



Recognition of vowels by cochlear implants using a fuzzy logic

Stéphane Gallégo, Emmanuel Perrin, Christian Berger-Vachon, Lionel Collet, Eric Truy

► To cite this version:

Stéphane Gallégo, Emmanuel Perrin, Christian Berger-Vachon, Lionel Collet, Eric Truy. Recognition of vowels by cochlear implants using a fuzzy logic. International IMSEE Symposium Fuzzy Systems and Neural Networks, pp.103-116, 1994. <hal-04407804>

HAL Id: hal-04407804

<https://hal.science/hal-04407804v1>

Submitted on 3 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC0 1.0 - Universal - International License

RECOGNITION OF VOWELS BY COCHLEAR IMPLANTS USING A FUZZY LOGIC

S.Gallego, E.Perrin, C.Berger Vachon, L.Collet, E.Truy

Hearing & Voice laboratory (URA-CNRS 1447)
ORL Clinic.
E.Herriot Hospital
69437 LYON CEDEX 03 (France)

ABSTRACT

In this work the authors have studied the application of a fuzzy logic for the modelling of human behaviour. The subject is the recognition of phonemes by patients wearing a cochlear implant. The acoustic signal is coded by a cochlear implant and acquired on a working station and fed into the hard disk of a desktop computer. Parameters are then extracted from the signal to perform an automatic recognition of French vowels.

Four vowels, embedded in short sentences, have been taken in order to construct two 48-item lists. First list was used for learning and second for recognition. Ten speakers collaborated in this experiment leading to about a thousand utterances.

The lists were played to patients fitted with a cochlear implant. For each patient and for each speaker a confusion matrix was established.

With parameters taken out of the electrical signal delivered by the implant, 255 recognition models were tested.

Two automatic recognitions, using the learning and the recognition lists, were done with each model and the corresponding confusion matrices were calculated and compared with the matrix given by the patient.

Recognitions were based on a fuzzy logic and on an Euclidian distance.

Results showed that fuzzy logic led to results more close to the patients' decision than Euclidian distance.

Then, useful parameters appear to depend strongly upon the patient and this must be considered for the setting of the electrodes.

1/ INTRODUCTION

Processing of acoustical signal by cochlear implant patients is an important issue mostly when the coding performed by the machine is considered [2,11,12].

Several assumptions must be made, ranging from phonetic considerations at the acoustical stage to message understanding at a higher level.

In this work, results obtained with a fuzzy logic are studied compared to performances given by patients wearing a cochlear implant. They are discussed regarding the corresponding features calculated with a more classical distance, an Euclidian metric in this case.

1.1: Normal hearing

Ear is the first stage of the auditory mechanism. It can be seen in three parts:

- Outer ear receives acoustic vibrations of air and conveys them to ear-drum.

- Middle ear performs a transformation of acoustical energy into mechanical energy, using ear-drum and ossicles.

- Inner ear is mostly constituted by the cochlea which performs a transformation of mechanical displacements into nervous influx which will be further processed by the brain to take out intensity, tone and frequency of vibrations.

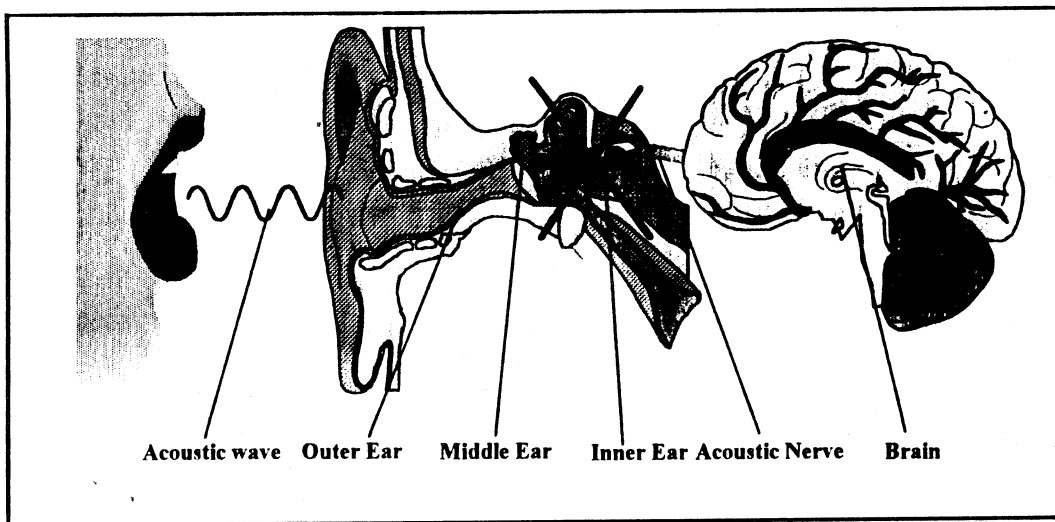


figure 1 : steps of hearing mechanism

1.2: Cochlear implant

The aim of a cochlear implant (CI) is to replace a non functioning cochlea [1,4]. CI provides a direct stimulation of the auditory nerve. A biphasic pulse is delivered on the ear at the ends of the auditory nerve bearing a tonotopic information.

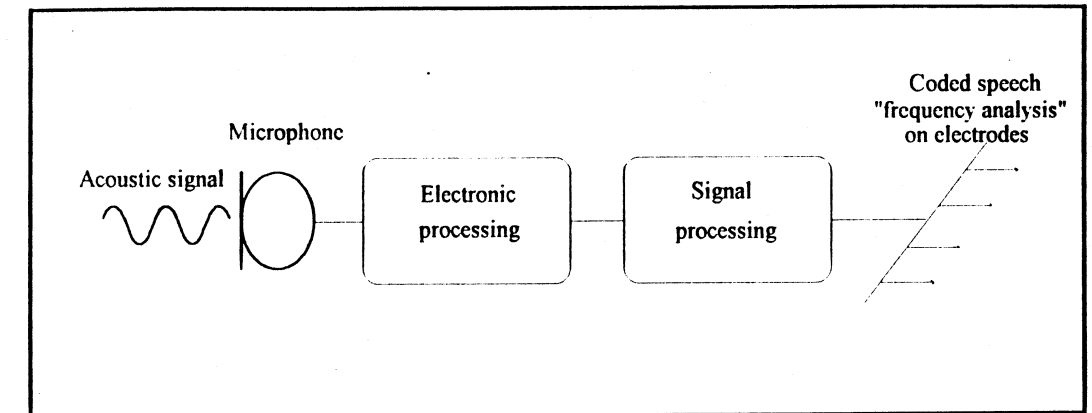


Figure 2 : Block diagram of a cochlear implant

CI is mainly made of two parts :

- An 'external part' where processing of the signal is made.
- An 'implanted part' distributing pulses on the nerve ends.

Principle of the CI used in this work (Nucleus System) is based on the classical properties of the voice [3]. As vowels only are taken, it is classically admitted that they can be represented by their first formants. Formants of the voice are the frequencies with highest energy.

In the cochlea duct, a 22-electrode array is inserted. Each electrode faces ends of the cochlear nerve. In order to simulate different frequencies the electrodes are stimulated according to the formants detected in the acoustic signal. As the automatic detection of formants is still an open question, only maximums situated in two frequency ranges are taken, giving the number of the electrode to stimulate.

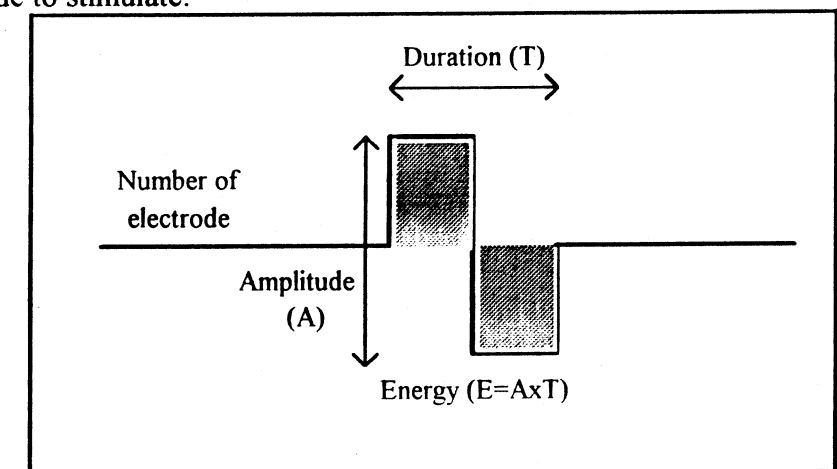


Figure 3 : Structure of a stimulating pulse

Each stimulation is a biphasic pulse. The surface (length X amplitude) of the pulse represents the energy attached to the formant. Stimulation rate is the voice pitch(fundamental frequency). This working is basically true when vowels are uttered; in other circumstances, coding is more complex and needs to be studied more thoroughly.

Four parameters can be taken out of each pulse:

- Number of the stimulated electrode (F).
- Amplitude (A) and duration (T) of the pulse.
- Surface (AxT), called energy (E) of the pulse.

As each formant is represented by two pulses (P1 and P2), eight parameters were used to characterise a vowel.

1.3: Research on electrical parameters

Direct analysis of electrical parameters has already been done by several research groups, mostly in Europe.

Dillier (University of Zurich) [5] compared different coding strategies (PES, CIS, CA, IP, ...) and the influence of the frequency range until 8 kHz. It turned out that the use of a 8 kHz bandwidth led to better results than a 4 kHz bandwidth.

Doering (Medical Faculty, Aachen, Germany) [6] showed that there is a slight difference according to the signal processing strategy used, regarding LPC (linear predictive coding) coding or the Nucleus principle for instance. This was done for vowels / ϕ /, /a/ and /o/ in German.

G.Collet and C.Berger Vachon (E.Herriot Hospital, Lyon) suggested that formants position is not the only cue used by patients for performing the distinction between vowels. This work was done with the Nucleus machine taken in the WSP (wearable speech processor) mode which uses the pitch and the first two-formants.

The present study is a continuation of this work with a larger set of parameters leading to 255 models according the combinatory theory.

M.Maillot and C.Berger Vachon [10] established that classical analyses of speech, using a formant detection done by ILS (interactive laboratory system) software, did not match perfectly the recognition performed with the maximums detected by the Nucleus system.

II/ MATERIAL AND METHODS:

2.1: Acquisition of the signal

The acoustic signal undergoes several stages before it is processed according to automatic recognition methods.

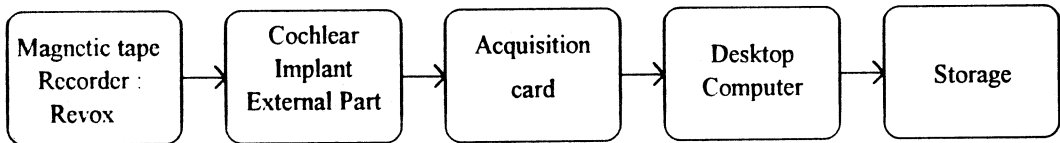


Figure 4 : Block diagram of the acquisition system.

Acoustical material had been recorded on a Revox tape recorder. It was then played at the input of a cochlear implant (Nucleus WSP III), which delivers pulses to an acquisition card. Parameters of the pulses were then detected in the signal and fed into a 386 IBM Desktop computer before being stored on hard and floppy disks attached to the system.

The core of the system, the acquisition card, works with a Texas TMS-25 which is controlled by specialised routines written in assembler language. The surrounding programs, written in Turbo Pascal, control the data flow from implant to storage volumes.

The eight parameters taken for the vowels are :

F1 F2 A1 A2 T1 T2 E1 E2.

where:

F indicates the electrode number attached to a formant,

A is the amplitude of the pulse,

T is the duration of the pulse,

E is the energy.

1 and 2 give the number of the formants.

It is not necessary to take the eight parameters to perform a recognition ; according to the phonetic theory, only F1 and F2 should be important for a distinction between vowels. Delattre's triangle (figure 5) shows this situation.

In order to establish the importance of the parameters, all the possible combinations of the eight parameters were studied, leading to 255 possibilities. Obviously, the '00000000' combination meaning that no parameter is taken has been excluded.

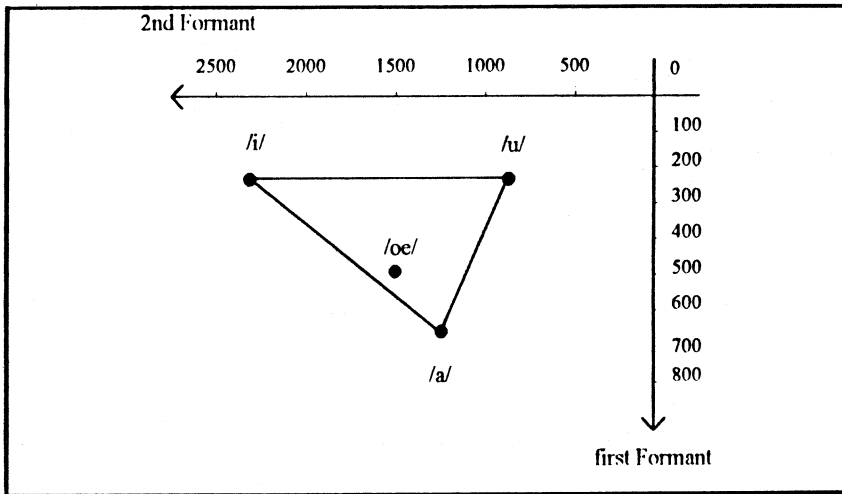


Figure 5 : Vowels represented in the Delattre's triangle

2.2: Acoustical material

Four oral vowels were considered in this work.

They are /i/, /u/, /a/ and /ə/ written in phonetic notation.

Each vowel was randomly repeated 12 times in two 48-item lists.

Each vowel was embedded in a sentence:

‘ c’ est /v/ ça’
 (‘it is /v/ that’)

where v stands for the vowel.

First list was the training list and second was the recognition list.

Ten speakers (five males, five females) took part in this experiment, and uttered, each one, two lists involving 980 vowels.

Four patients collaborated with this work ; each 980-vowel set was analysed by the CI of each patient with its own setting. Altogether 4,000 vowels with their own parameters were recorded on the mass storage volumes.

2.3: Recognition with the patient

Each patient was asked to perform a distinction of the vowels, for each speaker. He had to establish ten confusion matrices. He first listened to the training list, in order to adapt his understanding possibilities. Then, for each utterance, he had to give his best choice for the four vowels chosen for the experiment.

Out of his responses, a confusion matrix was automatically established by the computer. Ten confusion matrices were calculated for each patient (one for each speaker).

2.4: Automatic recognition

Two strategies were studied, the fuzzy logic [8,9] and the Euclidian distance [7].

2.4.1: Fuzzy logic

Let us suppose that k parameters are taken in the model.

For each parameter a table was constructed with the training list, giving for each vowel and each value of the parameter, the number of times the value was observed.

The parameters range was divided into 32 same length intervals, so that, a number between 1 and 32 was possible for each vowel.

At the decision stage, a value x_i ($i=1,...,k$) was observed for the i th parameter.

The system took the following decision :

- a score was attached to each vowel; it was read on the x^{th} line of the table.

- all the scores were summed up for each vowel, on the k parameters.

- the utterance was then classified to the vowel with the highest score.

A confusion matrix was then constructed with this recognition.

As 255 combinations of the k parameters can be considered , 255 confusion matrices were established for each speaker, and for each patient, leading to more than 10,000 matrices.

Table I shows the establishment of the vowels scores for a parameter.

Classes Value for 1 parameter	/a/	/i/	/u/	/e/
1	1a	1i	1u	1e
x	x _a	x _i	x _u	x _e
32	32a	32i	32u	32e

Table I : Scores of the vowels for a parameter ; x was observed.

2.4.2 Euclidian distance:

Let us suppose, once more, that k parameters are taken on the model. The recognition will be considered in a k -dimensional space.

For each vowel and for each parameter a value ranging from 1 to 32 is observed. Thus, each utterance leads to a point in this k -dimensional space.

With the learning list, 4 centres of gravity were established for the vowels.

An utterance in the recognition list gives a vector x to be classified ; this vector will be classified to the nearest centre of gravity, according to the Euclidian formula:

$$D^2(x, v) = \sum_{i=1}^n (x_i - v_i)^2$$

where:

x is the sample to be classified,

v is the vowel (class),

i is the i^{th} dimension (i^{th} parameter),

k is the dimension of the space,

and $D(x, v)$ is the distance of utterance x to class v .

x is classified to the vowel giving the shortest distance. Again, more than 10,000 matrices were constructed with the strategy.

Figure 6 shows this decision process.

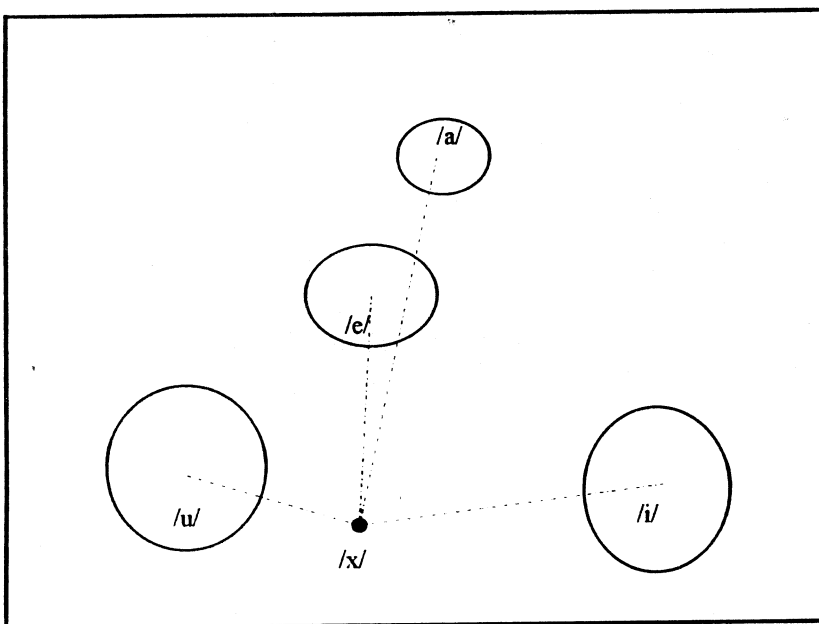


Figure 6 : classification according to the Euclidian distance

2.5: Evaluation of the importance of the parameters

As the aim of the work is to simulate the patients' recognition, a Hamming distance was calculated between the matrices established with the models and the patients' performances.

The following formula was used:

$$H(M, P) = \sum_{i=1}^4 \sum_{j=1}^4 |m_{ij} - p_{ij}|$$

where :

M is for the model and P for the patient.

m_{ij} is the recognition in cell(i,j) of the confusion matrix given by the model, and p_{ij} is the corresponding value obtained with the patient.

The absolute value of the difference was taken.

For each comparison, the following indications must be given:

- description of the model.
- name of the patient.
- name of the speaker.
- protocol (fuzzy or Euclidian).

An example of this calculation is given in figure 7

Mode F2F1, patient CO, speaker Ga.

Computer.(Model)

CAL	A	I	U	E
A	6	0	0	6
I	0	12	0	0
U	0	0	11	1
E	0	0	0	12

Patient

PAT	A	I	U	E
A	5	0	4	3
I	0	10	1	1
U	0	0	12	0
E	1	0	1	10

Error=Hamming distance: 37.5

Figure 7 : calculation of the Hamming distance between two confusion matrices.

III/ RESULTS AND DISCUSSION

3.1: Influence of the fuzzy logic

Distances between matrices have been calculated (table II) using Euclidian distance and fuzzy logic.

Results are given for each patient and averaged over all the speakers.

It can be seen that patients' recognition is closer to fuzzy results than those obtained with Euclidian distance.

Average of Hamming distance is 85 with the fuzzy logic and 105 with Euclidian distance ; results are always better when fuzzy representation is used.

This trend is seen again (table III) on the overall results. Difference seem to be bigger with females' voices than with males' ones, but this has to be proved to be significant. Also, F2 turned out to be the best parameter, but the energy of the formants cannot be ignored, mostly when fuzzy logic is considered.

1 first parameter **2** second parameter **3** third parameter

Patient	Method	F2	A2	E2	T2	F1	A1	E1	T1	Σ
AM	Euclidian	95	109	106	97	102	125	106	100	105
BA	Euclidian	116	116	113	113	114	120	114	113	115
CO	Euclidian	76	100	92	100	92	113	96	100	96
LA	Euclidian	105	108	105	102	105	115	105	102	106

Patient	Method	F2	A2	E2	T2	F1	A1	E1	T1	Σ
AM	Fuzzy	79	83	78	79	76	82	78	79	80
BA	Fuzzy	94	94	91	93	98	94	91	93	94
CO	Fuzzy	76	81	82	84	77	78	81	84	80
LA	Fuzzy	90	88	87	85	87	88	87	84	87

Table II : Mean distances between models and patients' recognition for both fuzzy and Euclidian methods

1 first parameter **2** second parameter **3** third parameter

Speakers	Method	F2	A2	E2	T2	F1	A1	E1	T1	Σ
Female	Euclidian	95	107	101	100	100	116	103	101	103
	Fuzzy	80	84	80	81	79	83	80	80	81
Male	Euclidian	101	110	107	106	106	121	108	107	108
	Fuzzy	89	90	89	90	90	89	89	90	90
Total	Euclidian	98	108	104	103	103	118	105	104	106
	Fuzzy	85	87	84	85	85	86	84	85	85

Table III : Overall comparisons between methods and patients' recognition

3.2: Parameters influence.

Best parameters were calculated, single, pairs and triplets, for Euclidian and fuzzy methods . (table IV)

Sets differ with the patient , and the energy factor cannot be ignored, mostly for the first formant which is predominant. This fact emphasizes the importance of the setting (not only the threshold) of the electrodes.

Difference between patients should be more deeply studied, mostly with the consideration of patients' history and settings of electrodes.

Furthermore, if a transformation of the signal is made by a computer, before delivering pulses to the patient, results given by the theory must be carefully considered prior doing any work that can be done with the patients.

Classical results such as :

-best single : F2.

-best pair : F1F2.

-best triplet : F1F2 (A1 or A2)

were observed with patient CO who is a star patient (very good use of CI) in our clinic.

Six best parameters:

Patient	1st	2nd	3rd	4th	5th	6th
AM	T2	F2	T1	F1	A2	A1
BA	T1	T2	A2	F1	F2	A1
CO	F2	A2	F1	A1	T1	T2
LA	T2	F1	T1	A1	A2	F2

Six best pairs:

Patient	1st	2nd	3rd	4th	5th	6th
AM	T2T1	T2F1	F2T1	F2T2	F1F2	A2T2
BA	A2T1	A2T2	A2A1	A1T1	T2A1	T1T2
CO	F2F1	F2A1	F2A2	F2T1	F2T2	F1A1
LA	A2T2	T2A1	A1T1	T2T1	A2T1	F1A1

Six best triplets:

Patient	1st	2nd	3rd	4th	5th	6th
AM	T2A1T1	A2T2T1	T2F1T1	A2T2F1	T2F1A1	F2T2T1
BA	A2A1T1	T2A1T1	A2T2T1	A2T2A1	A2F1T1	A2F1A1
CO	F2A2F1	F2F1A1	F2A2A1	F2T2F1	F2E2F1	F2F1T1
LA	A2T2A1	F1A1T1	A2T2F1	T2F1A1	A2F1T1	T2A1T1

Table IV(a)

Six best parameters:

Patient	1st	2nd	3rd	4th	5th	6th
AM	E1	T1	E2	F2	F1	T2
BA	E1	E2	T1	T2	F2	A2
CO	F2	A2	E1	A1	E2	F1
LA	T2	T1	E2	E1	F2	A1

Six best pairs:

Patient	1st	2nd	3rd	4th	5th	6th
AM	F1E1	E2F1	E1T1	E2T1	F2F1	T2F1
BA	E1T1	A1E1	E2A1	A2E2	E2T1	T2E1
CO	F2F1	F2A1	F2E1	F2A2	A2A1	A1E1
LA	E1T1	A2T2	E2T2	A2T1	E2T1	E2F1

Six best triplets:

Patient	1st	2nd	3rd	4th	5th	6th
AM	F2F1E1	F2E2F1	E2T2F1	A2E2F1	T2F1E1	E2F1A1
BA	A2E2T2	T2E1T1	A2T2E1	A2E2E1	F2A2E1	A2T2T1
CO	F2F1A1	F2A2F1	F2E2F1	F2F1E1	F2A1E1	F2E1T1
LA	A2T2T1	E2T2T1	T2E1T1	A2E1T1	A2T2E1	A2E2T1

Table IV(b)

Table VI : Best sets of parameters to approximate the patients recognition
(a : Euclidian, b : Fuzzy)

IV/ CONCLUSION

Fuzzy logic introduced an interesting means to approximate human performances, compared to classical Euclidian distances. Further work will be done in order to improve this first step.

Second formant turned out to be the best parameter able to perform the distinction between studied vowels. This was mostly seen with a star patient and when Euclidian distance is taken.

Amplitude of the formants proved to be an important factor for the recognition.

Classical coding (F1F2) turned out to be the best pair only for the star patient. Generally speaking, it can be said that each patient has its own strategy. Analysis of this point will be an important issue of the research in the field of cochlear implant.

ACKNOWLEDGEMENTS:

Authors wish to acknowledge the supports of institutions and people who back them in their research: the Civil Hospitals of Lyon, the Rhône-Alpes Region, the French Council for Research (CNRS), API company, the University of Lyon, and Professor A.Morgon from the Edouard Herriot hospital.

REFERENCES

- [1] Beliaeff (M), Dubus (P), Leveau (J.M), Repetto (J.C), Vincent (P)
Sound processing and Stimulation coding of the Digisonic DX10 15-channel Cochlear Implant.
Advances in cochlear implant, Edited Hochmair (University of Innsbruck Austria). (1994) : 198-203.
- [2] Berger Vachon (C), Collet (L), Djedou (B), Morgon (A) :
Model for understanding the influence of some parameters in cochlear implantation.
Ann. Otol. Rhinol. Laryngol (1992). 101 : 42-45
- [3] Blamey (P.J), Clark (G.M) :
Place coding of vowels formants for cochlear implant patients.
J.Acoust.Soc.Am (1990) 88(2) : 667-673
- [4] Clark (G.M) et al : The university of Melbourne - Nucleus Multi-electrode Cochlear Implant.
In Advances in OtoRhinoLaryngologie (1987) vol 38. Ed. Karger.
- [5] Dillier (N), WaiKong (L), Hans (B)
A High Spectral Transmission coding strategy for multi-electrode cochlear implant.
Advances in cochlear implant, Edited Hochmair (University of Innsbruck Austria). (1994) : 152-157.
- [6] Doering (W.H.), Schneider (L)
Electrical vs. acoustical speech patterns of German phonemes using the Nucleus CI-system.
Cochlear Implant. Acquisition and controversies. Edited Bernard Fraysse (Toulouse) (1989) : 243-253.

[7] Diday (E), Lemaire (J), Poujet (J), Testu (F)

Eléments d'analyse de données.

Ed. Dunod (1982).

[8] Kaufman (A).

Introduction à la théorie des sous-ensembles flous.(I,III and examples.).

Ed. Masson (1975).

[9] Kaufman (A).

Les logiques Humaines et Artificielles.

Ed. Hermes (1988).

[10] Mailliot (M), Berger Vachon (C), Truy (E), Morgon (A).

Overlap of Vowel Clusters Through a Nucleus Cochlear Implant.

Advances in cochlear implant, Edited Hochmair (University of Innsbruck Austria). (1994) : 260-265.

[11] Wilson (B.S), Finley (C.C), Farmer (J.C), Lawson (D.T) :

Comparative studies of speech processing strategies for cochlear implants.

Laryngoscope (1988) 98 : 1039-1077.

[12] Wilson (B.S), Lawson (D.T), Woford (R), Eddington (D.K) :

Coding strategies for multichannel cochlear prostheses.

Am.J.of Otol. (1991) 12(Suppl) : 56-61.