БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

ПРОБЛЕМИ НА ТЕХНИЧЕСКАТА КИБЕРНЕТИКА И РОБОТИКАТА, 48
PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, 48

София . 1999 . Sofia

On the Use of Wave Transformation for Determination of Voice Signals Emotionality*

Ivan Moustakerov, Atanas Gotchev, Zdravko Nikolov

Institute of Information Technologies, 1113 Sofia

Introduction

One of the most interesting areas in the study of voice signals is the influence of the psycho-physiological state on the voice signal characteristics. New methods for signals digital processing have been recently developed that could be examined with respect to their applicability for similar purposes. One of them is the method for signals analysis with the help of wave transformations.

In the time period when the vocal signal can be considered stationary (10–20 ms), one phoneme is generated. The transitions towards different phonemes and their intonation are traced by the so called trajectories of the formants and the contour of the main tune. Investigations have been accomplished, in which it is noted that the formants trajectories reflect different transitions among different sounds and depend on the phrase studied, while the contour of the main tune depends on the intonation. The study of the emotional status is done with respect to any condition accepted as normal. This research uses methods for formants separation (by linear prediction, quick Fourier transformation or kepstral analysis [1]) or methods determining the main tune (auto-correlation, cepstral methods, transitions through a zero level and so on).

Characteristic features of wave transformations application

The advantages of wave functions use in the analysis of sharp changes in the signals are well known. The good localization in time and in frequency enables the exact time determination of the short high-frequency components and exact frequency determination of low frequencies components. It can be accepted that the wave transformations are close to the physiological mechanisms of hearing and are appropriate to investigate speech signals from the view point of their psycho-physical (including emotional) characteristics.

* The present investigations are sponsored by the National Fund for Scientific Research, Contract No TH-467/94.

For this purpose from all the existing basic wavelets, the most appropriate for analysis must be selected [3, 6, 7, 8, 9, 13]. They have to be with good frequency and time localization, linear phase and acceptable from a computing view point.

The wavelets of Battle-Lemarie have been selected in the case. Their application follows the classical diagram of wave analysis [9], and namely – a scaled function is selected, the frequency characteristics of the corresponding wavelet and quadratic-mirror filters are defined, the coefficients of the pulse characteristics of the filters also, which can be used to compute the decomposition. The application of this type of wavelets for analysis with different levels of resolution are based on the use of polynomial spline functions of 2p+1 order. They have been studied at first by Lemarie [11] and independently by Battle also, and then used by Mallat [4, 12]. The approach and the denotations of Mallat [4] will be used to explain the way of filter coefficients computing.

The linear space of all the functions of $L^2(R)$ is considered, that are p-times differentiable and are equal to a polynom of $2p+1$ order within each interval $[k, k+1]$. The scaled function, connected with this space, is described in the frequency domain by the expression:

(1) $$\Psi(\omega)=\frac{1}{\omega^n\sqrt{\Sigma_{2n}(\omega)}}, \qquad n = 2p+2,$$

and the function $S_n(\omega)$ – as

(2) $$\Sigma_{2n}(\omega) = \sum_{k=-\infty}^{\infty} \frac{1}{(\omega + 2k\pi)^n}.$$

The frequency characteristics of the low-frequency quadratic filter $H(\omega)$ is connected with the scaled function by the equality:

(3) $$\Psi(2\omega)= H(\omega)\,\Psi(\omega),$$

from which $H(\omega)$ is obtained as:

(4) $$H(\omega)=\sqrt{\frac{\Sigma_{2n}(\omega)}{2^{2n}\,\Sigma_{2n}(\omega)}}.$$

The computing of the function $S_n(\omega)$ can be realized finding the $n-2$ derivative of the function

(5) $$\Sigma_2 = \frac{1}{4\sin^2(\omega/2)}.$$

Wavelets formed by cubic splines (p=1, n=4) have been used in the present investigations. In this case it is necessary to obtain $S_s(\omega)$ and $H(\omega)$ by formula (4). After mathematical transforms it is obtained:

(6) $$H(\omega) = 2(1-u)^4\,\frac{315-420u+126u^2-4u^3}{315-420v+126v^2-4v^3},$$

where $u=\sin^2(\omega/2)$, $v=\sin^2\omega$.

The pulse characteristics of the filter $h(n)$ is obtained by the inverse Fourier transformation. It is symmetrical with respect to the null and is of infinite length, but it attenuates exponentially and can be limited to a finite number of coefficients for practical calculations. The coefficients of the mirror filter are obtained with the help of the relation [4]

(7) $$g(n) = (-1)^{(1-n)}\,h(1-n).$$

The algorithm of Rioul has been chosen as a basis of the software realization with the use of mirror-quadratic filters for wave transformation computing.

The filter coefficients of Battle-Lemarie can be applied in algorithms, enabling the computing of expansions with redundancy (pseudo-continuous) [5, 10]. They avoid the decimation of the high-frequency output filters and in going down the low-frequency part of the expansion pyramid, make filtrations separation, followed by adequate time re-ordering of the output coefficients. It should be noted that due to the noncausal character of the filters g and h, continuation at the ends of the filter sequences is necessary. Mallat [4] recommends a mirror continuation. In other research studies continuation of the two two end values ($x(-L) = ... = x(-1) = x(0)$ and $x(\mathrm{NofPts}+L) = ... = x(\mathrm{NofPts}) = x(\mathrm{NofPts}-1)$) is applied, where L is the filter order and NofPts is the filtered sequence length. Adding of zeroes is also possible, which has the advantage, that it does not bring additional (non-existing) energy to the signal.

The computing expenses are determined by the filter length and the length of the sequence analyzed. For a discrete wave transformation at one point, for the current scale, $2L$ multiplications are required (L is the number of the filter coefficients for filters of equal lengths). Since the input points for the successive scales decrease by a degree of two due to the decimation, $L(1+1/2+1/4+...+1/2^{j-1}) = 2L(1-2^{-j})$ multiplications are necessary for a discrete wave transformation at one point and J scales. The multiplications number declines towards $2L$ with the increase of the number of scales. The total number of multiplications is $2LN$ for N input samples (proportional to the input data number).

Experimental investigations of test voice signals of emotional character with the help of Battle-Lemarie wavelets

The method described, using Battle-Lemarie's wavelets and Rioul's algorithm is accepted as a base for the design of a software module for wave transformations, with the help of which experimental study has been accomplished. The module realizes wave decomposition for quick wave transformation with the help of Rioul's algorithm using quadratic-mirror filters. The following coefficients of a quadratic filter of Battle-Lemarie are used:

$H(0) = 0.766130$	$H(8) = 0.008685$
$H(1) = 0.433923$	$H(9) = 0.008201$
$H(2) = -0.050202$	$H(10) = -0.004354$
$H(3) = -0.110037$	$H(11) = -0.003882$
$H(4) = 0.032081$	$H(12) = 0.002187$
$H(5) = 0.042068$	$H(13) = 0.001882$
$H(6) = -0.017176$	$H(14) = -0.001104$
$H(7) = -0.0017982$	$H(15) = -0.000927$

57 files with test signals have been analyzed (27 with positive emotion, 21 – with neutral and 9 – with negative emotion). After the wave transformations, the signals for seven successive scales have been obtained, that are equal in length to the input signal and are of real values. They are called octave signals. There were obtained 7 files for each scale or totally 399 files that can give qualitative idea about the character of the expansions of the vocals selected.

The maximums existing in the octave of the basic tune (characterized with values close to the maximum), can be used to determine the alterations within the period of the main tune, which on its side can give the emotional component of the signal studied.

These files, that contain signals of the octave in the main tune of a voice signal, generated by one and the same source with different emotional content, have been chosen. This enables the use of any base for comparison. The number of the similar characteristic files has been 12 totally for the vocals "E" and "O", pronounced by different sources with different emotions. The software module, decomposing the signals studied, has been thus expanded that it computes the length of the main tune period (the distance between two successive maximums). The alteration of the period of the basic tune has been recomputed as an alteration in the frequency of the main tune (the discretization frequency/length of the period) and visualized graphically. The trajectories of the main tune frequency of one and the same source with different emotions in it are given in the (Figs. 1–8).

The following conclusions can be made on the basis of the trajectories of the main tune frequency of vocals spoken with different emotions by one and the same source (Fig.6):

1. There is difference in the curves, characterizing the separate emotional states. The curve, corresponding to positive emotion, is characterized with higher frequencies in the main tune than that with negative emotion, while the curve of neutral emotion is between both.

2. At this stage of the research, only qualitative differences can be accounted, since the quantitative alterations depend on the individual characteristics of the person, pronouncing the voice signal being analyzed.

3. These qualitative differences can be also observed when taking an equal base - one and the same vocal, pronounced by one and the same person with different emotions.

4. Though the qualitative difference is preserved, there is quantitative difference even at repeated pronouncing of one and the same sound with the same emotions by one and the same person – as seen from the graphs for the vocal "O" of the source A in the first and second variant.

An attempt has been made to compare the pronouncing of one and the same sound (the vocal "O") with one and the same emotion by different sources. The corresponding graphs of the trajectories of the main tune frequency are shown in (Fig. 7).

The conclusions from this are as follows:

1. There is again a tendency observed towards higher frequency of the main tune at positive emotion.

2. The individual differences in the pronunciation lead to the decrease of the frequency differences between the neutral and negative emotion.

3. It follows, that the determination of the different emotionality of a voice signal, due to quantitative differences in the main tune frequency would be relevant enough at the presence of an equal base for comparison – one and the same signal source.

The individual differences in the pronouncing of one and the same vocal with one and the same emotion by different sources can be seen in the simultaneous visualization of the trajectories of one and the same emotion for different sources (Fig. 8).

The conclusions made are:

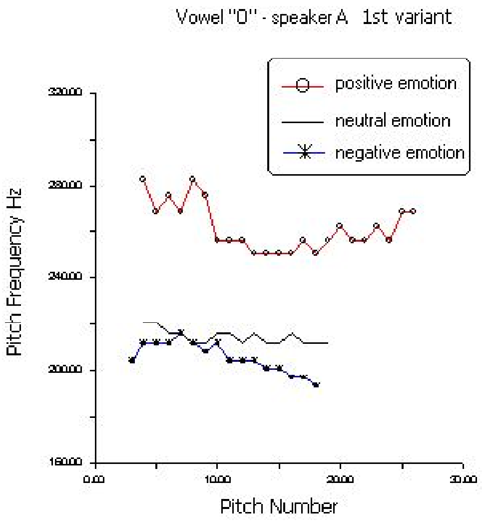
1. The tendency mentioned for increased frequency of the main tune at positive emotion is here confirmed.

2. The digital values are highly dependent on the speaker's individuality.

3. The extracting of adequate conclusions for the voice signal emotionality depends on the relevance of the base accepted for comparison.

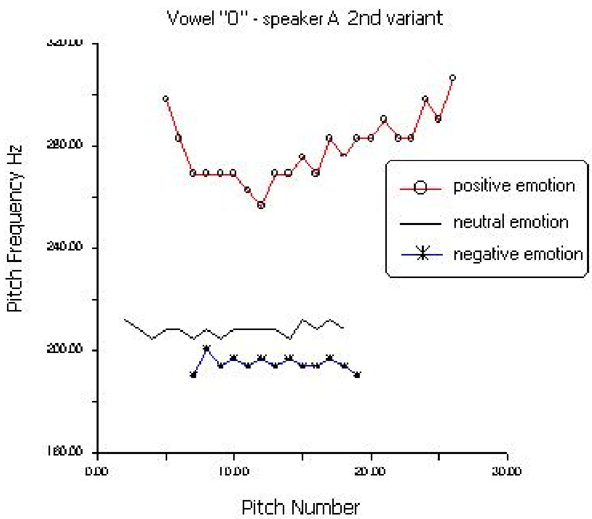
The following conclusions can be done concerning the analysis method selected:

1. The Battle-Lemarie's wavelet is described by filter coefficients and the analysis is realized with the help of efficient computing algorithms. Signals for successive octaves are obtained that are of real values and can be used for qualitative comparison.



Vowel "O" - speaker A 1st variant
positive emotion
neutral emotion
negative emotion
Pitch Frequency Hz
Pitch Number

Fig. 1



Vowel "O" - speaker A 2nd variant
positive emotion
neutral emotion
negative emotion
Pitch Frequency Hz
Pitch Number

Fig. 2

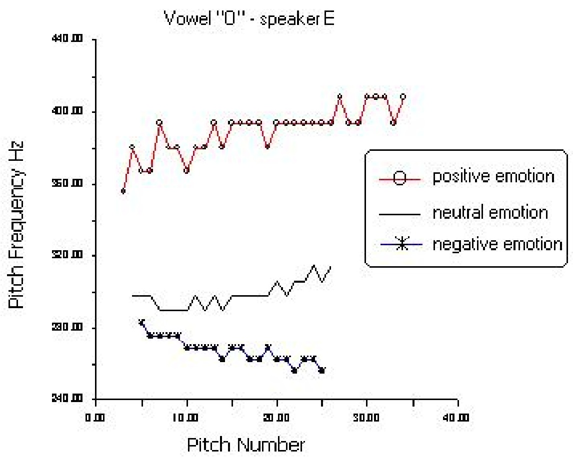


Fig. 3

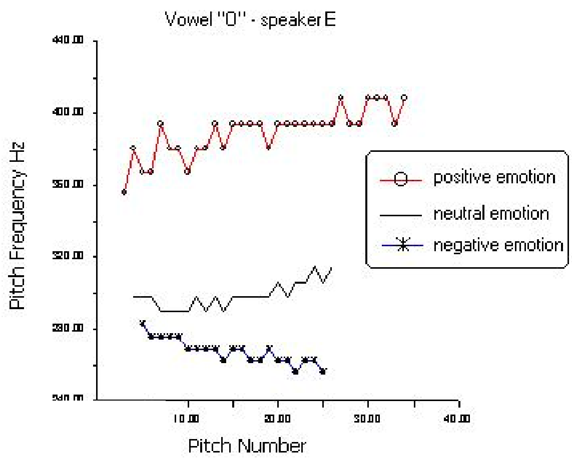


Fig. 4

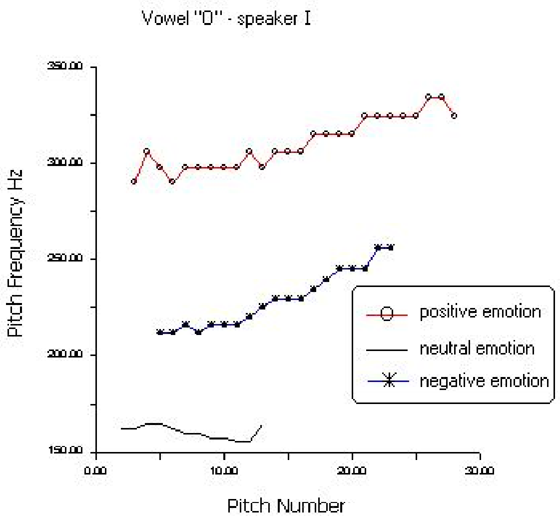


Fig. 5

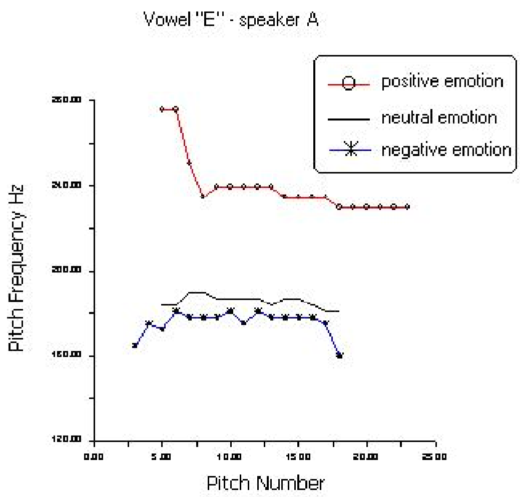

Vowel "E" - speaker A
positive emotion
neutral emotion
negative emotion
Pitch Frequency Hz
Pitch Number

Fig. 6

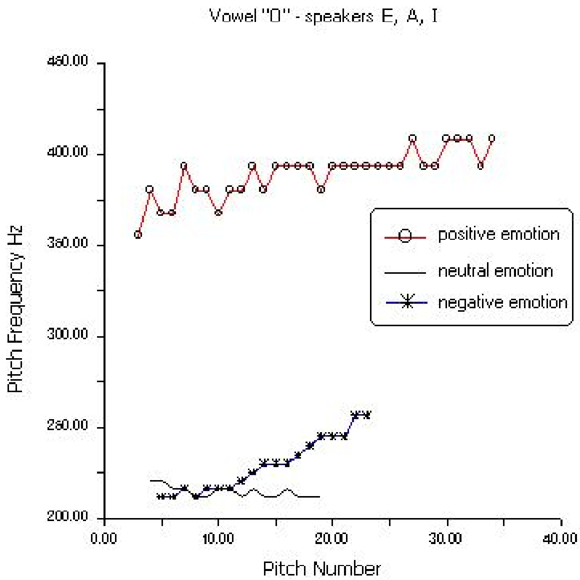

Vowel "O" - speakers E, A, I
positive emotion
neutral emotion
negative emotion
Pitch Frequency Hz
Pitch Number
200.00
300.00
400.00
0.00
10.00
20.00
30.00
40.00

Fig. 7

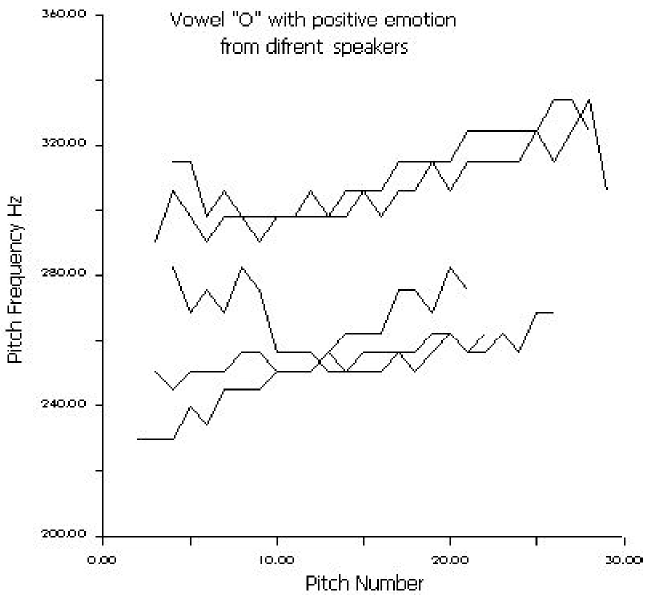


Fig. 8

2. The wavelets are appropriate for determining the moments of glottal closing and opening and the period of the main tune as well, the alterations of which can on their side be used to define the emotionality of the voice signal.

3. The method used allows the programming and automation of the investigations to some extent.

4. The non-stationarities in the signals analyzed do not influence the analysis. The presence of noise in the signals analyzed does not influence the low (low-frequency or of high scale) permissions used to define the alterations in the main tune period as a qualitative measure of the emotionality. The wave transformation here is better than the curve of the minimized error due to linear prediction, which is much more sensitive to noise.

Conclusion

The applicability of the wave transformations in the analysis of voice signals emotionality is proved. There exist qualitative differences in the trajectories of the main tune frequency at emotionally different speech signals. When a base of comparison is available (one and the same source, one and the same vocal, known trajectory of the neutral status), the quantitative differences in the different trajectories can be practically applied.

References

1. Marple, S. L. Digital Spectral Analysis with Applications. Englewood-Cliffs, N. J., Prentice-Hall Inc., 1987.
2. Rabiner, L., B. Gold. Theory and Application of Signals Digital Processing. Moskow, Mir, 1978 (in Russian).
3. Louve, M. Sur le Fonctions aluatoires stationnnaaires de second ordre. – Rev. Sci., **83**, 1945, 297-310.
4. Miyoshi, Y., K. Yamato, R. Mizuguchi, M. Kanaguda, O. Kakusho. Analysis of speech signals of short pitch period by a sample-selective linear prediction. – In: IEEE Trans. ASSP. **35**, 1987, 1233-1239.
5. Monsen, R. B., A.M. Engebretson. Study of variations in the male and female glottal wave. – J. Acoust. Soc. Am., **62**, 1977, 981-993.
6. Morikawa, H., H. Fujitsaki. System identification of the speech production process based on a state-space representation. – In: IEEE Trans. ASSP, **32**, 1984, 252-282.
7. Pinto, N. B., D.G. Childers, A. L. Laiwani. Formant speech synthesis: improving production quality. – In: IEEE TRans. ASSP, **37**, 1989, 1870-1887.
8. Queroz, R. On Lappped Transform. Ph.D. Thesis. University of Texas at Arlington, 1994.
9. Rothenberg, M. An interactive model for the synthesis of voice sounds. – STL-QPSR. KTH, 1981, 155-165.
10. Rissanen, J. Universal coding, information, prediction and estimation. – In: IEEE Trans. Inf. Theory, **30**, 1984, 629-636.
11. Saito, N. Simultaneous Noise Suppression and Signal Compression Using a Library of Orthonormal Bases and ht Minimum Description Length Criterion. Wavelets in Geophysics 9E. Foufouls-Georgiou and P. Kumar, eds. San Diego, CA, Academic Press, 1994, 299-324.
12. Smith, M., T. Barnell. Exact Reconstruction for Tree-Structured Subband Coders. – In: IEEE Trans. ASSP, **34**, 1986, 434-441.
13. Gotchev, A. A method for the Investigation of a Voice Source. Ph.D Thesis, IIT-BAS, 1996.

Применение волновой трансформации при определении эмоциональности речевых сигналов

Иван Мустакеров, Атанас Гочев, Здравко Николов

Институт информационных технологий, 1113 София

(Р е з ю м е)

Публикация связана с определением психофизических особеностей сигналов голоса пре помощи характеристики основного тона сигнала. Показано применение метода волновых трансформаций, который декомпозирует сигнал в так названных октавных сигналах. Предложена гипотеза, что изменения в октаве основного тона характеризируют эмоциональное состояние сигнала. Представлены программные средства для этого анализа.