MICROPHONHHES AS FUNDAMENTAL SEGMENTS OF SPEECH WAVE PRIMARY SEGMENTATION - AUTOMATIC SEARCHING FOR MICROPHONEMES

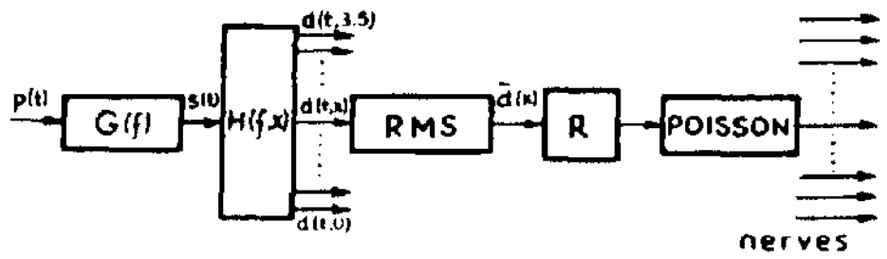
Mgr inz. Andrzej Dziurnikowski
Institute of Organization, Management and Control Sciences, >/arsaw, Poland

Abstract

This paper concerns the stage of acoustic analysis in speech recognition of Speech Understanding System SUSY; subsystem AKORD, Acoustic analysis in this system is based on dynamic spectral analysis of the speech wave. It corresponds with the function of the hearing organ expressed by an analog nodel of human ear. Acoustic analysis in the AKORD system uses an FFT algorithm but is not strictly based on accepting constant intervals of 10-20 ms; analysis does not cause inaccuracy at the level of parametrization because *of* a considerable dispersion of parameters of particular segments. A conclusion was made in the AKORD system that a properly conducted process of segmentation will help in solving some problems of speech wave analysis, particularly recognition and time compression. This paper attempts to choose the primary segment (microphonem) in the most optimum way and to Indicate the algorithm of automatic segmentation based on microphonemes as the dynamic segments.

Introduction

System AKORD was designed for research work concerning the analysis and synthesis of speech waves by means of a ZAM-41 computer (1,2,3) - during the initial phase of research that system was helpful in defining the most essential problems of speech wave analysis- It also made it possible to inculcate and test some algorithm models of speech analysis. However, before accepting a concrete algorithm model of speech analysis it was necessary to realize different functions of the hearing organ. Those functions are expressed by an analog model of the human ear (4). Generally, the track of a given sound in the ear may be presented as follows: the external ear, the ear-drum, the internal ear and its bones (malleus, incus, stirrup). The movements of the stirrup bone cause translocation of cochlea septum :
-its resiience is not constant through its whole lenght (4,5). That vibration is subsequently trans nitted through the *nerve* cells on the septum to ~30 000 auditory nerve fibrils. Figure 1 shows analog model of human ear presented by P.Kolers (4) . Although this mod^l is only approximate, a comparison of the characteristics obtained experimentally showed a considerable convergence (4). Ne accept the conclusion, with no deeper consideration of the ear model presented here, that the internal car is the centre of sound spectral analysis *S(t)* . Tts results are made average



P(t) -external auditory duct pressure, G(f) -linear block transforming pressure P(t)into voluminal translocation sCt)of the stirrup, H(f,x) - this function connects translocation s(t)with translocation of septum d(t,x)at the point x [cm] from the stirrup, RFIS - realises the mean-square function of input signal d(t,x)for short duration (~10 ras).

Figure 1

in time for the period Ts (~ 10ms).The model of the auditory organ presented here is a basis in this paper, although its imperfection is realised. But it is not a novelty. The majority of studies concerning the analysis of speech waves have been based *on the* spectral analysis *of* the examined signals (6). Some of them used the dynamic spectral analysis, yet they were strictly based on accepting constant intervals of 10-20 ms. Although .the acceptance of that constant interval made it possible to realise some functions of analysis more simply, it caused some inaccuracy at the level of parametrization due to a considerable dispersion of parameters of particular segments(12). This fact resulted in substantial difficulties connected with the segment extraction for recognition. It seems relatively easy to find a relevant set of distinctive phoneme features and so the segment corresponding with phonemes would be the most suitable for the recognition (8) . The number of segments determining particular classes also points to the same conclusion; in the case of phonemes there would be no more than 40. Yet, difficulties connected with phoneme segmentation as well as those resulting from the lack of univocal dependence between the parameters of some concrete realization of the stochastic process and the *process* itself, make it too difficult to work out a highly efficient automatic system of speech recognition. Oome authors consider it simply impossible to develop studies bas^d on phoneme analysis and recognition (7,8) pointing to the practical difficulties of segment extraction. It is possible to conclude that a solution of this problem will become a basis for the development of some methods of speech analysis by means of particular speech sounds of a given natural language.

The process of segmentation when properly conducted will help in solving some other problems of speech wave analysis e.g. time compression, coding, code ciphering and recognition. Many scientists have noted the role of segmentation in the process of speech wave analysis (6,7,8,9) . Yet, they were too strict in their choice of segments derived from the analog model. Naturally, therefore, they did not obtain the type of results which could be acquired through a segmentation based on the segments chosen in the most optimum manner. This paper attempts to indicate the algorithm of automatic segmentation based on microphonemes as the dynamic segments.

Flicrophoneme as the Fundamental Segment of Voiced Speech~Waves

We are concerned with voiced speech,since only for that class of speech is possible to extract fundamental segments in a simple way, keeping in mind the purpose this extraction could *serve.* It is also possible to resolve the stochastic signal into the class of stationary signals (as considered in frequency domain) by means of some simple methods, for examples, Fourier's transformation. It corresponds with the model of the ear presented above. If we consider the objective function in speech wave analysis the following requirements arise:

- for the purposes of information recognition Cits content) it is necessary to accept segment S_R^i, $i = 1,2,3,\ldots$, making it possible to realize the set of relations m; for the correct classification, while considering the subset of parameters $t_R \subseteq t$ defined for a given segment. It may be presented as the following condition:

$$1 \qquad m_i : t_{k_{il}(S_R^i)} \longrightarrow t_{Ri} \; ; \; i=1,2,\ldots$$

meaning that there is a relation m_i which makes it possible to transfer the set of parameters defined for any "i" realization of the class K_i, $i = 1,2,\ldots,m$ ("i" segment S_R) into the set of parameters t_{Ri} relevant to class K_i .

- for the purposes of compressing and coding a speech wave, it is necessary for the extracted segments S_K^i to satisfy condition No 1. Considering the time compression, the best solution (if possible) would be to extract such segment S_K so that they will be realization of the quasi-periodic waveforms within a given class K_i . That quasi-periodicity should be conceived as the periodicity which is discrete from the point of view of segments (time quanta of real signals) . In other words the segment S_K would constitute a string of the segments s_K

$$2. \qquad S_K = \langle s_K^j \rangle \; ; \; j = 1,2,\ldots$$

In the time domain t (continuous) it would be possible to present it in the following way :

$$3 \qquad \overline{S_K(t)}^{t_K} = 1(t_K - t) \cdot \sum_{i=1}^{n} \left\{ S_K(t) \cdot 1\left(\sum_{r=0}^{i} T(r) - t \right) \cdot 1\left(t - \sum_{r=0}^{i-1} T(r) \right) \right\}$$

where: $\overline{S_K}^{t_K}$ means the segment S_K defined for a period of time t_K (conceived as a set of samples of a signal in the interval of $[0,t]$), n is a function of time t and it is calculated according to the following condition

$$4 \qquad \sum_{j=1}^{n} T(j) \leqslant t$$

It is illustrated in Figure 2.

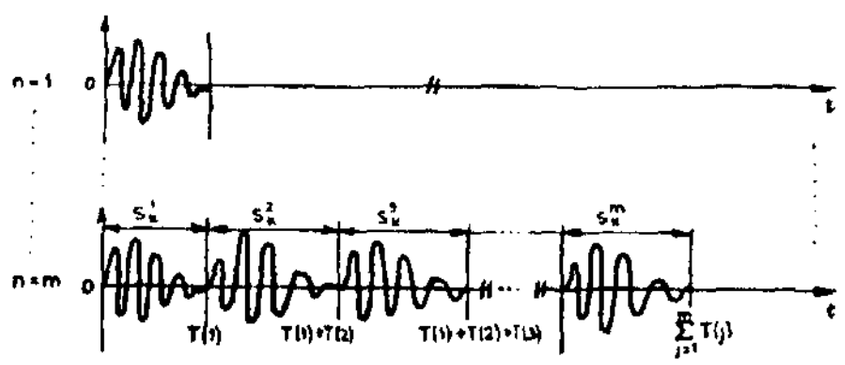


Figure 2

Condition No 1 determines the time for the segment *SK* ; it is tk .

Considering the function of compression and coding, *i*K should be the maximum length of time of the segment S_K satisfying condition No 1. Fundamental segments s_K^j forming quasi-periodicity of the segment SK will be called primary segments. Each of them is of different duration T(j) and they form quasi-periodicity within the segment SK : - see formula No 3. With regard to the function of speech wave time compression, primary segments should satisfy the following condition :

$$5 \qquad \overline{S_K}^{T(j)} = \min_t S_K$$

This condition states that a number of primary segments *n* should be maximum.

- for the purposes of speaker identification it is necessary to indicate such segment S_I^i ; $i = 1,2,3,\ldots$, for the subset of parameters afterwards defined Cidentifying the speaker) to satisfy condition No 1 in consideration of a given class of speakers Kp and also :

$$6 \qquad \overline{t_{Ip}}^{S_I^i} \equiv t_{Ip}' \; ; \; i=1,2,\ldots$$

This condition indicates that the subset t_{Ip} of parameters identifying the speaker specified in consideration of a freely accepted segment Si (free in the sense of time) is the equivalent of the subset of parameters t_{Ip}' • These parameters are defined by means of one of the possible methods and the subset is calculated with some optional duration of speech for a given speaker, belonging to tho class Kp ,

- in the process of reducing the noise the most essential parameters are : the speech spectrum and particularly the average spectrum of a speech wave. Let us denote the subset of those parameters : t_s . With this approach it is necessary to define the segment So such as :

$$7 \qquad \overline{t_s}^{S_o^i} \equiv t_s' \; ; \; i=1,2,\ldots$$

This condition means that it is necessary to indicate such a segment So for the calculated parameters of the average speech spectrum, to fix univocally the average spectrum parameters calculated for

a speech wave of any duration and at any moment. In this case it is necessary to indicate a segment which would satisfy condition No 7 and which would at the same time be the most favourable considering the minimal time interval. This approach shortens the process of indicating the parameters ts considerably. With regard to these considerations, it is necessary to state that in the process of speaker identification and noise reduction, the choise of the best segment should be carried out considering its length of time. The shorter the segment Si or So , the larger the possibilities of realizing both processes. Therefore, it will be possible to.indicate the minimal segment of a speech wave and through that to identify the speaker. The segment chosen in this way must satisfy condition No 6 and 7, corresponding with condition No 1 in that the set of parameters is indicated for a ftiven speaker p . But for the purposes of compression and codification it is necessary for the segment SK to be maximum and also to satisfy condition No 1 as in the case of recognition. In this case the aim is also to find the segment SK(J) which would satisfy conditions No 1,3 and 5. We must search for a primary segment satisfying the following condition :

$$8 \qquad \bar{t}^{\bar{S}_{\kappa}^{T(j)}} \equiv \bar{t}^{S_n^j} \quad ; \quad j=1,2,\ldots$$

within a given class Ki , i =1,2,...,m. Because it is well known that the acceptance of the segment S« fixes the number of classes K_t^* tCt^fCS*)), the minimation of the number of classes Ki is the additional condition that should be taken into consideration while choosing the segment SR . If we accept syllables and words as the segment $S«_f$ the number of classes in which it would be necessary to include the analysed segments would be tens and hundreds of thousands. Therefore it seems that it is necessary to accept a phoneme as the segment S_n or a primary segment which is the equivalent of phoneme considering the condition No 1. The number of classes with a segment so defined is m ■ 40 for the Polish language. With this number of classes a given phoneme as a segment S* is the longest segment and it would fix the segment 3K within the class Ki . Yet in spite of its usefulness for the purposes of recognition and compression, it would not be the most favourable segment because the condition 5 would not be satisfied. And only finding the primary segment that would meet all the requirements would make it possible to state the existence of the most favourable choice of the speech wave segment, in terms of recognition of the content and time compression of a speech wave •

The number of classes according to condition No 8 would be retained. The additional effect of such an approach on speech wave analysis would be the obtaining of a dynamic analysis of parameters defined within the primary segment "5K . Concerning the two remaining goals, speaker identification and reduction of noise, we may accept the primary segments "SK as fundamental only if it is possible to determine the parameters £r/and £a * on the basis of any parameter combination estimated within a given segment *»K . In the case of speaker identification is advisable to estimate parameters for the dynamically analysed speech wave. Spoed and changes in the larynx vibrations are some of the dynamic parameters characterizing the speaker and his articulation : let us denote the length of time of the larynx tone : TK • It is a function of time for a given voiced speech wave. Therefore it is advisable to accept the primary segment ^5K; its duration is T(j) - TK . Tt is different from the majority of algorithms assuming T(j}= const . This approach would also give the phase accordance of the analysed segments (11) • The primary segment is based on pitch period equal to TK corresponding to the analog ear model (TK 3 10 ms) and is called a microphonerae in this paper. Our studies have entirely confirmed the pertinence of the accepted reasoning and also the fact that a microphoneme satisfies all the above conditions (12,13) •

Primary Segmentation
Automatic Searching for Microphonemes

Since a voiced speech wave results from the vibration of the vocal cords and is therefore only approximately periodic and since it depends on the individual characteristics of the speaker, it is not possible to assign definitely the frequency which could be considered fundamental for all segments 3 of a speech wave. In other words, the length of time of microphonemes is : T(IW const. The purpose of primary segmentation is the extraction of microphoneme sequence from the continuous speech wave of a given speaker by means of an algorithm realized as a computer programme. The solution of the problem concerning the extraction of quasi-periodic segments of a speech wave is complicated by the fact that not only quasi-periodicity is troublesome but also because the speech waves vary in amplitude and shape. In assigning microphonemes the algorithm uses some of the methods and remarks included in Reddy* s algorithm : Pitch Period Determination (10) • It also completes Reddy's algorithm with the elements connected with the acceptance of the microphonemes as a primary segment and with other parameters appropriate for Polish speech as well as the accepted method of their representation (1) • The algorithm PPD was used as an element to indicate the location neighbourhood of the expected end of a given microphoneme. The algorithm extracts microphonemes and indicates the places of expected microphoneme ends of the analysed speech wave and also gives a proportional indication of the voiced signal content in the analysed speech wave*

As we have already assumed, we shall be searching for the microphonemes only in that part of the speech wave which is voiced according to condition No *3»* Therefore the first step of the present segmentation

is to find and indicate voiced segments of the analysed speech wave. Because the whole procedure of primary segmentation operates on the time signal of the speech wave, the first step makes use of amplitude and frequency criteria • At this stage these criteria make it possible, as well as in the PPD algorithm, to distinguish three fundamental kinds of signals : silence, noise and a quasi-periodic signal. The algorithm has been adapted to the processing of the input signal which is first quantized. It assumes the frequency of sampling : f_p =12 kHz, which makes it possible to represent the primary speech wave quite well. The analog-digital converter working in the AKORD system performs signal quantization on 256 levels, enabling the notation in the form of numerical data ranging from £-128, +127] . This fact has its reflection in the values of coefficients indicating the amplitude thresholds. Initially the following assumptions have been acceoted as in Reddv pro :

1. In any examined speech wave (with the same speaker) the length of neighbouring microphonemes cannot differ more than 20%
2. When the frequency of the larynx tone is contained within 70-450 Hz the frequency test f_p =12kHz will produce a length of microphoneme within the bounds of 26-171 samples.
3. The examined speech wave will be also accessible in the possible process of correction. This assumption is easy to realize in the case of computer processing.

The succeeding speech wave segments of D = 256 samples (21,3 ms) undergo the amplitude criterion. The accepted length D results from the fact, that in the segment S· longest microphoneme s^i_k should be contained $(\max(T(j)\cdot f_p)<D)$.
The segments S^i_D , i =1,2,... are accepted as silence segments if the maximum amplitude within a given segment does not exced the threshold value σ_1 = 8 , σ -times.
In other words a variable α^i

$$\alpha^i_{\sigma_1} = \#\left\{x_i : x_i \geqslant \sigma_1 \wedge x_i \in S^i_D\right\}$$

(where $x_i \in S^i_D$ means the samples o: the segment i ; its length is D whereas #{ } means the cardinality of the set)
assings some segments S_D accepted as silence in the case of the condition $\alpha^i_{\sigma_1} < \sigma$. In this algorithm σ=5 was accepted. At the next stage only those segments S^i_D are accepted for which $\alpha^i_{\sigma_1} \geqslant \sigma$. All those segments (of the length D) having $\alpha^i_{\sigma_2} \geqslant \sigma$ where σ_2 =43 are accepted as the segments of quasi-periodic character. All the remaining segments undergo the frequency criterion. Initially they are included among the class of high frequency signals or among the class of voiced signals or "conditional noise". Frequency criterion is based on the estimation of zero-crossing parameters. The segments S_D for which the number of zero-crossings is larger than σ_3= 96 (and that with f_p = 12kHz equals the frequency excceding 2,5kHz) are considered high frequency noise segments, .faereas the segments for which the number of zero-crossings exceeds σ_4 = 50 C 1,2kHz) and their maximum amplitude is $\max\{x_i\}<\sigma_2; x_i \in S^i_D$ are accepted as "conditional noise".
"Conditional noise" may be accepted as a voiced segment or as high frequency noise. If there is a group of "conditional noise" segments in the close neighbourhood of the segment accepted as high frequency noise, the whole group is accepted as high frequency noise. Otherwise the "conditional noise" segments are considered and included among the class of voiced segments.
If we denote the segment S» accepted as silence - O, voiced segment - 1, "conditional noise" - 2, and high frequency noise - 3, the signal represented by the following sequence of the approprite classes of the segments S_D

00011100011122211122331122331 11
is classified as
00011100011111111133331133331 11

During the next stage of the algorithm we make use of the results of the previous stage. Only those segments which have been accepted as voiced are exposed to processing. It is necessary to search for indexes t for which the momentary values correspond with the significant amplitude of a given microphonemes. As well as PR) algorithm the following terms are introduced - local maximum and local minimum, absolute maximum, significant maximum and minimum and also significant extreme. Some modifications concerning the semantic content of the above terms have been introduced and also the algorithm of their indication has changed.
If we denote the vector representing the input signal as X then:

$x_i \in X$ is a local maximum if
$(x_i > x_{i-1}) \wedge (x_i \geqslant x_{i+1}) \wedge (x_i \geqslant 0)$
and $x_i \in X$ is a local minimum if
$(x_i < x_{i-1}) \wedge (x_i < x_{i+1}) \wedge (x_i < 0)$

We shall denote them $\max_t x$, $\min_t x$ respectively* On the basis of local maxima the absolute maximum, the significant maxima and significant minima are indicated, we denote them max^x, min^x, A. The absolute maximum A is calculated for the succeeding segments of the length D1 = 8D C it makes about 170.6 ms). Here, it is necessary to point out that a division of a signal into segments of the length Dt as dictated by the limitations posed by the ZAH-41 is different from the principles given in the algorithm PPD and aims at using as few memory cells as possible and achieving compatibility with the system AKORD (1,3) . Accepting the segments of the length D>i greatly inffluences the process of dynamic analysis of the microphoneme periods. It is also extremely significant in the process of correction, during the final stage of algorithm. The indication of the period of the larynx tone for such segments allows independence from fairly essential changes of the length of the larynx tone owing to some changes in articulation, contrary to the indication of the expected period of the larynx tone estimated continually through the whole signal X.

It is particularly important for long signals considering the initially accepted assumption No 1. The indication of segments of duration D_1 influences the first stage of the algorithm in which 8 segments S are indicated successively for each step. Within every segment $S^k_{D_1}$, $k = 1,2,\ldots$ the absolute maximum A_k is assigned.

$$A_k = max \left\{ max_{ln}\, x \right\}_k$$

where $\{max_{ln}\, x\}_k$ means the set of local maxima indicated within the segment $S^k_{D_1}$. Having A_k , we indicate the significant maxima and minima within the sets of local maxima and local minima. The significant maximum must satisfy the following conditions:

$$W1.\ \left(max_{2n}x \geqslant 0.9A_k \vee \begin{cases} max_{2n}x \geqslant \frac{max_{2n-2}x + max_{2n-1}x}{2} & ;n \geqslant 2 \\ max_{2n}x \geqslant 0.6 \cdot A_k & ;n < 2 \end{cases}\right) \wedge$$

$$\wedge\ G_2 \leqslant \rho(max_{2n}x, max_{2n-1}x) \leqslant G_3 \wedge \rho(max_{2n}x, max_{2n-1}x) = min\,\rho$$

where $\rho(max_{2n}x, max_{2n-1}x)$ indicates the distance between the previous and the current significant maximum. That distance should be minimum within the defined interval. If the condition W1 is not satisfied the first "essential" local maximum is chosen according to the condition:

$$W2.\ max_{2n}x \geqslant u \cdot max\left(max_i x : G_2 \leqslant \rho(max_i x, max_{2n-1}x) \leqslant G_3\right) \wedge$$

$$\wedge\ G_2 \leqslant \rho(max_{2n}x, max_{2n-1}x) \leqslant G_3 \wedge \rho(max_{2n}x, max_{2n-1}x) = min\,\rho$$

If the significant maximum has been indicated considering the condition W1 and if that condition is satisfied:

$$\rho(max_{2n}x, max_{2n-1}x) > 2 \cdot G_2$$

then we make an attempt at finding the additional significant maximum $max^a_{2n}x$ according to the condition W2 within the interval satisfying the following inequality:

$$G_2 \leqslant \rho(max^a_{2n}x, max_{2n-1}x) \leqslant \rho(max_{2n}x, max_{2n-1}x) - G_2$$

This algorithm is different from the one presented by Roddy, and thus the significant extremes will have different meanings than in the algorithm PPD. Yet owing to this difference, it is possible to eliminate errors of the type of amplitude fluctuation caused by low frequency ($T \approx 2T_k$)
The algorithm of searching for the significant minimum is analogous to the search for the significant maximum, Speech wave analysis has shown that because the envelope of the larynx tone is of a rather differentiated character, not all significant maxima univocally assign the places of extremes of the microphoneraic envelope. It can be observed that whenever there exists a significant maximum that we accept as the extreme of the microphoneme envelope, there also exists a corresponding significant minimum within a small neighbourhood. This can be explained by the occurence of the greatest turbulence in the speech wave at the moment when the vocal tract is excited by the release of the pressure resulting from the vibration of the vocal cords, the moment in time when the microphoneme begins • Therefore, as the "significant extremes" only those significant maxima are chosen which have the corresponding significant minima in their neighbourhood (4.5 ms - 42 samples as 1/4 of the longest expected microphoneme) . Thus obtained, microphoneme markers undergo the logical processing. During the stage of the logical processing we make use of the already presented assumption that microphonemes are searched for in the quasi-periodic signal and that the following inequality is satisfied:

$$0.8 \leqslant \frac{T(j)}{T(j-1)} \leqslant 1.2$$

This assumption makes it possible to eliminate the errors of the second stage, e.g. finding an extreme in a place where it should not be or not finding it in the place where it should be. In order to eliminate such errors an expected value on the average period of microphonemes T_{sp} is calculated. It is calculated as a discrete statistical distribution of distance among the selected extremes within the segment $S^k_{D_1}$, consistent with the earlier accepted settlements. We shall denote it T^k_{sp} . Thus calculated, T^k_{sp} is used for the stage of correction on the same analogous basis as it was done by Reddy (10) . The acceptance of T^k_{sp} calculated along the fixed length D_1 of the segment $S^k_{D_1}$ makes it possible to eliminate the errors, resulting from intonation changes of longer statements. Thus the correction's efficiency has been considerably increased while retaining all of Reddy's algorithm and the coefficients accepted by him.

The process described above concerning the logical processing is iterated for the same segments according to need. The retrieved markers indicate the significant places within the microphoneme and those regions make a basis for finding the ends of microphonemes. The author will consider the zero-crossing point or the point where the signal assumes a value which is the closest to zero, the end or begining of the microphoneme. The end or begining of microphoneme is indicated in the region where the envelope of the signal assumes a minimum. Such a choice for the place of the microphoneme end reflects the physical aspect connected with the excitement of the vocal cords and also it corresponds with the results of the spectral analysis. The spectral analysis concerning microphonemes showed great parameter stability.
The agreement of phase as well as the great relevance of microphoneme and phoneme characteristics, belonging to the same speaker and to the same class of speech sounds, proves the correct segmentation. In accordance with the phenomenon of vocal cords vibration we shall search for the place of the microphoneme end (or the begining) before the maximum of turbulence assigned by the significant extremes, /e .search for the ends of the microphoneme before the significant extremes or before the proceeding local maxima. We accept a principle that the end of the microphoneme is calculated directly before the local maximum which is the furthest on the left and its value is not less than 1/2 of the

value of the succeeding significant extreme. Yet there is a condition where the distance between the indicated end of the microphoneme and the succeeding extreme should not be larger than $0.5\ T_{sp}$. Let us denote this distance as E^j ; where j is a succeeding number of the microphoneme. It is used for indicating the average distance between the expected end of the microphoneme and the succeeding extreme.

$$E_{SR} = \frac{\sum_{j=1}^{m} E^j}{m} \quad ; \quad m = max\, j$$

This value is used for the correction of erroneously calculated ends of microphonemes. Such errors may appear in the situation shown in Figure 3.

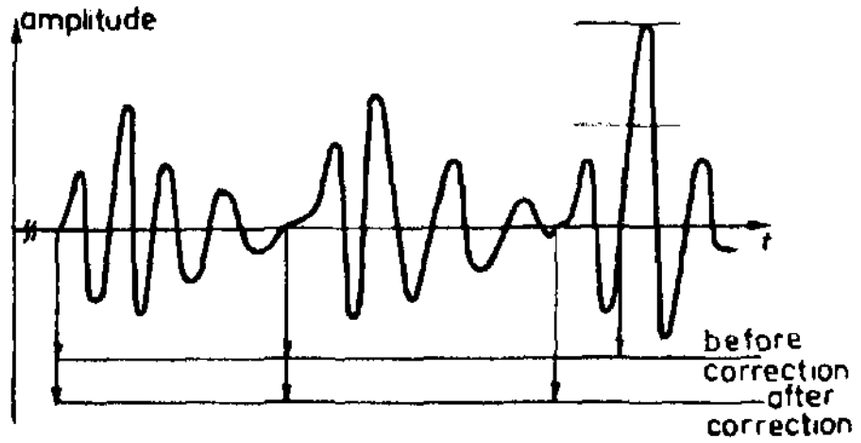


Figure 3

All the ends of microphonemes which satisfy the condition:

W 3. $$|E_{SR} - E^j| < \frac{1}{16} T_{sp}$$

are considered correct.

In the other cases, a process ox correction is undertaken which may lead to indicating a new and better end of the microphoneme. The algorithm of correction aims at finding a new end of the microphoneme, satisfying condition W3 . This phase ends the stage *of* correction for a given microphoneme or the stage of indication of the microphoneme end, which is the least distant from the expected end, marked by E_{SR} . The ends of microphonemes are calculated during the stage of correction according to the analogous principle as before but the values of succeeding local maxima need not exceed half of the value of the respective extreme. Because the process of correction may be repeated, there is a condition indicated, concerning the character of this process. The process or correction should be "unilaterally convergent", by which we mean that a new end of the microphoneme will be considered correct, and the stage of correction in which it was calculated will be considered essential, if it is placed closer to the expected end than the previous incorrect one and also, if it is on the same side, in relation to the expected end, as the end considered incorrect. In the case of "oscillation" of the correction proccess towards the expected end, we accept the end of the *microphoneme* indicated during the previous stage of correction as correct. As a consequence of a possible correction we obtain the markers of microphonemes considered the results of the algorithm of primary segmentation.

Summing up, the algorithm of primary segmentation is divided into four stages:

Stage I - The indication of a quasi-periodic signal and its extraction from an input signal.
Stage II- The calculation of the significant extremes of a voiced signal.
Stage III - The correction of the extremes calculated in stage II
Stage IV - The calculation of the ends of microphonemes
part I - An indication of the expected ends
part II - The correction ; an indication of the resultant ends of microphonemes.

Conclusions

The procedure described in this paper was used for the analysis of 10 single words * with satisfactory results. In the majority of cases this algorithm found more microphonemes than a man could have done on his own. In particular this cases concerned word endings. It was impossible to carry on more exacting research on longer signals because of the ZAM-41 computer's slowness. Due to this difficulty, the above described algorithm is processed in Fortran on the H-6030 computer. This new programme will thus make it possible to test a larger number of longer utterances. Therefore, by working with more and longer utterances, it will be possible to indicate if this algorithm is better, in any cases, than those previously used for frequency analysis of speech sounds.

References

1. A. Dziurnikowski, M. Zdrojek : Organizacja współpracy urządzenia anlogoweo-cyfrowego AKORD z maszyną cyfrową ZAM-41
Papers of Symposium : Metody bezpośredniego wprowadzania i wyprowadzania informacji tekstowej i obrazowej w systemach informatycznych, Jabłonna 1973.
2. L. Macherzyński : Określenie widma przebiegu wprowadzanego do EMC przez urządzenie analogowo-cyfrowe AKORD i wyprowadzanie przebiegów o zadanym widmie
Papers of Symposium : Metody bezpośredniego wprowadzania i wyprowadzania informacji tekstowej i obrazowej w systemach informatycznych, Jabłonna 1973
3. A. Dziurnikowski : System AKORD, SUP-1 CAM-TRA
Internal raport of M.U.C.C. 1972.
4. P. Kolers, Murray e Den : Recognizing Patterns- Studies in living and Automatic Systems
MIT Press, Cambridge, Mass.,1968

* This words are: ala, rojal, rzeka, wąch, jodła, nora, żyba, masz, foka, ryba.

5. S. Zwicker, R.Feldkeller : Das Ohr als Nachrichtenempfanger
Stuttgart ,1967
6. Pierre Vicens : Aspects of Speech Recognition by Computer
(a dissertation) Com. Scien. Dep.
Stanford University, 1969
7. E.A. Sapozkow : 3ygnal mowy w telekomunikacji i cybernetyce
./arszawa, VNT, 1966
8. R. Tadeusiewicz : Sioniczna koncepcja mowy
Papers of Symposium: Zietody Bezposredniego wprowadzania informacji tekstowej i obrazowej w systemach infor-
r.iatycznych, Jablonna, 1973.
9. J.L. Flanagan : Speech Analysis, Synthesis and Perception
Berlin, Springer-Verio^, 1965.
10. D.R. Reddy : Pitch period determination of speech sound
Communication of the ACM, 1967, vol.10
11. G.Modena, C. Scagliola, E. Vivalda : Influence of phase handling on short time speech spectra
The 8 International Congres on Acoustic, London, July 1974, Vol. 1. p. 238. Contributed Papers
12. A.Dziurnikowski : Segmentacja pierwotna sygnaiu mowy i ekstrakcja mikrofonemow
Papers of Symposium : Komputerowe systemy przetwarzania danych doswiadczalnych, Kazimierz, 1974
13. A.Dziurnikowski : System rozumienia mowy SUSY
Papers of Symposium : Metody rozpoznawania, klasyfikacji i wyszukiwania infornacji, Jadwisin 1975.