A SCORE-TO-SINGING VOICE SYNTHESIS SYSTEM FOR THE GREEK LANGUAGE

Varvara Kyritsi
University of Athens,
Department of Informatics and Telecommunications,
Athens, Greece
grad0554@di.uoa.gr

Anastasia Georgaki
University of Athens,
Department of Music Studies,
Athens, Greece
georgaki@music.uoa.gr

Georgios Kouroupetroglou
University of Athens,
Department of Informatics and Telecommunications,
Athens, Greece
koupe@di.uoa.gr

ABSTRACT

In this paper, we examine the possibility of generating Greek singing voice with the MBROLA synthesizer, by making use of an already existing diphone database. Our goal is to implement a score-to-singing synthesis system, where the score is written in a score editor and saved in the midi format. However, MBROLA accepts phonetic files as input rather than midi ones and because of this, we construct a midi-file to phonetic-file converter, which applies irrespective of the underlying language.

The evaluation of the Greek singing voice synthesis system is achieved through a psycho acoustic experiment. The thirty participants are given two samples and are asked to evaluate them according to vocalness, naturalness, intelligibility and expressivity, in a 1-5 scale MOS test. The results are encouraging, with naturalness and expressivity being ranked above 2.5 in average. Thereafter, we create a new diphone database, from the same text corpus but with a different human voice. The text is recited at constant pitch. After repeating synthesis and evaluation, the quality of the produced voice is evidently improved, with naturalness and expressivity being ranked above 3 in average.

1. INTRODUCTION

The last thirty years, there has been a special research interest on the synthesis of the singing voice. With the term singing voice synthesis (SVS) we mean the production of human-like singing voice by a computer. The input data for a synthesis program are usually a score in some standard format and lyrics given in ordinary text or phonetic notation. SVS systems are classified into two main categories: the ones which use a waveform synthesizer and the ones which use the modification and concatenation of recorded units of natural singing voice. [15].

Singing voice as well as speech originates from the voice instrument. However, they differ significantly in terms of their production and perception by humans. Speech production aims at the exchange of messages, whereas singing voice production aims at the use of the voice as a musical instrument. In singing, vowels are often sustained much longer than in speech, the pitch range is wider and higher and the musical quality of the voice is more critical than the intelligibility of the lyrics [15], [11]. Consequently, a SVS system must include control tools for various parameters, as phone duration, vibrato, pitch, volume etc.

Since 1980, various projects have been carried out all over the world, through different optical views, languages and methodologies, focusing on the mystery of the synthetic singing voice [10], [11], [13], [14], [17], [18], and [19]. The different optical views extend from the development of a proper technique for the naturalness of the sound quality to the design of the rules that determine the adjunction of phonemes or diphones into phrases and their expression. Special studies on the emotion of the singing voice have not carried out.

In this article, we develop a score-to-singing voice synthesis system for the Greek language, which is not the first attempt that has been made to synthesize Greek singing voice: in the frame of the AOIDOS research program, developed in the Department of Computer science, University of Athens, a comparative study between Byzantine singing and bel-canto has taken place [26].

The rest of the paper proceeds as follows: in the second paragraph, we briefly highlight the on going research on singing voice synthesis, in order to show the innovation of our system. In the third paragraph, we concentrate on specific MBROLA features, which we have chosen among other techniques. Moreover, we describe the existing Greek diphone database and the creation procedure of a new database. In the fourth paragraph, we present our SVS system, focusing on the midi-file to phonetic-file converter. In the last paragraphs, we evaluate our SVS system and suggest ideas for further research on the synthesis of the Greek singing voice.

2. THE CURRENT RESEARCH ON THE SYNTHSIS OF SINGING VOICE

2.1. Synthesizer models and techniques

Three synthesizers, already established in the world of computer music research, are: MUSSE/RULSUS [3], CHANT [2], [12] and SPASM/SINGER [4]. Recently, new synthesis systems as FLINGER, LYRICOS [11], VOCALWRITER, VOCALOID [17], CANTOR [33],

MBROLA [5], [8], [23], MAXMBROLA [23] extend the idea of usability and performability of computers.

All of the above models differ in their synthesis technique. During the last twenty years, the development of flexible synthesis techniques has become a special research focus, in order to solve the code of *naturalness* and *vivacity* of the singing voice in the lower or higher frequencies during singing. Techniques like FM synthesis [1], the formant model [3], [6], the FOF synthesis [2], [15], the physical modeling [4], the synthesis by concatenation of sampled sounds [9], [12], [19] are some of the highest popularity in the world of computer music.

2.2. The complexity of the voice signal

Attempting an evaluation to the research models, mentioned above, the technical problems are related to the complexity of the vocal signal and more specifically to:

a) The huge quantity of parameters and data, describing the complex voice singing model, related to the incapacity of the machines to elaborate them satisfactorily (for example, in order to have an entire command of the Greek language we must sample about 2300 sound units, just for one type of voice only). This is one of the major cues for differentiating the vocal signal from an ordinary instrumental signal, as the voice is closely connected to the human being, not only from the physiological point of view, but also from the acoustic one. In other acoustic signals, it is not necessary, in the same detailed manner to describe the formant trajectories or the microvariations of the signal (which in the case of the voice, are related very closely with the biological function of the vocal apparatus).

b) The specifity of the voice concerning biological functions of the human body (organic and physiological) which affect the timbre, the intensity and the articulation of the voice. For example, aleatoric microvariations, due to the stress or other factors, influence the periodicity of the vocal signal.

c) The fact that every language has its proper phonetic rules and phonemes, renders very difficult the creation of an international phonetic database (a large vocabulary of phonemes and diphones in several languages and an interdisciplinary connection between them), hindering the commercialization of a universal vocal synthesizer.

2.3. The necessity of using concatenative synthesis in SVS

The concatenative synthesis, which was initially developed for speech synthesis, has now been applied to singing voice synthesis, due to the large improvement of quality in the produced voice [15].

Concatenative synthesis is based on the concatenation (or stringing together) of segments from a database of recorded speech. A main advantage of the concatenative synthesis is that different voices can easily be produced by means of different databases. On the other hand, the dependency of the output voice on the choice of database can turn out to be a serious limitation.

There are three main sub-types of concatenative synthesis: unit selection synthesis, diphone synthesis and domain-specific synthesis [35]. Diphone synthesis uses a minimal speech database that contains all the diphones for the target language. A diphone is a speech segment, which starts in the middle of the stable part of a phoneme and ends in the middle of the stable part of the next phoneme. When the diphone is used as the basic unit, the concatenation points are at stable parts of the phonemes. This facilitates some smoothing operation to be performed at synthesis time, which reduces possible discontinuities at concatenation points.

3. THE MBROLA PERFORMABILITY

MBROLA (Multi-Band Re-synthesis OverLap Add) is a widely used speech synthesizer, based on the concatenation of diphones. MBROLA takes as input a phonetic file, giving the list of phonemes with some prosodic information (duration/pitch), and outputs an audio file containing 16-bit linear samples at the sample rate of the diphone database [8], [23]. The audio output file format may be Raw, Wav, Au, or Aiff.

The extension of phonetic files is ".pho" and their format is very simple, as listed in Fig. 1. Each line begins with the name of a phoneme, followed by the duration in milliseconds and optionally, followed by one or more pitch points. Each pitch point is a pair of two values: relative position of the pitch point (in percentage of the phoneme duration) and pitch value (in Hertz). Pitch points define a piecewise linear intonation curve.

```
; bonjour
_ 51 25 114
b 62
o  127 48 170
Z 110 53 116
u 211
R 150 50 91
_ 9
```

Figure 1. Phonetic Files.

Although, three pitch points per phoneme are generally sufficient for the production of good-quality speech, the MBROLA synthesizer takes up to 20 pitch points per phoneme, which allows the reproduction of vibrato and portamento in singing voices and consequently the synthesis of acceptable singing voices.

Until now, only the Swedish used the MBROLA synthesizer for singing voice production [16]. The result was not so natural, because their diphone database was derived from spoken language. Among the various diphone databases (more than fifty) freely provided by the MBROLA project's web site [32], there are two databases[1] for the Greek language.

[1] We will examine only the GR2 case.

The main reasons why we choose the MBROLA synthesizer among others are:

a) The existence of diphone databases for the Greek language.

b) It is a text-to-speech (TTS) synthesizer, which means that it converts normal language text into speech (singing voice in our case), instead of rendering symbolic linguistic representations like phonetic transcriptions into speech.

c) It allows the synthesis of acceptable singing voices.

d) It is commercially available.

At this point, we would like to mention that apart from the MBROLA synthesizer, we have the alternative to use the MaxMBROLA external object. MaxMBROLA is a Max/MSP object, which is based on the MBROLA synthesizer and allows the real time synthesis of both singing voice and speech. Because real time synthesis is outside the scope of our work, we prefer to use the MBROLA synthesizer.

3.1. Greek diphone database GR2

A diphone database contains all the diphones of a language. The creation of a diphone database is achieved in three steps: creating a text corpus, recording the corpus and segmenting the speech corpus [8], [32] (see &3.2).

UoA-TtS-PA	SAMPA	Example	Hellenic transcription (ISO8859-7)(English)
_	_	(silence)	(παύση) (pause)
[consonants]			
p	p	patAta	πατάτα (potato)
b	b	balOni	μπαλόνι (baloon)
t	t	tirOpita	τυρόπιττα (cheesepie)
d	d	dInome	ντύνομαι (get dressed)
k	k	kalAmi	καλάμι (cane)
c	c	cerI	κερί (candle)
g	g	gremIzo	γκρεμίζω (blast)
q	gj	aqelIa	αγγελία (announcement)
f	f	fotinO	φωτεινό (luminous)
v	v	vuLAzo	βουλιάζω (sink)
T	T	Talassa	θάλασσα (sea)
D	D	DAskalos	δάσκαλος (teacher)
s	s	salAta	σαλάτα (salad)
z	z	zoGraficI	ζωγραφική (paint)
G	G	GAla	γάλα (milk)
j	jj,j	jortI, vjEno	γιορτή, βγαίνω (celebration, go out)
x	x	xarUmenos	χαρούμενος (happy)
C	C	CEri	χέρι (hand)
m	m	mATima	μάθημα (lesson)
M	mj	apaneMA	απανεμιά (calm)
n	n	nanUrisma	νανούρισμα (lullaby)
N	nj	NaurIzo	ναουρίζω (meow)
V	(-)	aVgaLAzo	αγκαλιάζω (bosom)
r	r	ropI	ροπή (torsion)
R	r	tRopI	τροπή (turn)
l	l	lAva	λάβα (lava)
L	lj	LOno	λιώνω (melt)
S	ts	SalakOno	τσαλακώνω (crumple)
Z	dz	ZamarIa	τζαμαρία (glass)
X	ks	XirAfi	ξυράφι (razor)
Y	ps	YAri	ψάρι (fish)
[vowels]			
a	a	aEras	αέρας (wind)
e	e	elpIDa	ελπίδα (hope)
i	i	irIni	ειρήνη (peace)
o	o	Oros	όρος (clause)
u	u	uranOs	ουρανός (sky)

Figure 2. UoA-TtS-PA alphabet.

A Greek diphone MBROLA database was produced (GR2) at the University of Athens from speech recordings and, primarily, for speech synthesis, consisting of 1081 diphones [36]. These diphones originate from a male voice. For the construction of GR2, the UoA-TtS-PA alphabet was used [31] (see Fig. 2). UoA-TtS-PA inherits elements from the SAMPA alphabet [31]. SAMPA (Speech Assessment Methods Phonetic Alphabet) is a phonetic alphabet, which was created by an international group of phoneticians and applied in many languages of the European community, including the Greek one. UoA-TtS-PA extends SAMPA by adding some phoneme groups in clusters. This way, the phonetic representation of text-to-speech synthesizers is enlarged, because complicated co-articulations are faced as single groups, rather than as the concatenation of discrete phones.

3.2. Creation of a new diphone database GR3

As proven by the experimental results (see &5.1.1), when the GR2 diphone database is used for the production of Greek singing voice, the result is clearly acceptable, but not satisfactory enough to be characterized as "natural". The main reason is that synthesis-pieces derive from speech recordings, rather than singing-voice ones. Additionally, the voice of the speaker is unfavorable to SVS (weak formants, bad timbre). As a result, there is a need for a new database.

3.2.1. Creating the text corpus

We use the same text corpus with the GR2 diphone database. We first form a list of all the phonemes of the Greek language. When a complete list of phones has emerged, a corresponding list of diphones is immediately obtained, and a list of words is carefully built, in such a way that each diphones appears at least once [32].

3.2.2. Recording the corpus

Our initial goal is to produce Greek popular music. We choose a professional male singer with strong formants. The text corpus is recited with the most monotonic intonation possible; it is digitally recorded and stored in digital format (22050 Hz). The result of singing this text corpus will be really disappointing, because the MBROLA's resynthesis procedure is performed in constant pitch [23], [32].

The experimental recordings were performed in an anechoic studio using a Rode K2 microphone and digitizing at 44.100 Hz with a resolution of 16 bits using model Yamaha AW16G professional audio workstation.

3.2.3. Segmenting the speech corpus

Once the corpus is recorded, all diphones are spotted automatically via the Diphone Studio tool [32]. The tool's output is a segmentation file, which is sent to the MBROLA team; the latter "mbrolizes" the database.

4. ARCHITECTURE OF THE GREEK SVS SYSTEM

The Greek SVS system consists of a score editor – Sibelius, a midi-file to phonetic-file converter, which applies irrespective of the underlying language, the MBROLA synthesizer and, initially, the GR2 diphone database (see Figure 3).

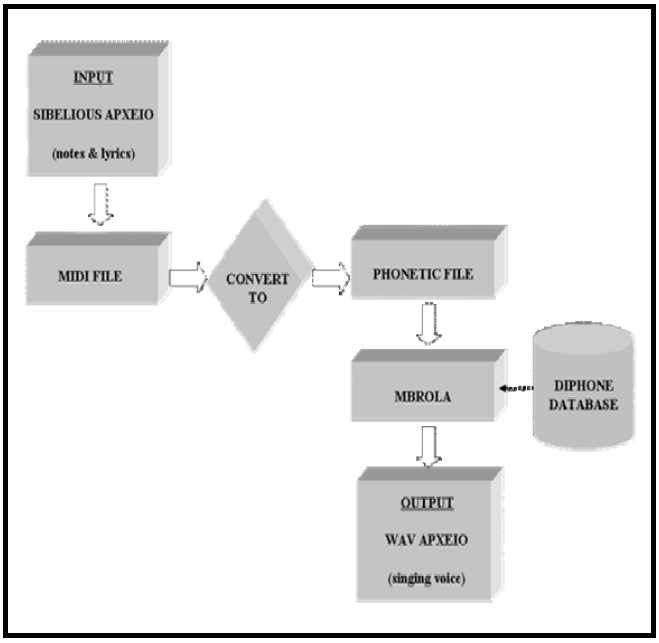

Figure 3. Production of Greek singing voice with the MBROLA synthesizer.

4.1. Functional description of the system

The system's input is a score. A score specifies the notes and the lyrics of a musical piece and is edited in a score editor. We use the Sibelius editor, because it allows the storage of the score in a midi file. However, midi files are not a possible input of the MBROLA synthesizer. They must therefore be transformed into phonetic files.

To date, the construction of phonetic files was a very time-consuming task. We construct a midi-file to phonetic-file converter to automate the conversion from midi files to phonetic. The user is only asked to choose a score, edit it in a Sibelius file and save it in a midi file.

The newly created phonetic files are being loaded on the *mbredit* tool, which is provided by the MBROLA project's web site [32]. The *mbredit* tool allows the modification of the average pitch, the fundamental frequency and the average duration. It also specifies the output format. After the specification of the output format and the completion of any modifications, the MBROLA synthesizer produces the singing voice.

4.2. Midi meta-events

Apart from the midi events, a midi file may also contain non-midi events. These non-midi events are called meta-events. The state byte of a midi meta-event has the hexadecimal value FF, which specifies its type.

When a midi file is inserted into a sound generator device, the listener listens only to the melody of a song, not to the lyrics. Lyrics are not considered to be midi events. However, if someone desires to include lyrics in a midi file, he or she can achieve this, due to the fact that the midi file format allows the insertion of lyrics in a midi file, in the form of meta-information. In other words, the lyrics of the songs are midi meta-events.

A lyric meta-event (FF 05 len text) includes only one syllable, which is expressed in the "text" parameter. The length of the syllable is expressed in the "len" parameter. A lyric message is always followed by a *Note On* and a *Note Off* message, which indicate the activation of the corresponding to the syllable note and the release of the same note, respectively.

4.3. MIDI-file to phonetic-file converter

The program takes as input a monophonic midi file of type 1 and transforms it into a phonetic file. The necessary data for the creation of the phonetic file are the phonemes, the durations of the phonemes and the pitches of the phonemes. These data are included in the *lyric*, *Note On* and *Note Off* midi messages. Subsequently, we must identify these messages and obtain the contained information.

The main routine of the program is called for every track of the input midi file and scans the tracks via an iterative loop. In each iteration, it searches for a *lyric* message, followed by a *Note On* and a *Note Off* message. The syllable is obtained from the *lyric* message, while the note number and the delta-time are obtained from the *Note On* and *Note Off* messages. The phonemes of the syllables will range from 1 to 5, due to a program's restriction. The obtained data can not be inserted into the target phonetic file as they are. We may know the phonemes, but the duration and the pitches of the phonemes are still unknown.

The phoneme's duration depends on the lengths of the lyric messages. There are two cases for a lyric's message length: to be equal to 1 or greater than 1. In the first case, the syllable is consisted of one phoneme and the duration of the phoneme is equal to the duration of the corresponding *Note On* message. In the second case, the syllable is consisted of more than one phoneme (2-5) and the duration of each phoneme is calculated as follows:

$$d_{mid\_pho} = d_{syll} / (5*(N_{pho}-1) \quad (1).$$

$$d_{last\_pho} = (4 / 5) * d_{syll} \quad (2).$$

In (1), N_{pho} is the number of the syllable's phonemes, d_{syll} is the duration of the syllable and d_{mid_pho} is the duration of the intermediate phonemes. In (2), d_{last_pho} is the duration of the last phoneme.

The duration of the phonemes is specified by the delta-times of the *Note On* messages. The format of the phonetic files imposes the duration of the phonemes to be expressed in milliseconds, so delta-times are transformed into milliseconds as follows:

duration (in ms) = delta-time * tempo / PPQN (3).

In (3), tempo is the percentage of time in micro seconds per quarter note and the PPQN is the number of pulses per quarter note.

4.3.1. Long Vowels

As it is already known, the MBROLA synthesizer is a speech synthesizer, which additionally can be used for the synthesis of singing voice. However, a speech synthesizer cannot maintain long vowels. This is an obstacle for the production of good quality singing voices. We solve this problem by dividing the phonemes of long duration (>256 ms) into various phonemes of smaller durations (200 – 300 ms). In general, it is a bad strategy to use phonemes of long durations. Even when they are divided into smaller ones, the quality of the result is bad.

4.3.2. Portamento

Portamento is the iteration of the same phoneme in different pitches. When portamento is met, the phoneme is written only once in the phonetic file. In subsequent iterations, the durations are added to obtain the total duration of the phoneme. The pitches, in which the phoneme is iterated, constitute the pitch points of the phoneme (up to 20).

4.3.3. Pauses

The delta-times of the *Note Off* messages are necessary for the calculation of the pauses' duration. The pauses always follow the *Note Off* messages. If the delta-time of a *Note Off* message is positive, there is a pause. In this case, the phoneme of silence "_" is written in the phonetic file. The duration of the silence-phoneme will be equal to the duration of the *Note Off* message.

The midi-file to phonetic-file converter is implemented via a software program, written in the C programming language.

5. SYSTEM EVALUATION

The Greek SVS system is tested to synthesize two famous Greek songs, the "Hartino to Feggaraki" and the "Aghia Nychta". The "Hartino to Feggaraki" song is a popular song written by the famous Greek composer Manos Hadjidakis. The "Aghia Nychta" song is the wide known "Silent Night" ("Stille Nacht"), a traditional and popular Christmas carol. The corresponding Sibelius files are shown in Figures 4, 5.

5.1. Psychoacoustic Experiments

The evaluation of the Greek SVS system is achieved through a psycho acoustic experiment. We adopt a test of MOS (Mean Opinion Score) to obtain subjective assessments for our system. The score was set to range from 1 through 5.

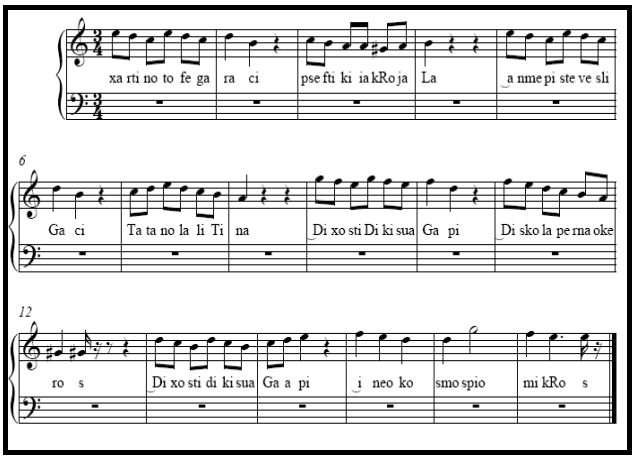

Figure 4. The "Xartino to Feggaraki" Sibelius File.

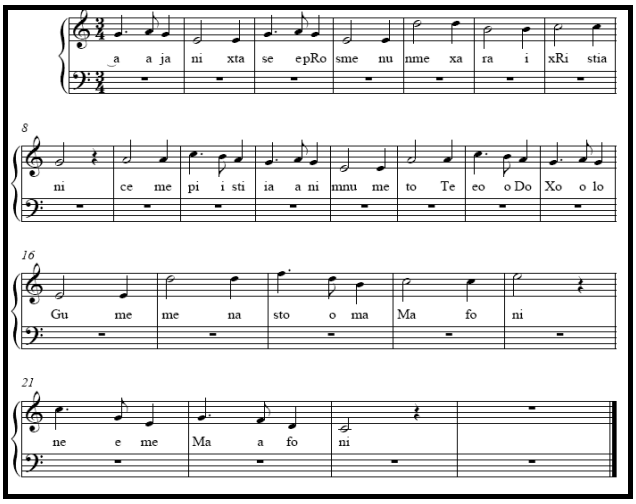

Figure 5. The "Aghia Nychta" Sibelius File.

The thirty participants, between 20 and 50 years old, who have no acquaintance with synthetic voice hearings, are given two samples and they are asked to evaluate them according to vocalness, naturalness, intelligibility and expressivity.

Expressivity is the capacity to express something. It can be an emotion, a sentiment, a message and probably many other things [20]. Naturalness is the attribute which quantifies the possibility of a voice heard to be a human real voice or a synthesized one. Vocalness can be defined as a signature of the voice, a vocal identification factor, which let us say if what we hear is a voice (natural or synthetic) or not. Intelligibility is the ability to comprehend the lyrics.

5.1.1. Using the existing diphone database

The *Mbredit* tool composes the aforementioned Greek songs by making use of the existing GR2 diphone database. In Figures 1, 2, we present the results.

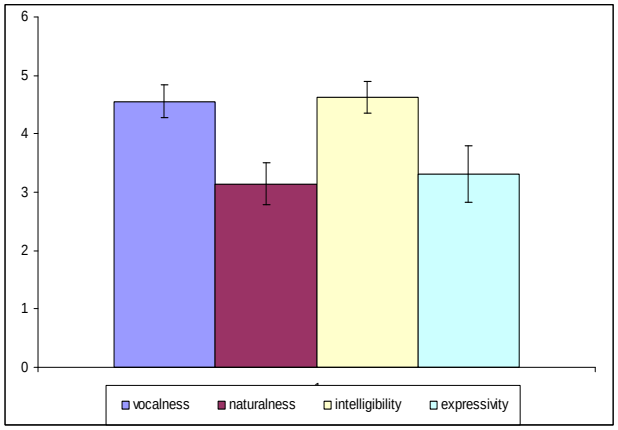

Figure 6. "Xartino to Feggaraki" sample.

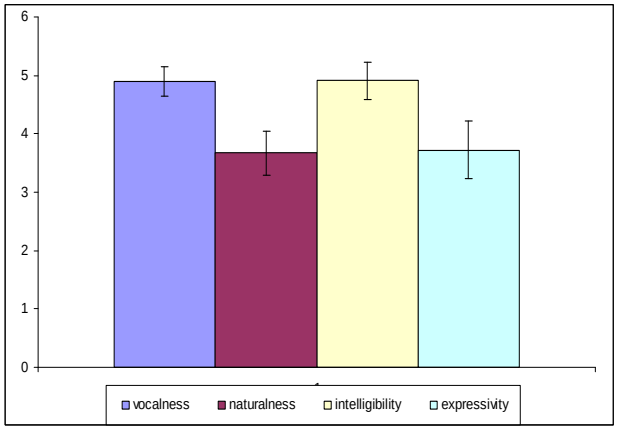

Figure 8. "Xartino to Feggaraki" sample.

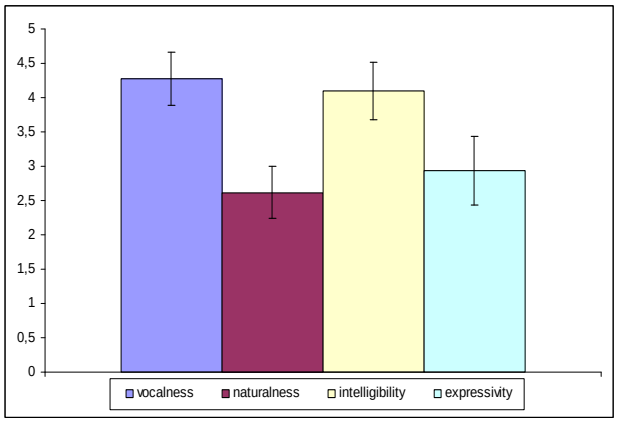

Figure 7. "Aghia Nychta" sample.

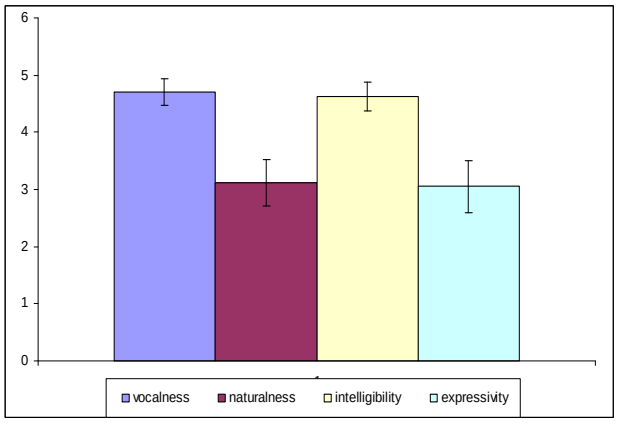

Figure 9. "Aghia Nychta" sample.

Participants ranked naturalness above 2.5 in average. The vocalness and the intelligibility were ranked above 4 in average, whereas the expressivity above 2.5 in average. The produced voice is vocally intelligible, but it suffers from the robotic sound effect.

5.1.2. Using the new diphone database

The Mbredit tool composes the same Greek songs by making use of the new GR3 diphone database this time. The psycho acoustic experiment is repeated with the same auditors, showing a clear improvement in the quality of the produced voice (see Figures 3, 4). In particular, the levels of vocalness and intelligibility are high, ranked above 4.5 in average. Naturalness and expressivity were ranked above 3 in average, implying that the deviation between synthetic and natural voice has diminished. However, the robotic sound effect persists and the accent sounds quite foreign. This is mainly attributed to the automated techniques used in diphone synthesis.

6. FURTHER RESEARCH

The musical cultural heritage of Greece is as diverse as its history. In the future, we shall experiment with different kinds of Greek music, demanding voice techniques other than this of popular singing. We plan to synthesize a) Byzantine music of the 15th century, using data extracted from the AOIDOS project [26], b) Greek ancient music, using the modal system of Mr. Stelios Psaroudakis [27], c) Greek folk music of 17th-19th centuries and d) Greek urban popular music. This work will take place in collaboration with the musicologists of the University of Athens.

In order to produce different kinds of Greek singing, we shall record both male and female voices, with different voice characteristics (formants, pitch, timbre, vibrato etc). The way in which the recordings will be carried out is a really important issue. The text corpus must be recited at the most monotonic intonation possible, slowly and clearly. As a consequence, the singers will have to be very attentive with the pitch and the articulation.

In order the users to be able to experiment with various diphone databases, we are going to develop a user

friendly graphic interface. Additionally, taking inspiration from the MIDI-MBROLA musical application [23], we plan to develop a similar musical application, based on the MaxMBROLA external object. This time, events from a midi accordeon, PHONODEON [25], will be used to compute the prosody in real time.

Last but not least, we would like to examine the perspective of creating a vocal synthesizer, which will be able to reproduce voices of good quality that can be extended in several registers and treated by different techniques. In general, our vocal synthesizer will be able to:

a) reproduce, not only a wide palette of Greek timbres (by the means of diphones), but also cover a wide range of registers whilst preserving timbre homogeneity between them.

b) be equipped with the proper rules of different vocal techniques and the appropriate modal musical systems (ancient Greek modes, ecclesiastic modes, etc).

c) have the possibility to combine singing techniques and other languages (for example combine the vocal technique of Byzantine singing with Latin language).

d) be used not only like a studio instrument, but also as a performance one (like the analog and digitals synthesizers).

7. CONCLUSION

There is an increasing research interest around the synthesis of singing voice, which can have diverse applications in the field of music synthesis. In this paper, we present a score-to-singing voice synthesis system for the Greek language. The system relies on the MBROLA synthesizer and an already existing Greek diphone database, produced for speech synthesis [36]. Subsequently, we create a new diphone database, using the same text corpus, but a different human voice. The text is recited at constant pitch. After having our system evaluated, results show significant improvement of the voice, produced with the new database. More specifically, in a 1 to 5 scale, naturalness and expressivity are ranked above 3 in average.

8. ACKNOWLEDGEMENTS

The voice of GR3 MBROLA database belongs to the fifty years old Dimitrios Delviniotis, who has a great experience in chanting music. We would like to thank Dimitrios Delviniotis (University of Athens) for his contribution in recording the database, Gerasimos Xydas (University of Athens) for his help in building the database and Nicolas d'Allesandro (Faculty Polytechnique de Mons) for his efforts on the mbrolization of the database.

This research was funded by the European Social Fund and Hellenic National Resources under the AOIDOS project of the Programme PYTHAGORAS II/ EPEAEK II.

9. REFERENCES

[1] Chowning J. "Computer Synthesis of Singing Voice", *In ICMC '81 Proceedings*, La Trobe University, Melbourne, 1981.

[2] Rodet X. et Al. "The Chant project: From Synthesis of the singing voice to synthesis in federal". *Computer Music Journal 8 (3) (pp. 15-31)*, MIT Press, 1984.

[3] Soundberg John. "Synthesis of singing by rule", *In Current directions of computer music research,* MIT Press, 1989[30].

[4] Cook P. "Spasm, a real-time Vocal Tract Physical Model Controller and Singer; the companion Software Synthesis System", *Computer Music journal, 17(1),* MIT, Boston, 1993.

[5] T. Dutoit and H. Leich, "MBR-PSOLA: Text-to-Speech Synthesis Based on MBE Resynthesis of the Segments Database," *Speech Communications, no 13, pp. 435-440,* 1993.

[6] Berndtsson, G. "Systems for synthesizing singing and for enhancing the acoustics of music rooms", *Dissertation, KTH, Department of Speech communication and Music Acoustics,* Royal Institute of Technology, Stockholm, 1995.

[7] Cook, P. R. (1996) "Singing Voice Synthesis History, Current Work, and Future Directions," *Computer Music Journal*, 20:2.

[8] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, O. van der Vrecken, "THE MBROLA PROJECT: TOWARDS A SET OF HIGH QUALITY SPEECH SYNTHESIZERS FREE OF USE FOR NON COMMERCIAL PURPOSES", *proc ICSLP'96,vol.3, p1393-1396.*

[9] Macon M. W., L. Jensen Link, J. Oliverio, M. Clements and E. B. George, "Concatenation-based MIDI-to-singing voice synthesis," *103rd Meeting of the Audio Engineering Society*, New York, 1997.

[10] Lomax K. "The Analysis and Synthesis of the Singing Voice", *PhD Thesis*, Oxford, 1997.

[11] Macon M. W., L. Jensen Link, J. Oliverio, M. Clements and E. B. George, "A system for singing voice synthesis based on sinusoidal modelling", *Proc. of International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 435-438*, 1997.

[12] Rodet X., Lefevre A. "The Diphone program: New features, new synthesis methods and experience of musical use", *proc. Int. Comp. Music Conference*, Thessaloniki, 1997.

[13] Gibson, I.S., Howard, D.M., Tyrell, A.M. "Real-time singing synthesis using a parallel processing system", *Proceedings of the IEE colloquium on Audio and music technology; the creative challenge of DSP, IEEDigest 98/470, 8/1-8/6*, 1998.

[14] Meron Y. "High Quality Singing Synthesis using the Selection-based Synthesis Scheme", *PhD thesis*, University of Tokyo, 1999.

[15] X. Rodet, "Synthesis and Processing of the Singing Voice," *Proceeding of the first IEEE Benelux Work-shop on Model-Based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgium, 2002.

[16] Marcus Uneson, "Outlines of Burcas – a Simple Concatenation-based MIDI-to-Singing Voice Synthesis System," *TMH-QPSR, Vol. 43 – Fonetik*, 2002.

[17] Lu H. –L. "Toward a High-Quality Singing Synthesizer with Vocal Texture Control", *PhD thesis,* Standford University, USA, 2002.

[18] Yamaha Corporation Advanced System Development Center. *New Yamaha VOCALOID Singing Synthesis Software Generates Superb Vocals on a PC*, 2003.

[19] Bonada J., Liscos A. "Sample-based singing voice synthesizer by spectral concatenation", *Proceedings of Stockholm Music Acoustics Conference 2003*, Stockholm Sweeden, 2003.

[20] Loic Kessous, "GESTURAL CONTROL OF SINGING VOICE, A MUSICAL INSTRUMENT", *Proc. of Sound and Music computing 2004*, Paris, October 20-22, 2004.

[21] Georgaki A. (2004b) "New trends on the synthesis of the singing voice ", *ICMC'04 Proceedings*, Miami , Florida, 2004.

[22] Georgaki A.(2004a) "Virtual voices on hands". *Prominent applications on the synthesis of the singing voice, Sound and music computing Proceedings of the SMC05, IRCAM*, Paris 2005.

[23] Nicolas D'Allesandro, Raphael Sebbe, Baris Bozkurt, Thiery Dutoit, "MAXMBROLA: A MAX/MSP BROLA-BASED TOOL FOR REAL TIME SYNTHESIS", *Proceedings of the 13th European Signal Processing Conference (EuSiPCo'05)*, 2005.

[24] Kyritsi Varvara, "Production of Greek Singing Voice with the MBROLA synthesizer", *Master thesis*, National and Kapodistrian University, Department of Informatics, Athens, 2006.

[25] Georgaki Anastasia, Zannos Ioannis, Valsamakis Nicolas: "Phonodeon: controlling synthetic voices though MIDI-accordion", *SMC*, 2005.

[26] G. Kouroupetroglou, D. Delviniotis and G. Chryssochoidis: "DAMASKINOS: The Model Tagged Acoustic Corpus of Chant Voices", *Proc. of the Conf. ACOUSTICS 2006*, 18-19 Sept. 2006, Heraclion, Greece.

[27] Psaroudakēs, Stelios (2005) «The Orestēs Papyrus: some thoughts on the dubious musical signs», *in Ellen Hickmann & Ricardo Eichmann (edd.), Studien zur Musikarchäologie IV. (Orient-Archäologie). Pp. 471-92. Rahden: Marie Leidorf GmbH*.

[28] *http//www.geocities.com/SunsetStrip/Balcony/7837/tutorial/miditut.html?20052*

[29] *http//www.borg.com/~jglatt/tutr/midiform.htm*

[30] *http//umsis.miami.edu/~kjacobso/speechsynth/speechsynth.htm*

[31] *http//www.phon.ucl.ac.uk/home/sampa*

[32] *http//tcts.fpms.ac.be/synthesis/mbrola.htm*

[33] *http//www.kvr-vst.com/get/984.html*

[34] *http//www.di.uoa.gr/speech*

[35] *http//en.wikipedia.org/wiki/Speech_synthesis#Concatenative_synthesis*

[36] *http//cgi.di.uoa.gr/~gxydas/mbrola/*