

A novel phonemes classification method using fuzzy logic

Ines Ben Fredj¹, Kaïs Ouni²

¹Department of Electrical Engineering, National Engineering School of Tunis, Tunis, Tunisia

²Department of Electrical Engineering, Higher School of Technology and Computer Science of Carthage, Tunis, Tunisia

Email address:

ines_benfredj@yahoo.fr (I. B. Fredj), kais.ouni@enit.rnu.tn (K. Ouni)

To cite this article:

Ines Ben Fredj, Kaïs Ouni. A Novel Phonemes Classification Method Using Fuzzy Logic. *Science Journal of Circuits, Systems and Signal Processing*. Vol. 2, No. 1, 2013, pp. 1-5. doi: 10.11648/j.cssp.20130201.11

Abstract: In this study, we will interest in phonemes classification of Timit database using Fuzzy Logic. The fuzzy method consists in the extraction of a three fuzzy-reference vectors: maximal, mean and minimal. To classify a phoneme request, we calculate its degree of membership to all defined classes. The class of a phoneme request is, then, the one which maximizes one degree of membership calculated according to reference vectors. Different techniques of speech analysis are used for evaluation. Results show that fuzzy logic can provide a significant issue when mathematical rigor is not present as to the signal processing since the retained recognition rates was 90,85%, 22,96%, 98,57% and 91,73% for respectively MFCC, LPC, PLP and RASTA PLP.

Keywords: Fuzzy Logic; MFCC; LPC; PLP; RASTA-PLP; Speech; Timit

1. Introduction

Usually, speech recognition systems suppose that speech signal is the encoding of some messages as a sequence of one or more symbols. To converting this encoded message to an oral statement, the continuous speech signal is firstly transformed into a set of vectors of discrete parameters and then recognized using a language model [1] [2].

Always, a language model of an automatic speech recognition system requires a kind of artificial intelligence for obtaining considerable results [3] [4].

In this paper, we propose a new language model of phonemes classification based on the concept of fuzzy logic evaluated with different extracting features such as MFCC, LPC, PLP and RASTA-PLP.

This paper is organized as follows:

We will be first presenting the speech analysis techniques used followed by the definition of fuzzy logic. After that, we will explain our phonemes classification approach and we will comment the experimental results. Then, other works of phonemes recognition are exposed for evaluation. Finally, we will finish with a conclusion and some perspectives.

2. Extracting Features

There are many techniques for extracting features from speech signal such as Mel-frequency cepstral coefficients

(MFCC), Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP), and Relative Spectral PLP (RASTA-PLP).

2.1. MFCC

The MFCC coefficients are mainly used for speech features representation. They are calculated by a Discrete Cosine Transform applied to the power spectrum' signal according to the Mel scale [5].

The algorithm of MFCC is as follows:

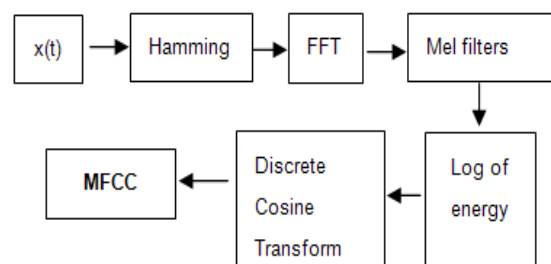


Figure 1. MFCC algorithm..

We take the Fourier transform of a signal windowed by the hamming window. We map the powers of the spectrum obtained above onto the Mel scale. We take the logs of the powers at each of the Mel frequencies [6].

We get the discrete cosine transform of the list of Mel log powers and apply the inverse Fourier transform to ob-

tain the MFCC coefficients.

2.2. LPC

Linear Predictive Coding (LPC) is a reliable speech analysis technique. It is also a useful method for encoding good quality speech at a low bit rate. It provides extremely accurate estimates of speech parameters [7].

LPC' algorithm is described in figure 2.

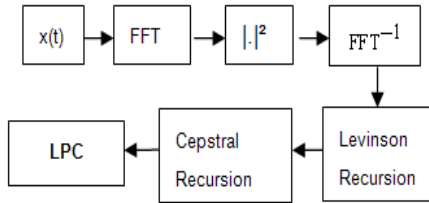


Figure 2. LPC algorithm..

First, Fourier transform of the signal is applied. Then, calculating the inverse Fourier transform of its module squared. Finally, we pass to Levinson and cepstral recursion for getting LPC coefficients.

2.3. PLP

PLP was studied by Hermansky in 1990 [8]. This technique is based on concepts from the psychophysics of hearing to derive an estimate of the auditory spectrum: the critical-band spectral resolution, the equal-loudness curve and the intensity-loudness power law.

The power spectrum is obtained with a Bark filter bank with a subsequent equal loudness pre-emphasis and a compression based on cube-root.

The auditory spectrum is then approximated by an autoregressive all-pole model.

The PLP algorithm can be presented as follows:

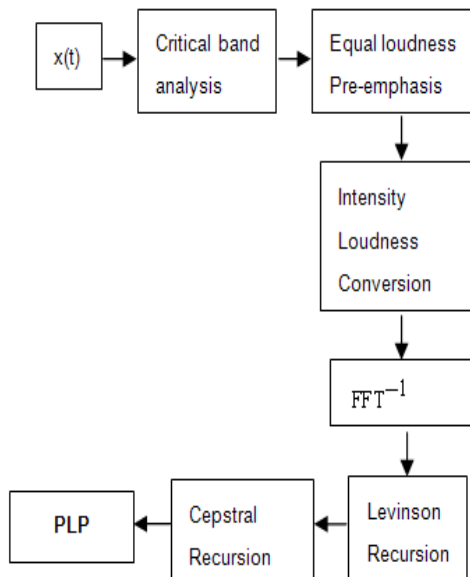


Figure 3. PLP algorithm..

2.4. RASTA-PLP

RASTA PLP was introduced by Hermansky and others in 1991 [9]. Rasta makes PLP more robust to linear spectral distortions. Rasta is based on the filtering of time trajectories of outputs from critical-band filters.

The algorithm of RASTA PLP is as follows:

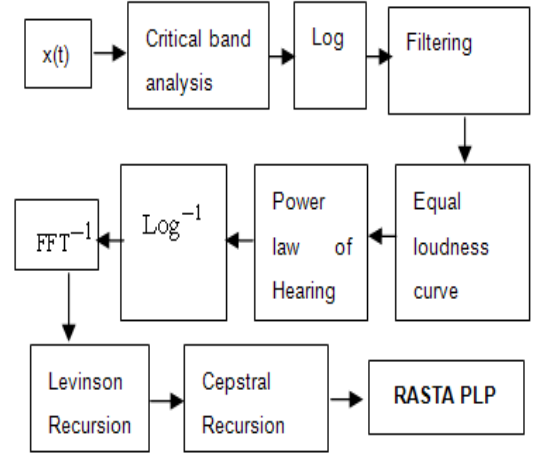


Figure 4. RASTA-PLP algorithm..

3. Fuzzy Logic

3.1. Review Stage

Fuzzy logic, proposed by Lotfi A. Zadeh in 1965 [10], is a theory that allows natural linguistic of problems to be solved rather than using numerical values.

Fuzzy logic formalizes and operates the same way human arguing does. It's based on the mathematical theory of the fuzzy groups [11]. This theory is a classic extension of the theory for the consideration of groups defined in an indistinct way.

Furthermore, Fuzzy Logic was conceived as a reliable method for sorting and handling data but has proven, also, to be an excellent choice for many control system applications since it mimics human control logic [12].

3.2. Principe

According to classical logic, decisions are binary having two-valued logic: true or false [13].

This point distinguishes fuzzy logic from classical logic. In fuzzy logic, a decision can be true and false with a degree of membership in each of these two cases.

Fuzzy logic describes uncertain and imprecise declaration using that each element belongs partially or gradually to defined sets [14]. For example, the statement "Today, it is a nice day" according to fuzzy logic, is 100% true if there are no clouds, 80% true if there are a few clouds, 50% true if there are a lots of clouds and 0% true if it rains all day.

To conclude, in fuzzy logic an element belongs to a "fuzzy" set (not strictly to one set).

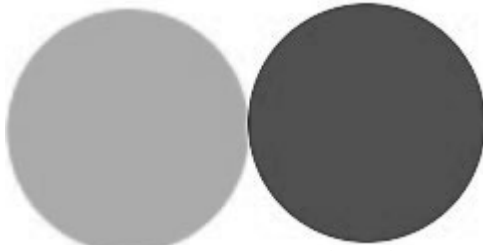


Figure 5. Representation sets using Classical Logic [13].

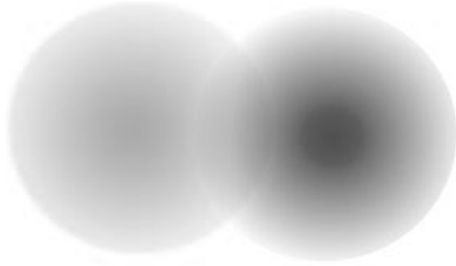


Figure 6. Representation sets using Fuzzy Logic [13].

4. Approach of Classification

4.1. Materiel

We have used the New England dialect DR1 of the TIMIT speech corpus [15].

TIMIT corpus was used to train and evaluate speaker-independent phoneme recognizers. It consists of 630 speakers, each saying 10 sentences.

We apply MFCC, LPC, PLP and RASTA PLP to obtain a database of cepstral parameters. Each vector of parameter contains 12 coefficients characterized by three middle windows. It was extracted from the speech signal with 256 sample frames and was Hamming windowed in segments of 25 ms length every 10 ms with a sampling frequency equal to 16000 KHz.

We have organized the database into seven homogenous phonemes classes which represent vowels, semivowels, affricates, fricatives, stops, nasals and others as illustrated in table 1.

Table 1. Classes' distribution of DR1_TIMIT database.

Class	Label
Affricates	/jh/ /ch/
Fricatives	/s/ /sh/ /z/ /zh/ /f/ /th/ /v/ /dh/
Nasals	/m/ /n/ /ng/ /em/ /en/ /eng/ /nx/
Semi-Vowels	/l/ /r/ /w/ /y/ /hh/ /hv/ /el/
Stops	/b/ /d/ /g/ /p/ /t/ /k/ /dx/ /q/ /bc/ /dc/ /gc/ /pc/ /tc/ /kc/
Vowels	/iy/ /ih/ /eh/ /ey/ /ae/ /aa/ /aw/ /ay/ /ah/ /ao/ /oy/ /ow/ /uh/ /uw/ /ux/ /er/ /ax/ /ix/ /axr/
Others	/ax-h/

4.2. Algorithm of Classification

The adopted fuzzy classification algorithm is based on the application of fuzzy reference vectors such as minimal, mean and maximal vector relative to each class of phonemes [16].

After features extraction, we obtain for each phoneme a matrix of 12 coefficients * 3 windows (these matrix are MFCC, LPC, PLP and RASTA PLP coefficients). We collect all phonemes according to the distribution described in Table 1 and we apply the algorithm of classification as following:

- First, giving a database of parameters (MFCC, LPC, PLP and RASTA PLP) according to each class (vowels, fricatives, stops.).
- Second, starting by extracting the references vectors: minimal vector V_{\min} , mean vector V_{mean} and maximal vector V_{\max} for all classes.
- Third, we compare the samples of every class with the various references vectors extracted. Those vectors are extracted in aim to calculate the degree of membership of every sample in all the classes using its norms.

The degree of membership “ $D_{r,c}$ ” of a vector “ V_r ” to a class “c” is defined by:

$$D_{r,c} = \frac{V_{\max_c} - V_{\text{mean_c}}}{V_r - V_{\text{mean_c}}}, \text{ if } V_{\text{mean_c}} \leq V_r \leq V_{\max_c}$$

$$D_{r,c} = \frac{V_r - V_{\min_c}}{V_{\text{mean_c}} - V_{\min_c}}, \text{ if } V_{\min_c} \leq V_r < V_{\text{mean_c}}$$

$$D_{r,c} = 0, \text{ otherwise}$$

- Then, we obtain for every sample one degree of membership concerning every class.
- Finally, we choose the class relative to the highest degree of membership.

To develop the various parameterization techniques (MFCC, LPC, PLP and RASTA PLP), we have used the implementation of Dan Elis described in [17]. In addition, the simulation of the classifier was performed under Matlab environment.

5. Experimental Results

Table 2 illustrates the rates of recognition of all classes using MFCC, LPC, PLP and RASTA PLP.

We notice that the interest recognition rate was obtained with PLP for the majority of classes. However, RASTA PLP gave remarkable results for many classes such as the Affricates, the Fricatives and the vowels. We observe also that MFCC has a good rate particularly with the Nasals, the semivowels, the vowels and the others. Generally, LPC was not the appropriate technique used with our classifier since modest rates it has given.

Table 2. Recognition Rates of all classes using MFCC, LPC, PLP and RASTA PLP.

Class	Number of Samples	MFCC (%)	LPC (%)	PLP (%)	RASTA PLP (%)
Affricates	510	60,00	43,92	99,00	98,23
Fricatives	5073	67,12	32,05	97,85	99,00
Nasals	3201	92,72	92,25	98,46	77,78
Semi-Vowels	4701	92,98	06,06	98,82	94,53
Stops	10031	92,21	24,08	98,29	84,43
Vowels	13669	96,94	07,33	98,96	94,91
Others	2899	99,82	24,04	98,65	97,24
Mean rate	40084	90,85	22,96	98,57	91,73

The number of samples varies from one class to another. The mean rate is relative to the number of samples and the techniques used in features extractions.

6. Similar State

For evaluation, we looked for other phoneme recognition works on the literature.

In [18], A. Strassberg, M. Molochnikov and R. Steinberg used a fuzzy inference system to recognize the plosive phonetic group which consists of the following phonemes: /b/ /d/ /g/ /p/ /t/ /k/ of TIMIT database.

They have shown that the knowledge of spectrogram readers can be exploited to build an expert system based on fuzzy logic.

Table 3 illustrates a comparison between the fuzzy inference system used and our classifier.

Table 3. Comparison Recognition Rates of plosives using fuzzy inference system and our classifier (Fuzzy Logic: FL).

Class	Rates (%) (spectrogram)	FL: MFCC (%)	FL: LPC (%)	FL: PLP (%)	FL: RASTA-PLP (%)
Stops	/b/	67,03			
	/d/	44,00			
	/g/	70,13			
	/p/	57,47			
	/t/	43,91			
	/k/	47,62			
	Mean rate	52,44	92,21	24,08	98,29

Furthermore, H.K. Tripathy, B.K. Tripathy and P.K. Das introduced fuzzy techniques for the classification of vowel data [19].

Experimental results have shown that the fuzzy inference system' performance was vastly improved over a standard MFCC features analysis of vowel recognition.

Table 4 shown the results obtained.

Table 4. Recognition Rates of vowels using fuzzy inference system.

Class	MFCC (%)	FL: MFCC (%)
Vowels	/a/	92,00
	/i/	82,00
	/e/	86,00
	/o/	93,00
	/u/	99,00
Mean rate	~ 90,40	92,21

In addition, a predictive neural network called Neural Predictive Coding (NPC) was presented in [20] by M. Chetouani, B. Gas, J.L.Zarader and C.Chavy.

This model is applied in non linear discriminated features extraction (DFE) applied to phoneme recognition.

Researchers also, have presented a new extension of the NPC model: DFE-NPC.

In order to evaluate the performances of the DFE-NPC model, researchers carried out a study of Timit phonemes recognition. Besides, they have compared the DFE-NPC with MFCC, PLP and RASTA-PLP.

A MLP (Multi Layered Perceptron) was used in classification.

Table 5 present Recognition rates obtained.

Table 5. Recognition Rates obtained with MLP classifier of some phonemes.

Class	MFCC (%)	PLP (%)	RASTA-PLP (%)	DEF-NPC (%)	FL: MFCC (%)	FL: PLP (%)	FL: RASTA-PLP (%)
Vowels	/aa/						
	/ae/	58,25	57,00	59,25	65,25	96,94	98,96
	/ey/						94,91
	/ow/						
	/b/						
Stops	/d/	62,00	64,00	64,00	73,00	92,21	98,29
	/g/						84,43
	/p/						
	/t/	66,60	66,00	68,33	70,30	92,21	98,29
	/q/						84,43

6. Conclusion

In this paper, we presented a novel method of phoneme classification based on the principle of fuzzy logic. The classifier consisted in the extraction of fuzzy reference vectors to determine the degrees of membership by class. For evaluation, we tested different techniques of features extraction such as MFCC, LPC, PLP and RASTA PLP. PLP gave the highest recognition rates for all classes. Though, MFCC and RASTA PLP were more satisfying than LPC.

This preliminary results show that our method is a promising approach for building a phoneme recognizer. Further work will focus on the improvement of the speech analysis techniques by adding first and second derivatives and energy to introduce temporal dynamic of parameters.

References

- [1] Anusuya, M.A., and Katti, S.K. "Front end analysis of speech recognition: a review", *International Journal of Speech Technology*, 2011, pp. 99–145.
- [2] Juang, B. H., and Rabiner, L.R.: "Automatic speech recognition - A brief history of the technology development" (Elsevier Encyclopedia of Language and Linguistics, 2005).
- [3] Gajsek, R., and Mihelič, F.: "Comparison of speech parameterization techniques for Slovenian language". *Proc. 9th International PhD Workshop on Systems and Control: Young Generation Viewpoint*, Slovenia, 2008.
- [4] Veeravalli, A.G., Pan, W.D., Adhami, R., and Cox, P.G.: "A tutorial on using hidden markov models for phoneme recognition". *Proc. Thirty-Seventh Southeastern Symposium on System Theory (SSST05)*, 2005, pp. 154 – 157.
- [5] Meyer, B.T., and Kollmeier, B.: "Complementarity of MFCC, PLP and Gabor features in the presence of speech-intrinsic variabilities". *Proc. Interspeech*, Brighton, 2009.
- [6] Hachkar, Z., Mounir, B., Farchi, A., and El Abbadi, J.: "Comparison of MFCC and PLP Parameterization in pattern recognition of Arabic Alphabet Speech", *Canadian Journal on Artificial Intelligence, Machine Learning & Pattern Recognition*, 2011.
- [7] Thiang., and Wijoyo, S.: "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot". *Proc. International Conference on Information and Electronics Engineering*, Singapore, 2011, pp. 179-183.
- [8] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Am*, pp 1738-1752, 1990.
- [9] H. Hermansky, N. Morgan, A. Bayya and P. Kohn, "Rasta PLP Speech Analysis," TR-91-069, Dec 1991.
- [10] L.A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *ACM'94*, March 1994, vol.37, no 3, pp.77-84.
- [11] L.A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no.3, pp. 338-353, June 1965.
- [12] L.A. Zadeh, "Making computers think like people," *IEEE Spectrum*, 8/1984, pp. 26-32.
- [13] <http://www.ferdinandpiette.com/blog/2011/08/la-logique-floue-interets-et-limites/>, accessed: November 2012.
- [14] B. Paris, J. Eynard, G. François, T. Talbert, A. Traore and F. Thiery, "Gestion des ressources énergétiques d'un bâtiment : contrôle flou," *IBPSA, France*, 2008.
- [15] http://www ldc.upenn.edu/Catalog/readme_files/timit.readme.html, accessed: November 2012.
- [16] A. Sadiq, R.O.H. Thami, M. Daoudi, J.P. Vandeborre, "Classification des Objets 3D Basée sur la Logique Floue," *Compression et Représentation des Signaux Audiovisuels (CORESA'2004)*, Lille, France, 25-26 mai 2004.
- [17] D.P.W.Ellis. PLP and RASTA (and MFCC, and inversion) in Matlab using `melfcc.m` and `inv_melfcc.m`. 143 <http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/> accessed: November 2012.
- [18] A. Strassberg, M. Molochnikov and R. Steinberg, "Recognition of plosives using Fuzzy-Logic," *Signal and Image processing Lab SIPL*, 2009.
- [19] H.K. Tripathy, B.K. Tripathy and P.K. Das, "A Knowledge based Approach Using Fuzzy Inference Rules for Vowel Recognition," *Journal of Convergence Information Technology*, vol. 3, no. 1, Mar. 2008.
- [20] M. Chetouani, B. Gas, J.L.Zarader and C.Chavy, "Neural Predictive Coding for Speech Discriminant Feature Extraction: The DFE-NPC," *European Symposium on Artificial Neural Networks ESANN*, pp. 275-280, Belgium, 24-26 April 2002.