the ideal observer. The values of T_{abs} estimated by the continuous FITA agree with expectations, starting out at values that reflect the H_f estimates and decreasing almost linearly with a decrease in SNR. T_{rel} estimates are more similar across FITA methods, which may explain why most researchers only report T_{rel} (e.g. Blamey *et al*., 1989; Van Wieringen & Wouters, 1999). Although T_{rel} is useful for evaluating the communication channel (Figure 1), H_f and T_{abs} are more suitable metrics for evaluating features. The continuous FITA has also removed the need to assign tokens to categories, thereby successfully addressing all three problems mentioned earlier.

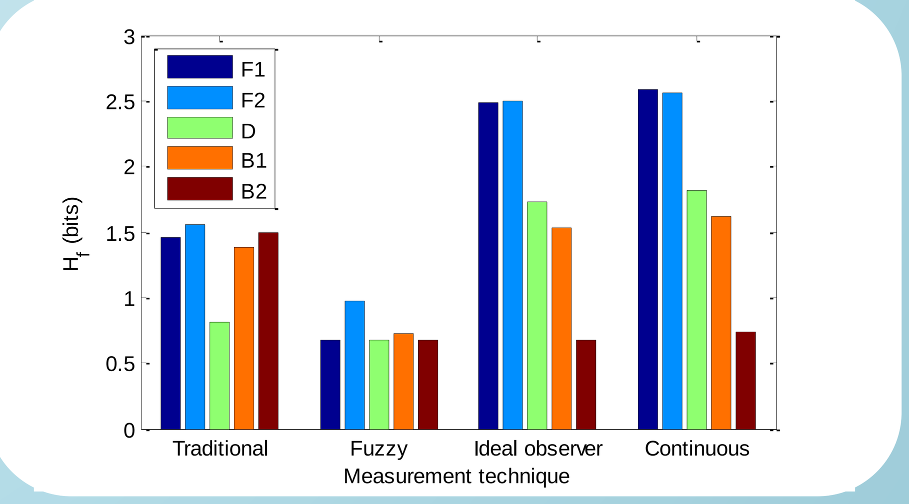


Figure 7. H_f of five features estimated using four techniques.

A multi-feature extension

The continuous FITA can be extended to estimate the information contained in combinations of features. This is done by defining *Z* as a multivariate random variable characterized by a multivariate random distribution. All parameters of this distribution can be calculated from feature values extracted from repeated recordings of each token by the same speaker. By subtracting the information metrics calculated for the individual features from those calculated for the combination, it is possible to estimate the degree of redundancy between features and the amount of unique information provided by a feature. With this extension, is it also possible to mathematically determine whether a set of features completely characterize the source of whether additional important features exist.

Conclusion

Three problems that arise when using the traditional FITA with continuous features, have been demonstrated and two alternative FITAs have been developed to address these. The fuzzy FITA addresses only the problem of manual category assignments and is preferred over the traditional FITA when repeated recordings of each token by the same speaker are unavailable. The continuous FITA addresses all three problems and is recommended for all studies involving continuous features if repeated recordings can be obtained. The traditional FITA is recommended for all studies involving categorical features. The continuous FITA can also estimate information metrics for combinations of features, from which redundancy, uniqueness and completeness measures can be derived.

References

BLAMEY, P. J., COWAN, R. S. C., ALCANTARA, J. I., WHITFORD, L. A., & CLARK, G. M. 1989. Speech perception using combinations of auditory, visual, and tactile information. Journal of Rehabilitation Research and Development, 26, 15-24.

HILLENBRAND, J., GETTY, L. A., CLARK, M. J., & WHEELER, K. 1995. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97, 3099-3111.

MILLER, G. A. & NICELY, P. E. 1955. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.

OOSTHUIZEN, D. J. J., & HANEKOM, J. J. 2015. Fuzzy information transmission analysis for continuous speech features. Journal of the Acoustical Society of America, 137, 1983-1994.

SHANNON, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal, 27, 379-423.

SWANEPOEL, R., OOSTHUIZEN, D. J. J. & HANEKOM, J. J. 2012. The relative importance of spectral cues for vowel recognition in severe noise. Journal of the Acoustical Society of America, 132, 2652-2662.

TIMME, N., ALFORD, W., FLECKER, B. AND BEGGS, J. M. 2014. Synergy, redundancy, and multivariate information measures: An experimentalist's perspective. Journal of Computational Neuroscience, 36, 119-140.

VAN WIERINGEN, A. & WOUTERS, J. 1999. Natural vowel and consonant recognition by Laura cochlear implantees, Ear and Hearing, 20, 89-103.

ZADEH, L. A. 1965. Fuzzy sets. Information and Control, 8, 338-353.

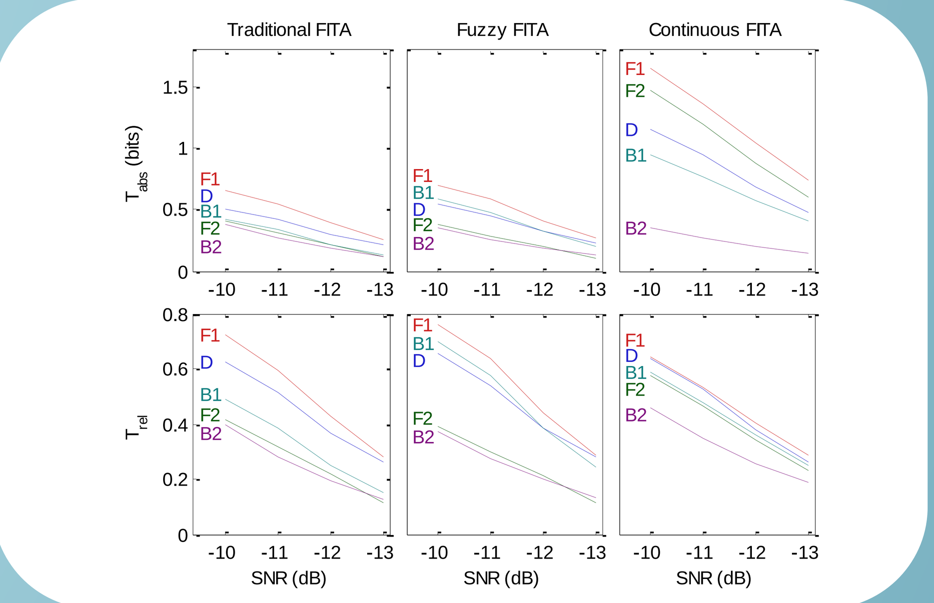
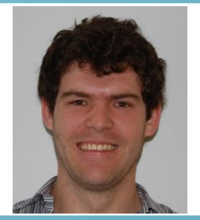


Figure 8. T_{abs} and T_{rel} as a function of SNR for five features estimated with the traditional and continuous FITA.

Bioengineering

Software tools available online:
www.up.ac.za/bioengineering



Dirk J J Oosthuizen
University of Pretoria

Johan J Hanekom, Ph.D
University of Pretoria
johan.hanekom@up.ac.za