APPLICATION OF AUTOMATED IDENTIFICATION METHODS OF BOW STROKES TO MUSICAL FOLKLORE RESEARCH

Dedicated to Professor
Jadvyga Čiurlionytė

Artūras Medonis
Academy of Sciences
of the Lithuanian SSR
Vilnius, USSR

Birutė Sinkevičiūtė
Dept. of String Instruments of the
Lithuanian SSR State Conservatoire
Vilnius, USSR

ABSTRACT

The purpose of the presented paper is to search for the common and distinctive features of an automated investigation of music/speech and simple speech signals. An algorithm of musical parameter estimation is based on the application of speech parameter recognition. Topical aspects of the automated ciphering in case the height of the musical folklore sounds is the same are analysed in our report.

INTRODUCTION

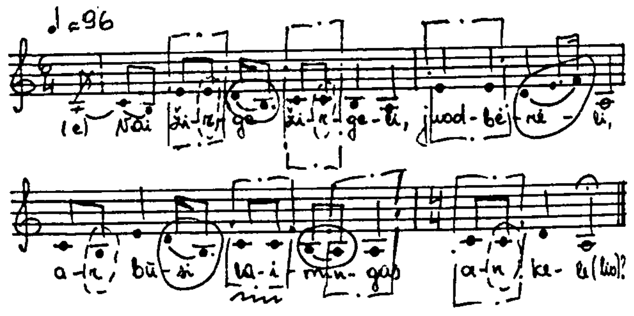
"Lithuanian people's songs art is exceptionally rich and various. The Lithuanian folklore Manuscript fund stores about half a million records of songs"./7 /

The most urgent problems of the Lithuanianfolklore songs were reviewed by Professor J.Čiurlionytė for the first time outside the republic /9, 10/. After that the songs are investigated and systematized in the following organizations:
- the Folk Music Study of the Lithuanian State Conservatoire (founded by Professor J.Čiurlionytė)/6, 8/,
- the Folklore Department of the Institute of the Lithuanian Language and Literature of the Lithuanian Academy of Sciences /7, 17/.

An evolution of folklore witnesses the improvement of folklore investigation methods was attained thanks to the application of new technical equipment (phonograph, tape-records). Lately folklorists often solve their problems with the help of computers, i.e. cipher, analyse and systematize the songs.

From the first stages in folklore investigations both the linguists and etnomusicologists are excited by these problems in the junction of the folklore music and linguistic text / 1,11, 17/. E.g., Vice-president of the Folklore Board of the USSR Composer's Union Alexeyev E. explores the interrelation between musical intonation and verbal inflection on the basis of Lithuanian folk-songs /1/.

Our report deals with a specific aspect of the above junction: it presents an estimation algorithm of musical durations; the algorithm is based on the application of characteristic speech parameters when the height of the sounds analysed is the same (Note example I, case /a/).



Note example I.
/a/ - a case when the height of the sounds sung is the same;
/b/ - consonants with independent musical height and duration;
/c/ - a melodic phrase performed by one and the same vowel;
/d/ - a case when both vowels of diphtong have independent musical durations.

The purpose of the presented paper is to discuss the common and distinctive features of automated investigation of music/speech and simple speech signals.

The 1986 International Computer Music Conference (ICMC) was held in Hague, the Netherland. October 20-24. We presented a report in which we discussed our main topic "On the identification of violin strokes in a real-time performance system" in this ICMC. Now let us review the report in short and to present our main material.

METHOD

Our paper /18/ deals with the automated identification of violin strokes in case the height of the sounds played is the same. E.g. it happens in the main theme of Concert No 2, E-dur by J.S.Bach (for violin and chamber orchestra). Estimations by the statistics

$$L_k=\left(\sum_{i=1}^{N} x_i^2/N\right)^{1/2}, \quad k=1,2,\ldots \qquad (1)$$

are obviously sufficient for sounds and pauses in a pizzicato case but it is not sufficient to identify martele strokes: (1) the estimates slightly differ from one another and the corresponding algorithm identifies the beginning and the end of the strokes unprecisely.

The violin plate bears a property to resonate the sounds played and to cease ringing gradually after stopping playing. Therefore a physical interpretation of the martele envelope is the following: at the beginning of a new stroke we listen to the sounds of new and earlier strokes resonated by a violin plate.

Analogous features of merging sounds are typical of speech analysis too: the signal is segmented in order to determine the limits of separate sounds. Segmentation of diphtongs is especially difficult as the Lithuanian linguist A.Pakerys notes / 15/.

We proposed not only to measure the sound parameters, but also to register the supplementary information obtained from a digital bow /4/.

Formalization of the above statement: let us consider a violin stroke identification function $F(y)$, where $y=\varphi(\nu,L,t)$ - is the function of violin sound determination; ν, L, t are the violin sound parameters: pitch, intensity and duration, correspondingly. The aim of the automated identification is to define the moments t_j:

$$F(y)=\begin{cases} a, & t<t_j \ (a\leftrightarrow \sqcap - m.n.^{*}) \\ b, & t>t_j \ (b\leftrightarrow V - m.n.) \end{cases} \quad \text{in case } \nu=const$$

The segments $X_{K1}, X_{K2},\ldots, X_{KN}: L_{K+1} \gg L_K$ ($k=1,2,\ldots$) were investigated for this purpose. The algorithm is sufficient for pizzicato stroke identification, but in a martele case the segments $k: L_{K+1} \gg L_K$ were not detected. Therefore, we proposed to consider the function $F(y,z)$, where $y=\varphi(\nu,L,t)$ and $z=\psi(t_i)$:

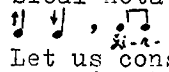
$$\gamma(t_i)=\begin{cases} 1, & \text{as } \psi'(t_i)<\psi'(t_{i+1}) \ (\sqcap - m.n.) \\ 0, & \text{as } \psi'(t_i)=0 \text{ or } \psi'(t_i) \ \nexists \\ -1 & \text{as } \psi'(t_i)>\psi'(t_{i+1}) \ (V - m.n.) \end{cases} \qquad (2)$$

Now we deal with the application of the above methods to the automation of Lithuanian musical folklore analysis. Automated ciphering of the original, in case the height of the sounds sung is the same, (note example I, case /a/) is an urgent problem. Case /b/ of the note example illustrates the situation when a consonant is in good agreement with musical duration.

* m.n. - in musical notation

A phonetic syllable in the speech analysis is determined as the least segment of speech torrent, unit of the pronunciation, which forms the words rhythmically and with emphasis/14/. In literature we did not notise a monosemantical definition of a phonetic syllable in the aspect of musical folklore analysis; let us denote it in terms of a "musical syllable".

There are two points of view between etnomusicologists on the above problem:
(a) musical syllable is not a structural sound which has no influence on the definition of etnomusicological parameters of the investigated melody: metre-rhythm structure, ambitus, the stable marginal sounds of the melodic vertical - the lower and upper tonic - etc.
(b) musical syllable is namely a structural sound which is an independent unit and therefore is fixed in musical notation.

There are many peculiarities in folklore singing and untraditional elements of musical notation, used to express them, e.g.  etc.

Let us consider a musical syllable duration (MSD) function $F(y)$, where $y=\varphi(\nu,L,t)$ is a function of a musical sound determination. The aim of the MSD identification is to define the moments t_j:

$$F(y)=\begin{cases} a, & t<t_j \\ b, & t>t_j \end{cases}, \quad \text{when } \nu=const$$

For this aim the segments $k: L_{K+1} \gg L_K$ are searched as they are presented in the methods for violin stroke identification. The algorithm is not always sufficient (e.g. sound intensity variation is often possible in the same vowel, as it is shown in fig2 for vowel "ė"). Supplement parameter has to be applied to musical parameter determination: classification parameter "voiced speech/unvoiced speech segments" are widely used in speech analysis/ e.g.16/ Conseguently it is expedient to consider the function $F(y,z)$: $y=\varphi(\nu,L,t)$, $z=\psi(\tau)$ and

$$\psi(\tau)=\begin{cases} 1, & \text{when } \tau \text{ is a voiced speech segment} \\ 0, & \text{when } \tau \text{ is an unvoiced speech segment.} \end{cases} \qquad (3)$$

The value of parameter τ can estimated using one of the numerous algorithms discussed in special literature /16, 19/.

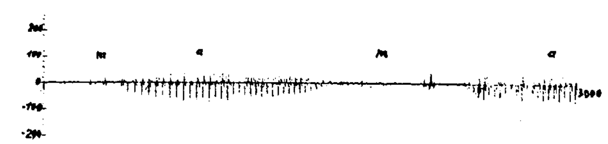


Fig.1 Acoustic signal of word sung " mama "

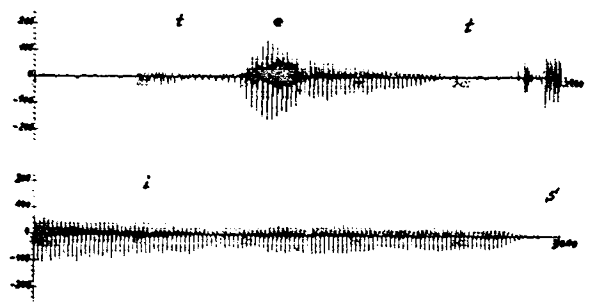
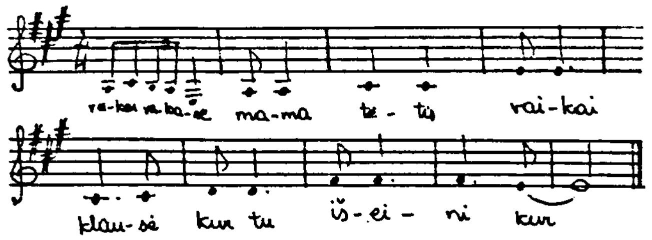


Fig.2 Acoustic signal of word sung "tėtis"

RESULTS

A musical fragment was introduced into the universal computer BESM-6 (Note example II). The sound waveform was sampled at 4 kHz because the fragment was a low register melody. For 100 ms acoustical segments the intensity was computed according to statistics (1) and an algorithm of a delay function $D(p)$ was applied to estimate the pitch. The estimates of ν and L were not sufficient for the definition of musical duration. Therefore, it is expedient to deal with the function with more parameters, i,e, to describe $F(y,z)$, where $z=\psi(\tau)$ .Estimation of the algorithm for segment vocalization is based on a common interpretation of energy and zero crossed frequency functions. An application of this methods gives more concrete results in comparison with the values of $F(y)$.



Note example II.

Pitch (Hz)	122.56	125.41	124.03	132.32	133.85	
Intensity	19.741	20.319	24.451	8.572	6.748	
Sound	-a-	-a-	-a-	-m-	-m-	

Pitch (Hz)	138.33	135.51	129.03	131.41	132.11
Intensity	17.642	20.809	17.642	19.581	20.499
Sound	-a-	-a-	-a-	-a-	-a

Table I. Estimates of parameters and of the sung word "mama"

x x x

In comparison with a real-time performance system it is sufficient to work in interactive regime when the musical folklore signal is processed. That allows to widely use spectral algorithms for pitch determination. In contrast, violin stroke identification requires the application of fast algorithms. One of such is presented in our paper /18/. It reflects an effective utilization of a delay function

$$D(p)=\sum_{k=1}^{M}|y_k-y_{k+p}|,\quad M<N,\quad p=0,1,\dots,N-M$$

As usual, in speech signal processing one of the three models is applied: excitation model, perception model or mathematical model/12/.

By extending the conclusion of Hess our point of view consists in that the application of the known pure mathematical methods is not sufficient in music/speech signal processing. Perception models, based on the musical knowledge, are preferable. The results below of our experiment done on the pitch determination illustrate this standpoint rather well.

Let us consider the following mathematical model of pitch determination

$$y(t)=f(t)+\xi(t),\quad 0\le t<\infty \qquad (4)$$

where $f(t)=\sum_{j=1}^{\infty}A_j\cdot\sin(2\pi\cdot\nu\cdot j\cdot t+\varphi_i)$

$y(t+T)=y(t)$, T is a period, $\nu=1/T$ is a pitch of the function $y(t)$; $\xi(t)$ is a stationary random sequence in a wide-sense. The aim is to estimate a parameter $\hat{\nu}=\hat{\nu}(y_1,\dots,y_N)$. We applied the following algorithms of spectral analysis:

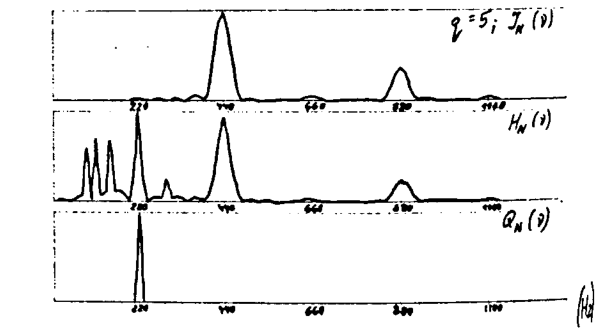


Fig.3

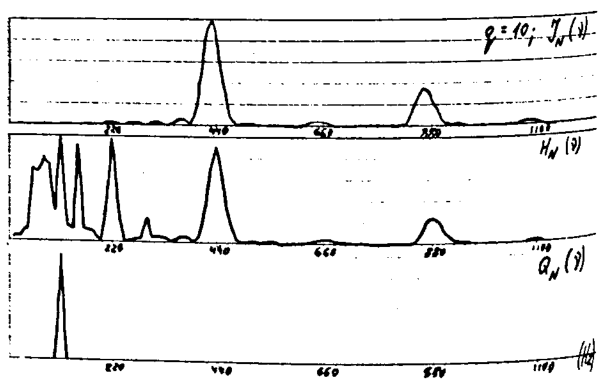


Fig.4

Fig.3, 4 illustrate graphic represantations of the functions $I_N(\nu)$, $H_N(\nu)$, $Q_N(\nu)$ obtained by theoretical series (8).

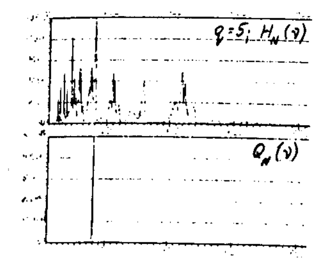


Fig.5

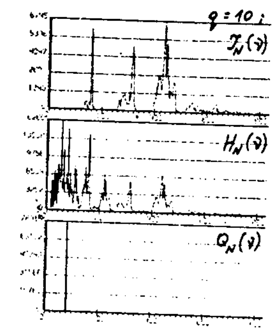


Fig.6

Fig.5, 6 illustrate graphic represantations of $I_N(\nu)$, $H_N(\nu)$, $Q_N(\nu)$ generated by a real series of sound "ė"

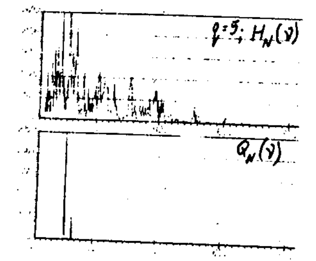


Fig.7

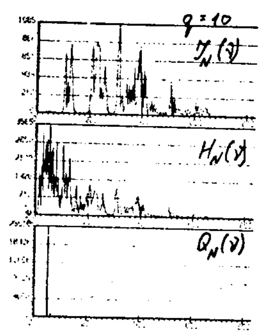


Fig.8

Fig.7, 8 illustrate graphic represantations of $I_N(\nu)$, $H_N(\nu)$, $Q_N(\nu)$ generated by a real series of sound "a"

$$\text{I. } I_N(\lambda)=\frac{\Delta t}{2\pi N}\left|\sum_{k=1}^{N}(y_k-\bar{y})\cdot e^{-ik\cdot\Delta t\cdot\lambda}\right|^2,\quad \lambda\ge 0:\ \lambda=2\pi\nu \quad (5)$$

is a periodogram of time series (4) at the points $\lambda_j=\frac{2\pi}{N'\cdot\Delta t}\cdot j$, where $j=0,1,\dots,N'-1;\ N'=2^{\ell}\ge N$ is a selective number by a user. The values of the periodogram are calculated on the basis of the algorithm of FFT.

$\hat{\nu}=\Delta t\cdot[\arg\max_{\nu} I_N(2\pi\nu)]$ -the pitch estimate (5') by the above algorithm

II. $H_N(\nu)=\sum_{j=1}^{q} I_N(2\pi\cdot\nu\cdot j)$ -the sum of periodograms (6)

$\hat{\nu}=\Delta t\cdot[\arg\max_{\nu} H_N(\nu)]$ -the pitch estimate (6')

III. $Q_N(\nu)=\prod_{j=1}^{q} I_N(2\pi\cdot\nu\cdot j)$ -the product of periodograms (7)

$\hat{\nu}=\Delta t\cdot[\arg\max_{\nu} Q_N(\nu)]$ -the pitch estimate (7')

We deal with one theoretical and two real time series. The modelling series is

$$y_k=\sum_{j=1}^{5}A_j\cdot\sin(2\pi\cdot\nu\cdot j\cdot k\cdot\Delta t)+\varepsilon_k,\quad k=1,2,\dots,N \quad (8)$$

The meanings of parameters are the following $N=2^{\ell}$, $\ell=10$ $\Delta t=0.02$ (MS)

$\nu_N=1/(2\cdot\Delta t)=25$ (kHz) the Naikvist's frequency $\nu=220$ Hz; $A_1=0.1$, $A_2=1.0$, $A_3=0.2$, $A_4=0.6$, $A_5=0.15$ $\varepsilon_k\in N(0,\sigma^2)$ where $\sigma=0.1$, $q=5$ and $q=10$. The results of application of algorithms (5)-(7) are shown in Fig.3-4. Pitch estimate $\hat{\nu}=220$ Hz corresponds to the given value ν in case $q=5$. In another case, as $q=10$, pitch estimate $\hat{\nu}=110$ Hz. The estimate is possible in a mathematical algorithm sense, but it is not logical in the aspect of musical theory. As it is well known /13/ in the theory, harmonics form consistent series of musical intervals: octava, quinta, quarta, b.tertia etc. There are following relations of harmonic frequencies in our example, as $q=10$: $220/110=2/1$ -octava, $440/220=2/1$ -octava, $660/440=3/2$ -quinta etc. Octava interval is repeated in our series, but it is impossible in the aspect of musical theory (as untertone does not exist in a musical sounds). Pitch estimate $\hat{\nu}=220$ Hz generates necessary series of musical intervals, therefore, it is true. This musical knowledge is laid in perception models, based on the application of algorithms of harmonic sieve type/2 /.

Real time series of the sounds "a" and "ė" corresponds to the words "mama", "tėtis" of a music/speech signal. Selection of a parameter q has the influence on the estimation result, as shown in Fig.(5)-(8). Therefore, the application of perception models is preferable again.

ACKNOWLEDGMENTS

We highly appreciate the attention of our colleagues: V.Baranauskienė, K.Augutis, T. Juzeliūnas, A.Dikčius (the Lithuanian State Conservatoire), V.Pikelis and J.Kazlauskaitė (Institute of Mathematics and Cybernetics of the Lithuanian Academy of Sciences).

REFERENCES

1. Alexeyev E.(1986). The Pitch Nature of Primitive Singing. Sovetsky Kompozitor, Moscow, -240p. (in Russian).
2. Allik J., Mikhla M., Ross J.(1984).Comment on "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception".JASA 75, p.1855-1857.
3. Ambrazas A.(1981). New methods for ciphering and catalogation of folklore melodies. Music, ISSUE 1, p.121-123(in Lith.).
4. Askenfelt A.(1984). Study your bowing technique! -Quarterly Progress and Status Report, No 1, Dept of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm.
5. There come nine-corn reindeer running. (1986). Folklore collection. Vilnius, 431p
6. Baranauskienė V., Medonis A., Medonis E, Sinkevičiūtė B.(1986). Some algorithms of Lithuanian folklore unanimous songs (in Lithuanian, to be printed).
7. Collection of the Lithuanian songs(1980) vol.I,The Songs of children. Prepared by the Lithuanian SSR Academy of Sciences, Vilnius, Vaga, -760p.
8. Četkauskaitė G.(1981). The Dzūkai Melodies, Vilnius, Vaga, -621p.(in Lith.).
9. Čiurlionytė J.(1949). Lithuanian folksong, Sovetskaya Muzika, No 6, 60-64p. (in Russian).
10. Čiurlionytė J.(1955).From the history

of the Lithuanian folk-song. Sovetskaya Muzika, No 2, 69-77p. (in Russian).
11. Čiurlionytė J.(1984). A song as the research object. Collection of papers "Daily works", Vilnius, Vaga, 46-53p.(in Lith).
12. Hess W.(1983). Pitch determination of speech signals. Berlin, Springer, -701p.
13. Kazlauskas J.(1975). Musical instruments and score. Vilnius, Vaga,-222p. (in Lithuanian).
14. Lithuanian Dictionary(1981). vol XIII, Vilnius, Mokslas, p.872-873 (in lithuan.)
15. Pakerys A.(1982). The Prosody of the standard Lithuanian language, Vilnius, Mokslas, -214p.(in Lithuanian).
16. Rabiner L., Shafer R.(1981). Digital processing of speech signals.
17. Sauka D.(1986). Lithuanian folklore, Vilnius, Vaga, -317p.(in Russian).
18. Sinkevičiūtė B., Sondeckis S., Medonis A.(1986). On the identification of violin strokes in a real-time performance system. Proceedings of the XII ICMC, 187-191p.
19. Zue V.(1985). The Use of Speech knowledge in Automatic Speech recognition. Proceedings of the IEEE, v.73, No 11,p75-91

APPENDIX (CONCLUSIONS)

The main points of the report are:
1) discussion of the problem of a searching for the common and distinctive features of an automated investigation of music/speech and simple speech signals on the basis of Lithuanian musical folklore (see note ex I)
2) parameters of the musical and speech processing are closely interconnected; therefore their common interpretation makes it possible to estimate musical duration of sounds more precisely in case their height is the same;
3) an exact definition of a term "musical syllable" is a topical problem in the aspect of automated ciphering of melodies. There are same specific features of music/speech sounds: sometimes consonants are in agreement with independent musical height and duration (note example I), necessary to express; another well-known case deals with the singing of some melodic fragment performed by one and the same vowel. Both the the specific elements are distinctive features of an automated investigation of music/speech and simple speech signals;
4) there are some common features of the above signals:
   a) acoustic parameter - pitch, intensity and duration - are most importance for both types of signals,
   b) a necessity of diphtong segmentation in music/speech and simple speech processing (Note ex.I /d/);
5) direct application of pure mathematical methods are not sufficient in music/speech signal processing.