PITCH CONTOUR ANALYSIS BASED ON DERIVED MODEL

PACS 43.70JT

Micallef Paul ; Sammut Mark Anthony
University of Malta
Dept. Communications and Computer Engineering
Msida
Malta MSD06
Tel: 356-32902520
Fax: 356-343577
E-mail: pjmica@eng.um.edu.mt

ABSTRACT

The paper describes a tool that is used in order to approximate a pitch contour of speech or chant. This is done using shapes called pitch primitives which when suitably concatenated result in an approximate synthesis of the original pitch contour. The approximate contour retains information regarding position and type of pitch primitive. This information can then be used to build statistical information on the original contour, relating types of pitch changes to position. The paper includes results obtained on Islamic and Byzantine chant.

INTRODUCTION

The aim of the work is to analyse the pitch contour or envelope of chant or of speech and to deduce the contour shapes for preferred ranges and rate of changes in the pitch. In this way some general conclusions can be obtained on the chant or on the speaker intonation. The statistical information is gathered using the proposed tool in a fast and efficient way. The statistical analysis can be used for the comparison of intonation of speakers and of languages. The analysis can also be applied to speech like waveforms such as voice recitative with no musical instruments. The pitch information is extracted using standard autocorrelation techniques over nonoverlapping segments of 20 ms., [1].

PITCH PRIMITIVES

In order to be able to look analytically at the intonation contour it is necessary to build a set of pitch primitives. These primitives are basic line shapes. By having a suitable set it should be possible to model the natural intonation contour. The possibility of having a known set of primitives that are concatenated together to obtain the final shape, also makes it possible to analyse the shape by finding the number and position of the primitives used in modelling a chant.
Figure 1 shows a subset of primitives used in this modelling. These shapes are themselves a subject of a statistical analysis based on postulating a given shape and then finding how much it appears within natural intonation. The present set was obtained based on an analysis of a set of spoken English sentences by Maltese speakers. It can be however easily adapted to any speech corpora or chant by adding a particular shape (pitch primitive) if it is necessary due to particular intonation contours in the corpus or in the chant.

In order to keep the size of the primitive set to a minimum, each primitive has two fundamental properties. These are a normalised shape referred in Y-axis relative to the start point as position with Y = 0. A ‘stretch’ of the shape is possible in the Y-axis only so that the slope of the primitive can change, but not its basic shape. Primitives are defined with 3, 5, and 7 points. The present set consists of about fifty primitives.

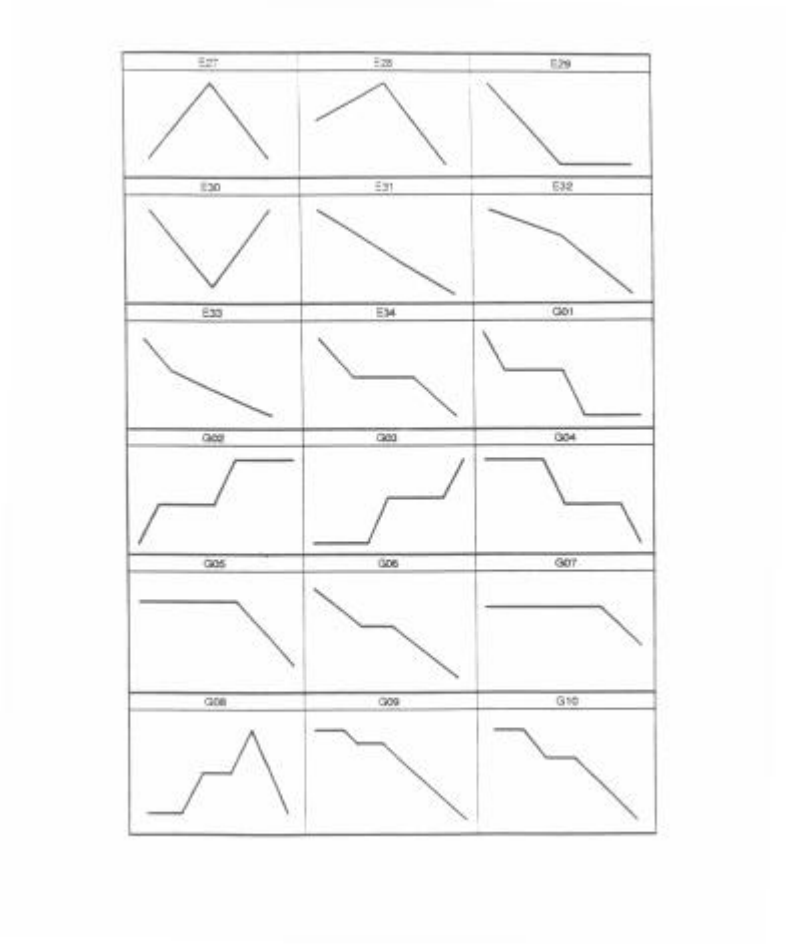

Figure 1

Figure 2 gives a few typical primitive definitions that can be related to some of the shapes in Figure 1. For example primitive E30 is a 5-point primitive with a V-shape. The 5-number definition relates the relative position of the 5 points relative to the start position, which is always 0.

E28	0, 1, 2, 0, -2
E29	0, -1, -2, -2, -2
E30	0, -1, -2, -1, 0
E31	0, -1, -2, -3, -4, -5
E32	0, -1, -2, -4, -6
E33	0, -2, -4, -5, -6
E34	0, -3, -3, -4, -5
G01	0, -1, -1, -1, -2, -2, -2
G02	0, 1, 1, 1, 2, 2, 2
G03	0, 0, 0, 1, 1, 1, 2
G04	0, 0, 0, -1, -1, -1, -2
G05	0, 0, 0, 0, -1, -2, -3
G06	0, -1, -2, -2, -3, -4, -5
G07	0, 0, 0, 0, 0, -1, -2

Figure 2

MODELLING WITH PITCH PRIMITIVES

In modelling, the real waveform is examined from the start and the best fit shape is applied. A weighting is placed on fitting with the longer (ie 7-point) shapes to keep the used set of primitives to a minimum. An error function based on distance of the natural pitch contour from the postulated primitive is used in order to arrive at the choice of the best primitive. Each primitive is examined for shape and stretch to obtain the particular best fit for that shape. This is given by

$$Error = Math.abs (target[i] - start\_pitch + (stretch * values_j[i]))$$

where	target[i] is the i^{th} value (i varying between 1 and 7 for a 7-point primitive) of the natural pitch array being matched;
$values_j$ [i]	value of i^{th} element of the j^{th} pitch primitive relative to start value of 0 for each primitive;
stretch	is the amplification factor ;
start_pitch	is the value of the natural pitch at the start position
Error	is the absolute value of the difference between each natural pitch value and that of the model

If the error is above a fixed threshold another primitive is considered in turn. The primitive with the best fit at a particular stretch is retained. This process is moved forward to the next part of the natural contour. At present there is no overall best error fit, but only a 'running' best error fit. Each primitive is independent of the previous or following. The endpoint of the primitive is not constrained. The next primitive analysis starts from the next real pitch value trying to find the next best fit. Therefore errors are not cumulative though there can be local anomalies in the shape.

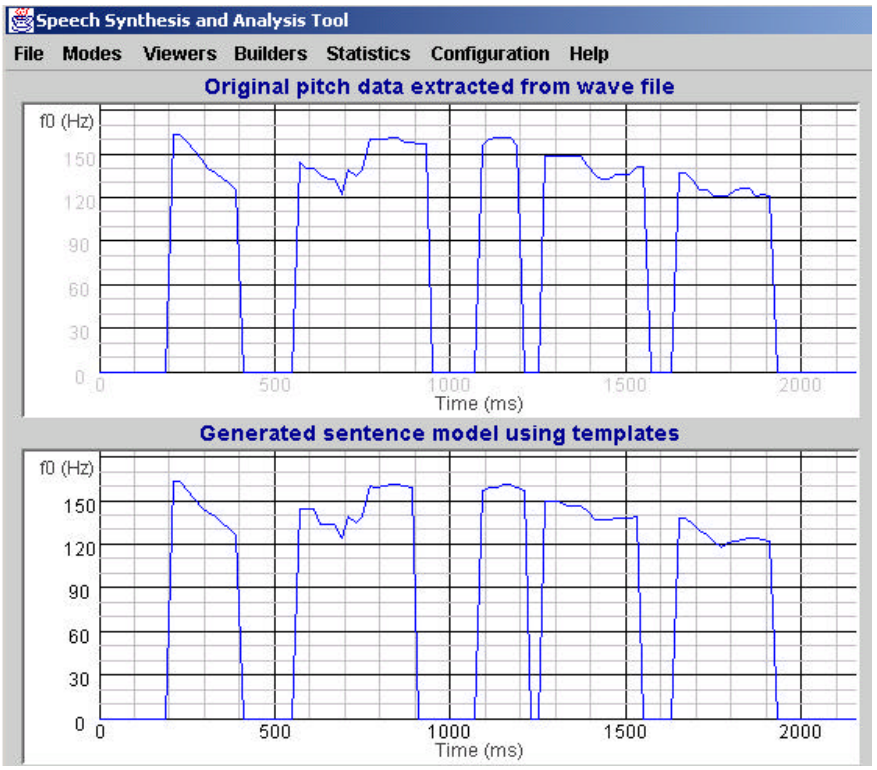

Figure 3 Sentence of continuous speech

Figures 3 and 4 show a natural and modelled intonation contour based on a sentence of continuous speech and part of a Byzantine chant, respectively. The modeled waveforms are quite similar to the natural intonation. As already mentioned any gross anomalies can always be rectified by editing the pitch primitive set. What is important is the fact that the modelled contour is known and defined in detail.

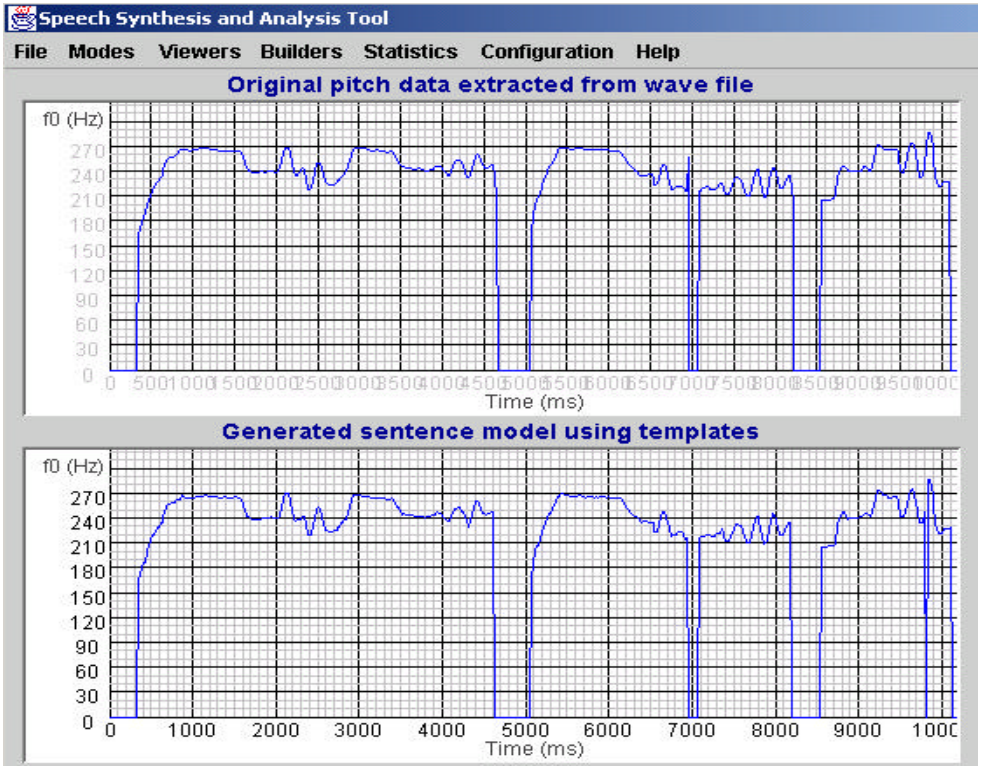

Figure 4 Byzantine Chant

Figure 5 gives a partial entry on the model for the Byzantine waveform. The entries refer to the primitive type, the stretch number, the initial real pitch, which for the purposes of the modelling would be the '0' position of the first pitch point in the primitive, and the start location, within the waveform in milliseconds. This data makes it relatively easy to obtain statistical information on the natural pitch contour based on the modelled contour.Typically 60 seconds of a chant require approximately 450 concatenated primitives to make up the overall shape.

Primitive	Stretch	Start Pitch	Start Time (ms)
E30	3	126.7	1690
G02	3	128.9	1790
G02	2	137.0	1930
E19	5	139.1	2070
E13	5	153.1	2170
E12	5	168.3	2270
E13	6	151.0	2370
E33	3	168.3	2470
G02	1	151.0	2570
G20	1	152.6	2710
G02	1	152.1	2850

Figure 5 Part of Modelled Contour for Figure 4

STATISTICAL ANALYSIS

The tool has been applied to continuous speech and to chant. Figure 6 gives the statistics of the most used primitives for a set of sentences spoken first as statements and then as questions. In this case an analysis of the most used primitive types shows that the percentage of primitives with a positive gradient and with a negative gradient are quite similar in a question, while the percentage of negative gradient primitives is considerably more in the statement. This is in line with the usual rise in the pitch at the end for questions, and the declination profile for the intonation of a statement, [2].

Statement			**Question**		
Pitch Primitive Type	**Percentage**	**Gradient of Pitch Primitive**	**Pitch Primitive Type**	**Percentage**	**Gradient of Pitch Primitive**
G01	8.06	negative	G01	7.16	negative
G02	5.65	positive	C08	5.43	positive
C08	5.38	positive	G03	4.69	positive
C09	4.57	negative	G02	4.44	positive
E06	4.30	negative	C02	4.20	neutral
G03	4.30	positive	E04	4.20	neutral
E13	3.76	positive	C09	3.95	negative
G09	3.49	negative	C04	3.70	positive
E04	3.23	neutral	E06	3.70	negative
E20	3.23	negative	E13	3.46	positive
G04	3.23	negative	G04	3.46	negative
G07	3.23	negative	G11	3.46	negative
G11	3.23	negative	C01	2.96	positive
Total % negative	33.3		Total % negative	21.7	
Total % positive	14.8		Total % positive	24.7	

Figure 6

The statistics on the stretch parameter are also indicative of the register range for a given waveform. Figure 7 shows a typical result obtained for primitive E05 when the tool was applied to an analysis of Byzantine Chant and Islamic chant, [3]. Note that there is a much more consistent use of higher stretch numbers in the Byzantine than in the Islamic again indicating the wider use of higher pitch range changes. The tool enables the statistical characterization of the register ranges in the waveforms.

Byzantine Chant		Islamic Chant	
Stretch Factor	Percentage of occurences	Stretch Factor	Percentage of occurences
1	0.4		
2	2.2	2	1.8
3	4.3	3	5.4
4	9.9	4	26.8
5	13.4	5	17.9
6	15.5	6	17.9
7	13.4	7	10.7
8	10.8	8	8.9
9	6.9	9	7.1
10	9.9	10	1.8
11	4.7	12	1.8
12	2.6		
13	3.9		
14	1.3		
15	0.4		
18	0.4		

Figure 7 Range of Stretch Factors for Pitch Primitive E05 from the analysis of Byzantine and Islamic chant

CONCLUSION

The analysis tool is able to give quick and accurate analysis of changes in intonation of speech or chant and the relative position within the time wavefom of the particular shape under investigation. This is achieved by modeling the natural waveform intonation pattern by a set of pitch primitives. The tool is at present being used to look at the intonation pattern of spoken Maltese to develop an intonation model for a Maltese speech synthesis tool.

ACKNOWLEDGEMENTS

Mr. David Attard, Mr. Alex Chircop, Ms Gillian Attard, Mr Aaron Culotta, Department of Computer Science, University of Malta, for the JAVA programming of the tool.
This work was supported by the EU under Fifth Framework INCOMED Project CAHRISMA.

REFERENCES

1. "COLEA: A MATLAB software tool for Speech Analysis", www.utdallas.edu/~loizou/speech/colea

2. "Prosodic Analysis of Natural Speech" in "Progress in Speech Synthesis", van Santen J. P., Sproat R.W., Olive J. P., Hirschberg J. editors, Addison Wesley 1996

3. Kerabiber Z et al, Anechoic Recordings of Islamic and Byzantine Chant, May 2000.