# TRACKING MONOPHONIC MUSIC FOR MODELLING MELODIC SEGMENTATION PROCESSES

Maja Serman, Dept. of CSIS, University of Limerick, Ireland.
Niall, J.L. Griffith, Dept. of CSIS, University of Limerick, Ireland.
Nikola Serman, Dept. of Power Engineering, FMENA, University of Zagreb, Croatia.

## INTRODUCTION

This paper describes research towards computational investigation and modelling of musical processes across cultures. It is focused on work towards a model of melodic segmentation. Research in music perception points to a key role for Gestalt principles in the basic grouping mechanisms for melodic segmentation (Lerdahl & Jackendoff, 1983; Narmour, 1989). The perception of some difference between regions (Deliège, 1987) is arguably the underlying principle of all grouping mechanisms. However, what constitutes a region and what guides our perception when differentiating regions? Usually, this question is simplified by the definition of melodic descriptors and grouping mechanisms in terms of Western Tonal Music (WTM). For instance, the melodic descriptor of pitch has 12 semitone values per octave, and it is within this space that low or high frequencies, or small or big intervals are defined. The Grouping Preference Rules (GPRs) proposed by (Lerdahl & Jackendoff, 1983) are an example of this approach. GPRs operate on a sequence of notes, and anchor the development of differentiation in any melodic descriptor to the percept of pitch change.

While there is little doubt of the importance of pitch change in segmentation processes, in non-western music other melodic descriptors can also carry perceptually significant changes while pitch remains constant (Titon, 1992). It is arguable that by using melodic descriptors defined in terms of WTM we impose perceptual categories before even starting to assess their universality, and the grouping mechanisms involved. WTM is a product of complex perceptual and cognitive processes, strongly influenced by development within western culture.

Using WTM notation for transcribing non-western music is a recognised problem in ethnomusicology (Sachs, 1962). Serman et al, (2000) discusses the influence of the transcription process on segmentation. Several types of analogue devices – known as the Melograph (Crossley-Holland, 1974), were made in the 1950's to address this issue. However, the aim of the device was not music perception research. While the importance of the idea for research into universal aspects of music perception was recognised (Crossley-Holland, 1974), the Melograph's contribution to comparative musicology was ahead of its time and interest in it passed as ethnomusicology developed in other directions. Most subsequent computational models of melodic segmentation rules have used notated melodies (or a MIDI encoding) as their input. The use of WTM notation brings into question cross-cultural investigation of melodic segmentation processes, and confines research to investigating grouping mechanisms that operate solely on changes that can be extracted from notation, i.e. pitch and duration.

## SEGMENTATION FROM PERFORMANCE – *MUSICTRACKER*

Assuming that it is desirable that the information relevant for the computational investigation and modelling of melodic segmentation processes across cultures should be extracted directly from the performed music, the *MusicTracker* project has been initiated. This involves a rather different set of problems to those presented by notation. If we use sound as the input, (as received by the auditory system), we are faced with the need to identify the perceptual structures that are built up from the sound. However, our perception of the descriptors of pitch, loudness and timbre, is not derived from a single parameter obtainable by computational signal analysis (Handel, 1995). Furthermore, in perceiving these descriptors a listener responds not only to the sequence, but is influenced by other aspects of the performance, e.g. the overall loudness of the sound, acoustic properties of the place, etc. Therefore, instead of attempting to quantify the melodic descriptors themselves by computational signal analysis and then map them into (possibly ill-defined) WTM categories, the emphasis in this project has been on estimating the *change* that occurs within a melodic descriptor that can then be used as input to a model of grouping processes. This seems justified (at least initially) by what is known of our nervous and cognitive system's sensitivity to change. The *MusicTracker* takes a digitally recorded music signal (.wav) as its input and calculates the values of "indicators": pitch, perceptive dynamics and timbre. The term "indicator" denotes that the value reflects the change of the corresponding melodic descriptor, rather than quantifying it directly.

### An outline of the *MusicTracker* basic functions

1. The signal is divided into equal chunks of 20ms duration, named "frames" and for each frame the values of "indicators" are calculated relying on the spectral analysis procedure.
2. The sequences of indicator values form discrete functions of time which can be presented graphically and/or can be stored into file for further manipulation,

e.g. for application of GPRs or for use as an input for some segmentation model.

3. The *MusicTracker* operation is controlled by adjustable parameters: pitch range, frequency resolution (from the WTM semitone to a few cents), the number of harmonics to take account of, the dynamic range of the signal, etc.
4. The setting of parameters and rough analysis of the results are supported by a set of "Frame Analysis" facilities.

## Melodic descriptor indicators

**The pitch indicator**: Though the perceived pitch of a musical sound is not completely determined by its fundamental frequency, this frequency is accepted as the pitch indicator of monophonic musical sound. Testing the *MusicTracker* with various musical sounds (including the human voice) has confirmed that the frequency of the first peak in the signal power spectrum can be accepted as an estimate of the fundamental frequency though the well known 'missing fundamental' phenomenon (Rossing, 1990) might call for a more sophisticated algorithm.

**The perceptual dynamics indicator**: Subjective perception of loudness depends mainly on sound pressure, but also varies with fundamental frequency and frequency content of the sound signal (i.e. spectrum) and its duration (Rossing, 1990). In the *MusicTracker*, the influence of the frequency content of the signal is taken into account, while the influence of the critical bandwidth across components and duration is being experimented with. The term *perceptual dynamics* in the *MusicTracker* denotes the ratio of the signal power of a frame and the *minimal perceptual power*. The latter is obtained by the correction of the minimal physical power in the record with respect to the frequency according to a simplified "pianissimo equal loudness curve" derived from Fletcher & Munson (Fletcher & Munson, 1933). The monophonic musical sound signals are comprised of a fundamental and its harmonics, therefore the following assumptions have been made:

A signal is treated as a sum of the pure sinusoidal components, each of them contributing to the total signal power proportionally to its spectral peak height.

Each of the signal components contributes to the perceptual dynamics of the signal proportionally to its individual perceptual dynamics.

In order to eliminate the influence of the pre-set physical dynamic range depending on the sound record quality, the values of the perceptual dynamics indicator of all frames are normalized with respect to the highest value found in the entire sequence. Thus, the perceptual dynamics indicator takes values in the range from 0 to 1.

**The timbre indicator**: Human perception of timbre seems to be influenced by the relative contribution of harmonics in a stationary sound signal, and by the dynamic attributes of onsets and decays of the tone components (Grey, 1977).

At present, the unambiguousness of the timbre indicator has been sacrificed in favour of a simple metric for use in melodic segmentation investigations. It has been defined as the weighted sum of relative heights of spectral peaks, i.e. the relative peak height of each harmonic is weighted by the logarithm (base 2) of the ratio between

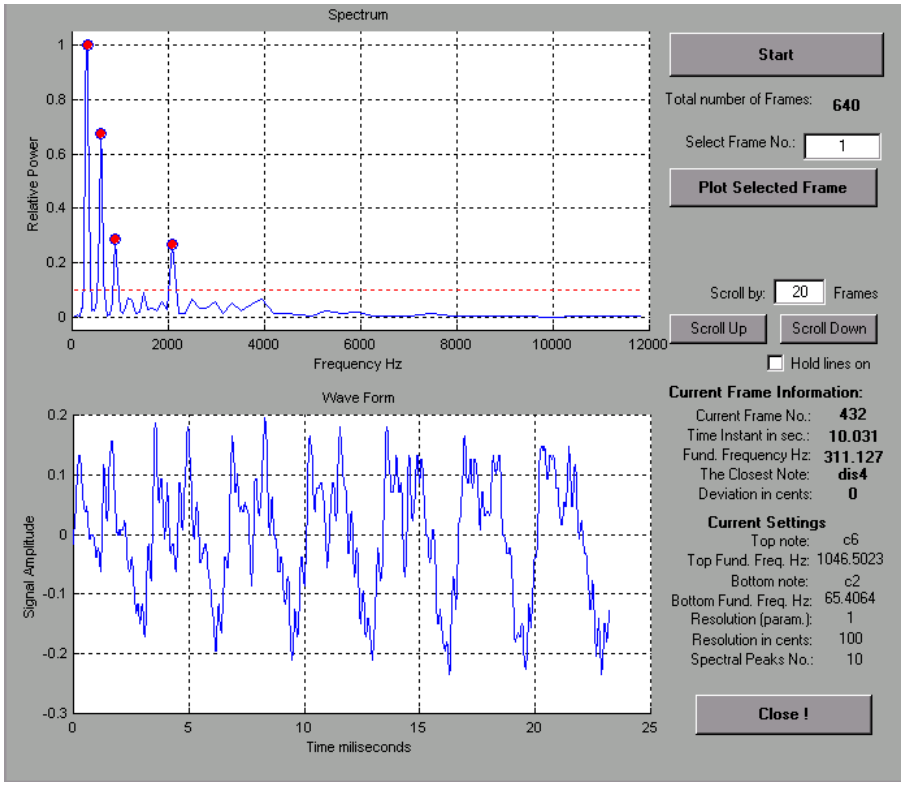


**Figure 1: The *MusicTracker* 'Frame Analysis' display window.**

the frequency of respective spectral peak and the fundamental frequency. In the case of a pure sinusoidal signal the value of the timbre indicator equals zero, and its value increases with the increasing presence of the higher harmonics. The timbre indicator values are finally normalized by the largest one found in the sequence of frames, so that the indicator has values between 0 and 1.

In order to investigate the usability of the unitary timbre indicator we have compared human judgments of timbral changes in monophonic melodies with the results from the *MusicTracker* graphs for two instruments. Six different timbres at the same pitch, commonly used by the shakuhachi players (Dai Shihan T. Inzan, April 2000, personal communication) are recognised by the *MusicTracker*, see figure 3. Secondly, the results from (Bloothooft and Plomp, 1984) experiments on variations in timbre of sung vowels comparing "vowels", "singers", "modes of singing" and "fundamental frequency", were confirmed by the timbre indicator results, these are illustrated in figure 4.

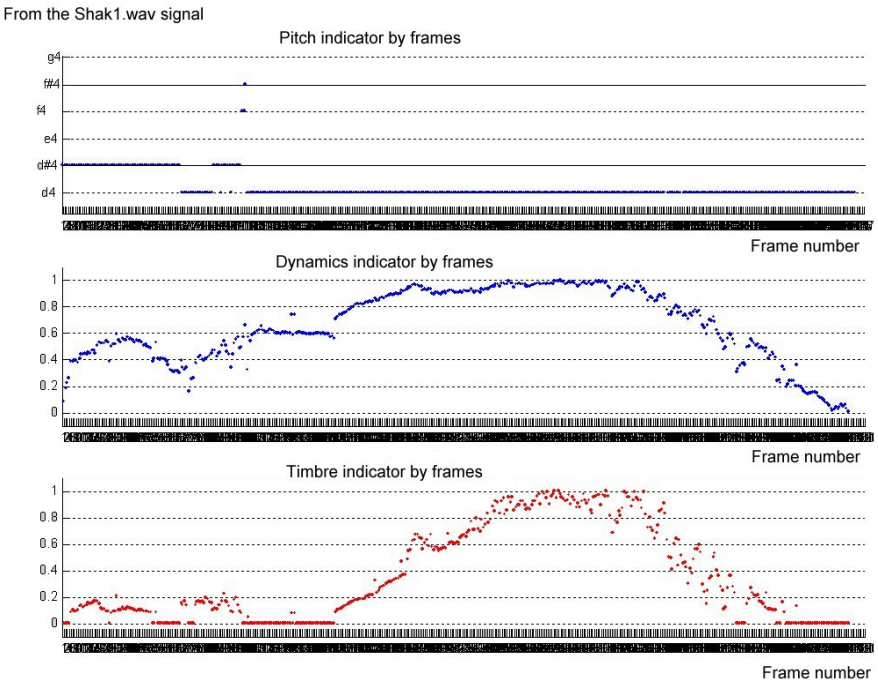


**Figure 2 *MusicTracker* indicators measured from a performance recording.**

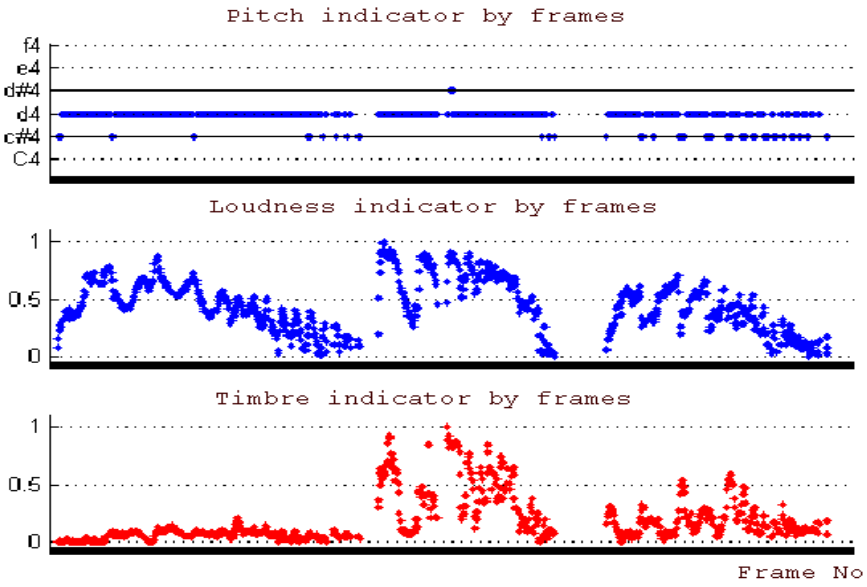


**Figure 3 Three different tone qualities played on a shakuhachi flute.**

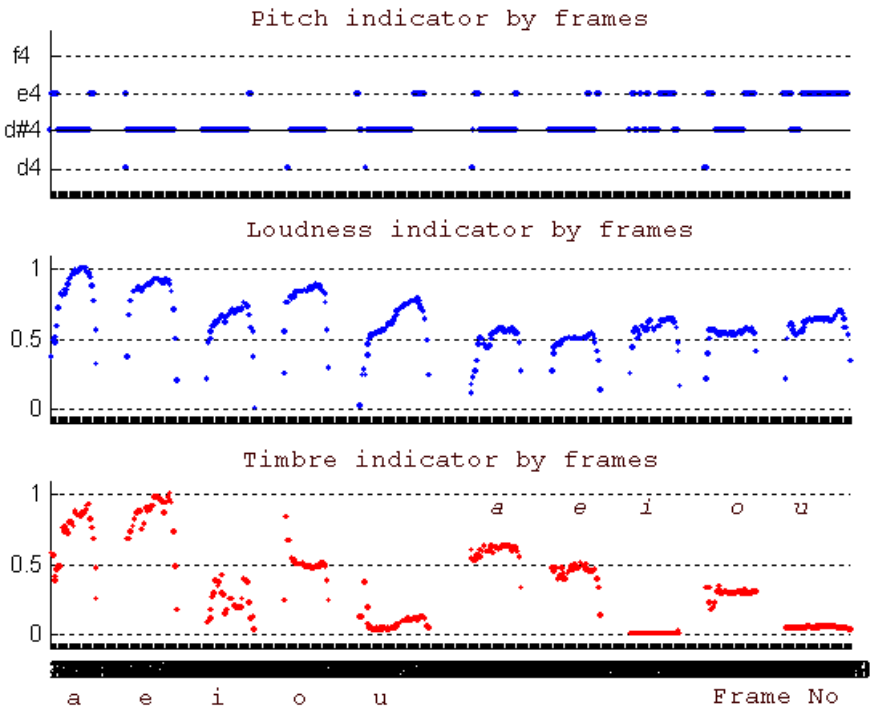


**Figure 4 Timbre measure for two modes of singing vowels – dark (below) & falsetto (italics above).**

## FURTHER DEVELOPMENT

There are three areas in which the *MusicTracker* is being developed. Firstly, the indicators of perceptual dynamics and timbre are defined by simplifications of the signal. This is especially so for the timbre indicator. Further experiments and comparisons will be made to decide at what level of detail the descriptors need to be represented, bearing in mind the goal of modelling segmentation in *monophonic* music. Secondly, the *MusicTracker* is being used to investigate the application of grouping mechanisms proposed by the theories of melodic perception. The *MusicTracker* allows this to be closer to a real performance and also allows comparison with results derived from notational transcriptions. Thirdly, the definition of the indicators as well as changes and the interaction between them are currently being analysed as part of the development of a cross-cultural monophonic segmentation model.

## ACKNOWLEDGMENTS

We would like to thank Ms Helen Arthur, Mr Tomizu Inzan - Master of the Tozan School of Shakuhachi, Mr Kevin Hayes, Mr Albert Llussa and Mr Drazen Loncar for their participation in the experiments, as well as for their knowledge and patience.

## REFERENCES

Bloothooft, G., & Plomp, R. (1984). "Spectral analysis of sung vowels. I Variation due to differences between vowels, singers and modes of singing". Journal of the Acoustical Society of America, 75(4), 1259-1264.

Crossley-Holland, P. (Ed.). (1974). Selected Reports in Ethnomusicology, 2(1)

Deliège, I. (1987). "Grouping conditions in Listening to Music: An Approach to Lerdahl and Jackendoff's Grouping Preference Rules". Music Perception, 4(4), 325-360.

Fletcher, H., & Munson, W. A. (1933). "Loudness, its definition, measurement and calculation". Journal of the Acoustical Society of America, 5, 82-108.

Titon J. T. (Ed.). (1992). *Worlds of Music: an introduction to the music of the world's peoples*. New York: Schirmer Books.

Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres". Journal of the Acoustical Society of America, 61(5), 1270-1277.

Handel, S. (1995). "Timbre perception and auditory object identification". In B. C. J. Moore (Ed.), Hearing, New York: Academic Press, 425-461.

Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. Cambridge, MA: MIT Press.

Narmour, E. (1990). The Analysis and Cognition of Basic Melodic Structures. Chicago: University of Chicago Press.

Rossing, T. D. (1990). The Science of Sound, Addison-Wesley.

Sachs, C. (1962). The Wellsprings of music. The Hague: Martinus Nijhoff.

Serman M., Griffith N. J. L., & Serman N. (2000). "Computational Modeling of Segmentation Processes in Unaccompanied Melodies", In *Proceedings of the 2000 International Conference on Music Perception and Cognition*. Keele: European Society for the Cognitive Sciences of Music.