First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

AUTOMATIC SEGMENTATION : WHY AND WHAT SEGMENTS ?

Mario Rossi

Institut de Phonétique (CNRS), Aix-en-Provence, France

ABSTRACT

I present and discuss the SAPHO (Segmentation by Acoustico-Phonetic knowledge) model implemented in the Awk language under the Unix system on a MASSCOMP computer. The system is devised as a speaker independent ASS (automatic speech segmentation), by previous recognition of the phonetic articulation manner. In all the ASR systems the phonetic knowledge is at least implicitely used. It has to be referred to explicitely. Following the Level Building procedure SAPHO supplies a hierarchized set of acoustic properties and segments, and phonetic properties and segments which fit the phonetic parsing of the acoustic wave. The amenability of this system is entailed by its modularity which allows a possible further architecture as distributed tasks.

I - INTRODUCTION

Segmentation procedures in Automatic Speech Recognition (ASR) have been and still are used by some and refused by others. Two questions arise immediately :

1.1 Given the lack of bijectivity between acoustic and linguistic levels, can automatic segmentation be considered justifiable, and, if so, do segmentation procedures yield useful information or, on the contrary, reach a deadlock ?

1.2 If the answer to the first question is that this segmentation is justifiable and reliable, what kind of items can this segmentation isolate (segments, phones, phonemes ?) in other words at which levels should segmentation be carried out ?

The answers to the above questions might give new insight into how we should proceed with the design of an Acoustic Phonetic Decoding module (A P D). We will provide an answer to these questions by presenting and discussing a procedure for the recognition of major phonetic classes: the SAPHO model (Segmentation of Acoustics by Phonetics).

II - THEORETICAL REQUIREMENTS

2.1 Many ASR systems have been or are still based on automatic segmentation (AS) procedures which are supposed to retrieve phonemic units directly from the acoustic parameters. These misconceived AS methods entail fatal and irreversible errors, and they have been rightfully criticized in the last decade as not congruent with either the linguistic theory or with language processing.

2.2 It is well known that bijectivity does not hold between the abstract phonemic representation and the concrete events, namely the acoustic and the articulatory ones. In particular, the phonemic units are defined by their discreteness which is not reflected at all in the acoustic wave. The acoustic wave is phonemically ambiguous. The question of the relationship between the phonetic and the acoustic level can also be considered. The phonetic level has been confused with the concrete acoustic and articulatory levels. In fact, it has to be distinguished from these last two as an abstract intermediate level of representation between the concrete speech events and the phonemic representation. Bijectivity also fails to hold between the phonetic and the phonemic strata. The retrieval of the phonemic units from the phonetic level requires linguistic knowledge about how they are realized at the phonetic lower level.

In the phonetic representation the units are still discrete. But because of the different rules of organization, i.e. the different logic of the acoustic and the phonetic levels, the discreteness of the phonetic units cannot be transfered into the acoustic continuum. Yet the acoustic signal contains discontinuities, but what are their phonetic values? In order to identify some discontinuities as a transition between a stop and a vowel, for example, we have to know the phonetic structure of the stops and vowels and the transformation rules from the phonetic to the articulatory and acoustic strata. Hence, the retrievability of the higher units from acoustic events is not a straightforward endeavour; it takes a complex phonetic and phonemic knowledge of speech coding. This requirement is emphazised by many scholars either phoneticians or engineers [2, 3, 4, 5, 6]. "It is our belief that acoustic-phonetic knowledge representation is a major roadblock in the design of advanced speech recognition systems that are meant to approach human performance" [6].

This assertion seems obvious. It is not. In fact, some scholars now fascinated by the results of Markov models and the neural networks believe and claim that knowledge based systems are out of date and must be abandoned on account of their inefficiency in comparison with the new models. It would be a pity to go back to the years where "most speech recognizers were based on an ignorance model of speech" [3].

2. 3 The question now is whether the information available in the empirical facts contains some reliable cues to the separation of the higher units ; in other words, are the higher units retrievable by the segmentation procedure?

Research on coarticulation demonstrates that coarticulatory processes mainly involve parameters and cues aimed at encoding place of articulation. Conversely, parameters specialized in coding the manner of articulation greatly escape the blurring effect of coarticulation. The relative resistance of manner of articulation parameters thus allows us to segment the so-called acoustic continuum. But we have to bear in mind that the correct processing and interpretation of these parameters and segments must be guided by the linguistic knowledge. The ultimate phonemic units are not at all recoverable without an explicit top down strategy which involves phonemic and lexical constraints.

III. METHOD

On the basis of what I said earlier, I shall present here a model of automatic macroclass (broad class) recognition, more precisely of acoustic segments built on phonetic knowledge. These segments are congruent with the structure of the phonetic units ; and are obtained by an indirect segmentation procedure on the basis of the sample properties.

This model is intended as part of a speaker independent ASR system. It is processed by a modular program written in the Awk language [1] under UNIX on a Masscomp computer.

3.1. Parameters and cues

The data files contain the spectra obtained by a 15 channel vocoder. The percentage of the energy in each channel gives the band contribution to the phonetic information .

3.1.1. The available primitive parameters calculated on each sample of10 ms are : zero crossing density (PZ) ; total energy (EN) ; first derivative of EN (DE); first spectral derivative calculated on the contributions (DK) ; low frequency energy (BF); high frequency energy (HF). The normalized cognates of EN, DE, BF and HF are respectivaly EN%, DE%, BF% and HF%.

3.1.2. The cues are tests on the relevant spectral energy distribution. The relevance of the tests is deduced from the phonetic knowledge. The openness, acuteness and flatness features are evaluated by 2 procedures : TRAITS2 and TRAITS3. The first is based on the binary response, the second on the fuzzy response to the tests. In each of the two procedures, the content of the cues is quite different. In the first binary based procedure, for each assumed feature, the cues are hierarchized and a coefficient assigned to each of them. The combination of the cues gives the profile and the value of the assumed feature. In the second fuzzy based procedure, each cue looks for the maximum energy of some bands and calculates the amount of the prominent energy (PE).

3.2. Paramater and Cue Processing.

In addition to the subroutines aimed at the parameter calculation in the PARAM processor, 14 independent processing modules are used for macroclass recognition. The first 7 modules mainly process the acoustic information driven by the phonetic knowledge. They are hierachized and some are called recursively by an acoustic phonetic supervisor called MODE. The other modules are called together with MODE by the phonetic supervisor, SUPERMODE.

These last modules in SUPERMODE are predominantly knowledge based processors which execute an expectation driven activity [2]. At each step, the intermediate results are preserved ; so that the history of the results given by the processors are available in immediate memory for further interpretation. The SAPHO module organization is represented in Fig.1 (where VOC = vocalic class ; MAX/MIN = rising/falling parts of VOC; the bracketted items represent the module headings).

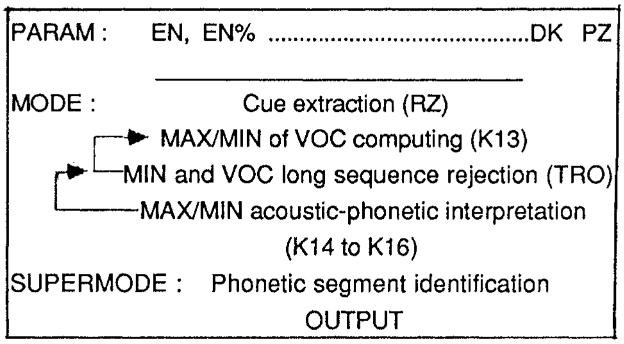

Fig.1

3.2.1. There is evidence for robust features and islands of reliability in the speech signal [3]. So the first step processed by the RZ module looks for gross acoustic classes : CSB, FRIC and VOC. CSB is identified in order to be congruent with the true consonants (obstruents). The FRIC is identified if some conditions are met comparing PZ, EN and derivatives. Later FRIC will be interpreted as either a fricative or a burst. The VOC will label the samples not recognized as either CSB or FRIC. The choice of a label is never naive, thus VOC remembers the vocalic feature. We know that in some way the voiced consonants are getting the properties of the vowels, let us say are "vocalized". This fact explains the choice of the VOC label which cover the cues of either the vowels or the "vocalized" consonants.

At this stage, it is worth bearing in mind that the resulting segmentation is not based on a previous arbitrary choice of some threshold, but on specific acoustic parameters embodied in the phonetic knowledge.

3.2.2. In the second step, the KL3 module looks for the increasing/decreasing values of EN. It codes rising and falling slopes. *min1* and *min2* are degrees of falling slope distinguished by a comparison between the EN and DE values.

Rising slopes contain the main energy maxima ; so together with the *min1* pieces they indicate the presence of a vowel. *min2* labels have to be reinterpreted as either the decreasing part of a vowel or a consonant. The KL3 module recursively processes the *min2* pieces, previously indentified as exceeding the mean length of the vowels.

3.2.3. In the following two steps the MODE supervisor calls the KL4, KL5 and KL6 modules in order to give a phonetic interpretation to the rising slopes and to the *min1* and *min2* portions. KL4 and KL5 are phonetic modules whose decision depends on context sensitive rules. The increasing slopes together with the *min1* labels are recognized as vowels and labelled respectively VOY (rising part) and VOX (falling part). The *min2* portions are interpreted as either liquids (LIQ) or consonants (CS) or the second half of vowels (VOX). The liquids could later be recognized as lateral or nasal consonants. The CS label refers to weaker consonants than those implied by the CSB label ; they can be interpreted later as CV (vocalic consonants) or as CSV (voiced obstruents).

3.2.4. The next step looks for possible missed consonants with the KL6, TROLO and KL3 modules. As in the other modules aimed at the splitting of a sequence, here the decision is not based on threshold criteria but on the relationship among several parameters. The consonants identified at the output of these last modules are immediately labelled as CV.

The output of the MODE supervisor is a sequence of VOY+ VOX (vowels), CV, CS, CSV, CSB and FRIC.

3.3 . Decisions by the phonetic supervisor

The phonetic supervisor SUPERMODE takes the symbolic output of MODE, and the results of the TRAITS2 and TRAITS3 processors. After KL8 has looked for the possible strong transitions (tr) comparing some cues with the transitions on the n-ary cues of the openness features (TRAITS2), the KL9, KL10 and KL11 processors try to identify the stronger stop bursts, the fine manner of articulation and the weaker stop bursts respectively.

We know that the stronger bursts for example of (k) and (t) before front vowels are characterized by high PZ values. Hence by the presence of FRIC'S. So the correct identification of FRIC'S as bursts assumes a great importance too in hypothesizing the place of articulation of the consonants. But identifying a burst under the FRIC's cannot be based on a PZ threshold ; it would give unreliable arbitrary results. The phonetic knowledge tells us there is a special relationship between EN and DK when FRIC occurs as a burst. The output`of KL9 is EXPL labels which stand for strong bursts. Later the EXPL labels will be used as a request to hypothesize the acute feature of the plosive and the subsequent vowel.

The KL10 module takes the CSB, FRIC and CS labels for a further interpretation, using the state of the n-ary cues of the openness feature, a comparison between EN and these same cues and a matching of the EN with the binary cues of the acuteness feature.

Finally the KL11 module using contextual knowledge looks for the weaker plosive bursts. Two degrees of plosive bursts are found below the stronger EXPL : EXP2 and expl. For example, one of the rules of KL11 is :

```
{ (if ((avant =="SIL")  && ($ 14 == "Ct-VX"))
        { $ 14 = "EXP2" ; print} }
```

IV. DISCUSSION

4.1. The strategy underlying the SAPHO model is a parsing of the acoustic and phonetic classes from the fundamental classes, vowels and consonants, to the subphonetic segments through the major properties.

The parsing of each of the main classes is structured as a set of ordered context free, context sensitive and construction rules.

At the phonetic level of representation we have larger units such as vowels, fricatives, plosives, and so on... matching the phonemic units. The SAPHO output provides some of these units, but it goes further. It supplies the units as realized, hence as perceived and noted in a narrow transcription ; it also provides the internal structure of some consonants.

4.2. Now it makes sense to set up the tokens to be obtained at each stage of an acoustic phonetic driven system. At the first step the tokens are the AP (acoustic properties); from the homogeneous AP sequences are deduced acoustic segments. Further parsing implies an expectation driven activity including more and more phonetic knowledge. At this step we are faced with phonetic segments which have to be processed by a top down phonetic driven procedure to find the phonetic units which fit the highest phonemic units. Although the first steps provide acoustic segments we must insist on the absolute necessity for the phonetic knowledge to drive the extraction of the acoustic properties for these to act as facts driving the computational processes ; so the hypothesized acoustic segments will be congruent with the higher phonetic units. The assumed hierarchy from the acoustic properties through the acoustic segments to the phonetic units is obtained in SAPHO with a strategy starting with the largest sequences. This strategy which reaches the phonetic segment stage by a splitting of the original sequences is justified by the requirement of reducing the uncertainity at the beginning of the process as much as possible ; starting with robust cues and islands of reliability and avoiding the location of strict boundaries between segments.

This strategy may be considered a translation of the Level Building algorithm from the isolation of words in a sentence to that of phonetic units in the speech wave [4].

V. RESULTS

A provisional assessment was conducted on 60 words pronounced by six speakers (three females and three males). A further assessment is being run on a corpus of isolated words and sentences from BDSONS (The French Acoustic Data Base).

The provisional assessment gives excellent encouraging results with a score of 97% : 1% of spurious segments and 2% of missed CV consonants on a total of 310 consonants. SAPHO cannot properly be assessed yet on the processing of clusters, because of their scarcity in the corpus.

VI. CONCLUSION

The SAPHO system is devised as a speaker independent ASS (automatic speech segmentation), by a previous recognition of the phonetic articulation manner. Phonemic units cannot be directly built from the acoustic signal and are not available at the output of SAPHO. In fact the major phonemic units in order to be isolated need all the phonetic information on the phonetic manner and place of articulation and the knowledge of phonological and lexical constraints. SAPHO demonstrates how the acoustic parameters ought to be phonetically driven in order to obtain acoustic properties and segments congruent with the internal structure of the phonetic units. The processors are concieved either as data driven with numeric computation or as expectation driven activities with symbolic computation. The recursivity in the acoustic and the phonetic supervisors at each step of the parsing ensures the likelihood of the decisions. The suitability and the reliability of SAPHO are corraborated by the accuracy of the results.

REFERENCES

[1] Aho, A.v., Kerninghan, B. W. and Weinberger, P.J., The AWK Programming Language, Addison-Wesley, Amsterdam, 1988.

[2] De Mori, R. and Laface, P., On the Use of Phonetic Knowledge for Automatic Speech Recognition, in De Mori and Ching Suen 1985, pp. 569-592, 1985.

[3] Lea, W.A., Selecting, Designing and Using Practical Speech Recognizers, in Haton 1982, pp. 331-368, 1982.

[4] Meyers, C.S. and Rabiner, L.R. , A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition, *I.E.E.E. Trans. Acoustics Speech and Signal Processing,* vol. 29, pp. 284-297, 1981.

[5] Ohala, J. , Linguistics and Automatic Processing of Speech, in De Mori and Ching Suen 1985, pp. 447-476, 1985.

[6] Zue, V. , Acoustic Phonetic Knowledge Representation : implications from Spectrogram Reading Experiments, in Haton 1982, pp. 101-120, 1982.

N	EN%	EN	DE	BF	BF%	HF	HF%	ZER	DK	PZ	RZ	K3	TRO	K3	TRO	K3	K6	TRO	K3	K4	K5	K6	K8	MOD
1	0	0	0	0	0	0	0	-60	0	0.00	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	OC+VX
2	0	0	0	0	0	0	0	-60	77	0.00	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	OC+VX
3	0	4	0	2	0	0	0	-36	93	0.00	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
4	0	5	0	4	0	0	0	-40	40	0.00	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
5	0	4	0	3	0	0	0	-41	54	0.02	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
6	0	2	0	1	0	0	0	-53	59	0.04	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
7	0	4	0	3	0	0	0	-51	31	0.03	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
8	0	4	0	3	0	0	0	-41	67	0.03	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
9	0	7	0	6	0	0	0	-38	67	0.03	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	OC+VX
10	0	13	0	11	1	0	0	-27	46	0.04	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	OC+VX
11	0	13	0	11	1	0	0	-27	29	0.03	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	OC+VX
12	0	9	0	7	0	0	0	-31	13	0.04	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	OC+VX
13	0	8	0	6	0	0	0	-32	48	0.04	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
14	0	7	0	5	0	0	0	-33	39	0.07	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
15	1	25	1	16	2	3	1	35	40	0.09	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
16	11	179	10	109	15	8	5	183	52	0.11	VOC	?	?	?	?	?	?	?	?	?	?	?	?	expl
17	24	379	13	189	26	45	29	753	39	0.10	VOC	24	24	24	24	24	VOY	VOY	VOY	VOY	VOY	VOY	VOY	-
18	48	737	24	456	63	95	62	1464	42	0.12	VOC	48	48	48	48	48	VOY	VOY	VOY	VOY	VOY	VOY	VOY	-
19	55	851	7	519	72	110	71	1697	42	0.12	VOC	55	55	55	55	55	VOY	VOY	VOY	VOY	VOY	VOY	VOY	-
20	88	1353	33	676	94	148	96	2691	53	0.14	VOC	88	88	88	88	88	VOY	VOY	VOY	VOY	VOY	VOY	VOY	-
21	100	1532	12	658	92	153	100	3044	25	0.12	VOC	100	100	100	100	100	VOY	VOY	VOY	VOY	VOY	VOY	?	-
22	69	1062	-31	679	95	74	48	2104	61	0.13	VOC	min1	min1	min1	min1	min1	VOY	VOY	VOY	VOY	VOY	VOY	VOY	-
23	56	872	-13	662	92	34	22	1719	39	0.14	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CV
24	56	870	0	713	100	52	33	1710	21	0.06	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CV
25	57	888	1	683	95	106	69	1746	19	0.05	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CV
26	47	733	-10	542	76	117	76	1441	27	0.06	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CV
27	27	414	-20	318	44	53	34	803	29	0.05	VOC	min2	min2	min2	min2	min2	+CS	+CS	+CS	+CS	+CS	+CS	+CS	+CS
28	9	150	-18	121	16	13	8	169	10	0.05	VOC	min2	min2	min2	min2	min2	+CS	+CS	+CS	+CS	+CS	+CS	+CS	+CS
29	4	70	-5	60	8	4	2	85	23	0.05	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
30	3	48	-1	39	5	2	1	67	30	0.04	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
31	2	45	-1	38	5	1	0	47	20	0.05	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
32	3	52	1	43	6	2	1	86	36	0.05	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
33	3	59	0	35	4	8	5	117	41	0.06	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
34	4	63	1	35	4	10	6	105	32	0.05	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
35	5	77	1	32	4	24	15	123	44	0.06	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
36	9	148	4	78	10	44	28	170	46	0.05	VOC	min2	min2	min2	min2	min2	CS	CS	CS	CS	CS	CS	CS	CSV
37	14	217	5	117	16	62	40	225	13	0.06	VOC	14	14	14	14	14	VOY	VOC	?	?	?	?	?	?
38	16	246	2	167	23	49	32	477	35	0.06	VOC	15	15	15	15	15	VOY	VOC	84	VOY	VOY	VOY	VOY	-
39	17	266	1	186	26	42	27	512	15	0.05	VOC	16	16	16	16	16	VOY	VOC	85	VOY	VOY	VOY	VOY	-
40	18	280	1	190	26	50	32	535	13	0.06	VOC	17	17	17	17	17	VOY	VOC	86	VOY	VOY	VOY	VOY	-
41	19	294	1	191	26	73	47	563	12	0.05	VOC	18	18	18	18	18	VOY	VOC	87	VOY	VOY	VOY	VOY	-
42	16	247	-3	165	23	56	36	469	15	0.05	VOC	19	19	19	19	19	VOY	VOC	min1	VOX	VOX	VOX	VOX	-
43	14	227	-2	138	19	68	44	429	33	0.05	VOC	20	20	20	20	20	VOY	VOC	min1	VOX	VOX	VOX	VOX	-
44	12	195	-2	109	15	54	35	175	17	0.05	VOC	21	21	21	21	21	VOY	VOC	min2	CS	CS	CV	CV	CV
45	10	168	-2	97	13	53	34	183	19	0.05	VOC	22	22	22	22	22	VOY	VOC	min2	CS	CS	CV	CV	CV
46	9	152	-1	94	13	42	27	177	38	0.05	VOC	23	23	23	23	23	VOY	VOC	min2	CS	CS	CV	CV	CV
47	8	133	-1	99	13	18	11	133	46	0.05	VOC	24	24	24	24	24	VOY	VOC	min2	CS	CS	CV	CV	CV
48	8	134	0	108	15	8	5	109	39	0.05	VOC	25	25	25	25	25	VOY	VOC	min2	CS	CS	CV	CV	CV
49	8	129	0	95	13	5	3	124	40	0.05	VOC	26	26	26	26	26	VOY	VOC	min2	CS	CS	CV	CV	CV
50	6	101	-2	59	8	13	8	108	31	0.04	VOC	27	27	27	27	27	VOY	VOC	min2	CS	CS	CV	CV	CV
51	4	73	-2	35	4	16	10	87	30	0.05	VOC	28	28	28	28	28	VOY	VOC	min2	CS	CS	CV	CV	CV
52	2	34	-2	17	2	9	5	50	51	0.04	VOC	29	29	29	29	29	VOY	VOC	min2	CS	CS	CV	CV	CV
53	2	31	0	15	2	8	5	41	27	0.05	VOC	30	30	30	30	30	VOY	VOC	min2	CS	CS	CV	CV	CV
54	2	37	0	15	2	13	8	73	55	0.04	VOC	31	31	31	31	31	VOY	VOC	min2	CS	CS	CV	CV	CV
55	3	52	1	14	1	26	16	114	37	0.08	VOC	32	32	32	32	32	VOY	VOC	20	VOY	VOY	VOY	VOY	-
56	3	58	0	5	0	40	26	160	46	0.08	VOC	33	33	33	33	33	VOY	VOC	28	VOY	VOY	VOY	VOY	-
57	6	101	3	4	0	80	52	239	25	0.08	VOC	34	34	34	34	34	VOY	VOC	42	VOY	VOY	VOY	VOY	-
58	8	131	2	6	0	102	66	240	40	0.12	VOC	35	35	35	35	35	VOY	VOC	43	VOY	VOY	VOY	VOY	-
59	9	145	1	13	1	92	60	275	40	0.16	VOC	36	36	36	36	36	VOY	VOC	44	VOY	VOY	VOY	VOY	-
60	7	116	-2	19	2	55	35	212	59	0.08	VOC	37	37	37	37	37	VOY	VOC	min1	VOX	VOY	VOY	VOY	-
61	10	156	3	24	3	87	56	240	42	0.14	VOC	38	38	38	38	38	VOY	VOC	42	VOY	VOY	VOY	VOY	-
62	11	169	1	32	4	101	66	261	27	0.08	VOC	39	39	39	39	39	VOY	VOC	43	VOY	VOY	VOY	VOY	-
63	10	160	-1	27	3	105	68	253	22	0.06	VOC	40	40	40	40	40	VOY	VOC	44	VOY	VOY	VOY	VOY	-
64	9	153	-1	27	3	97	63	227	16	0.04	VOC	41	41	41	41	41	VOY	VOC	min1	VOX	VOX	VOX	VOX	-
65	7	108	-2	25	3	65	42	176	19	0.04	VOC	42	42	42	42	42	VOY	VOC	min2	VOX	VOX	VOX	VOX	-
66	4	75	-3	24	3	42	27	129	27	0.04	VOC	43	43	43	43	43	VOY	VOC	min2	VOX	VOX	VOX	VOX	-
67	3	57	-1	18	2	31	20	112	15	0.04	VOC	44	44	44	44	44	VOY	VOC	min2	VOX	VOX	VOX	VOX	-
68	2	41	-1	12	1	22	14	115	27	0.04	VOC	45	45	45	45	45	VOY	VOC	min2	VOX	VOX	VOX	VOX	-
69	1	27	-1	6	0	15	9	85	37	0.04	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
70	1	16	0	5	0	7	4	-9	42	0.04	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
71	0	12	-1	5	0	4	2	-3	49	0.06	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
72	0	8	0	5	0	2	1	-27	54	0.00	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
73	0	7	0	5	0	0	0	-28	40	0.00	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
74	0	5	0	4	0	0	0	-40	38	0.02	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
75	0	4	0	3	0	0	0	-36	40	0.02	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL
76	0	4	0	3	0	0	0	-46	23	0.02	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	CSB	SIL

ANNEXE : Automatic segmentation of the word *Badiner* [badine] female speaker. Speech samples (every 10 ms) are numbered in the first column.