

The UOP Text-to-Speech System for Greek Speech Synthesis

P. Stathopoulou-Zois

Computer Laboratory, Department of Electrical & Computer Engineering
University of Patras, GREECE
pstath@ee.upatras.gr

Abstract

A Text-to-Speech system for synthesising the Greek language has been developed in Computer laboratory of the *University of Patras* (UOP). The system is composed of a core and different interfaces so that is compatible for Windows applications using Microsoft SAPI. The Greek TTS synthesizer uses waveform unit concatenation technology working with a sophisticated speech database and produces good quality speech synthesis of the Greek language. The system performs a text-to-speech conversion of any input text introduced to the computer and written according to ordinary Greek orthography.

The paper reviewing the system, makes emphasis on the one hand to the speech synthesis methodology, which is language dependent, on the other hand to its advantages. A careful linguistic study of the Greek language led to the construction of a speech database composed of a minimum set speech units, with well-defined properties. Finally in the paper are presented the first experimental measures for the evaluation of the system the results of which are very encouraging.

1. Introduction

The last two decades have been developed a number of different speech synthesis methods for various languages. Most efforts have been reported for languages spoken in large population countries and in countries with developed economies (e. g., English, French, German, Italian etc.). Many TTS products and applications for most of the above languages implemented with different design and quality specifications can be found now in the market.

There are many methods for synthesising speech by concatenation. Some use short phonetic segments like phonemes and their allophones, while others use transallophonic units such as diphones [1], [4] and [11], syllables [9], demisyllables [3] and clusters of vowels and consonants extracted from syllables [2], [5], [6] and [7]. Each of these phonetic units synthesis methods has certain advantages and disadvantages related to issues such as: the complexity of the synthesis rules, the size of the vocabulary, the produced speech quality and the amount of the computer memory required to store the synthesis units.

For synthesising the Modern Greek language a flexible method was developed by studying its phonetic characteristics, [14], [15], [16] and [17]. This method uses a combination of *syllables*, *phones* and *allophones* and was first introduced for the Italian language [2], [6] and [10], which has similar characteristics with the Greek. The speech synthesis process is being implemented in real-time, by concatenation of the above phonetic units. Our TTS system has the option to select the *PCM synthesis process* or the *LPC synthesis process*. In the *PCM synthesis process*, the synthetic Greek messages are produced by concatenation of the corresponding

phonetic units extracted from the *PCM* format phonetic database. With the *PCM* synthesis approach we do not have any intervention capability in order to improve the quality of the produced speech.

In the *LPC* synthesis process each phonetic unit is created by synthesis from the *LPC* parameters stored in the database and is concatenated with others to produce the Greek synthetic messages. With the *LPC* of each unit we can intervene in its parameters in order to improve the quality of the produced synthetic speech. We can develop algorithms, which can smooth discontinuities between phonetic units by matching dynamically the adjacent ones [4], and algorithms that control prosodic behaviour of the synthetic speech [18].

2. The UOP TTS architecture

The UOP Text-to-Speech synthesizer is composed of two basic modules: the Natural Language processing (*NLP*) and the Speech Synthesiser (*SS*). The system performs serial processing of data and the data exchange formats are well defined between modules. The processing blocks are sharply separated from each other and each module passes its output to the next module. This modular organisation makes it possible to control the operation of the synthesiser at low levels, make measurements concerning the stage of the TTS conversion and use the synthesiser in a developing environment for research purposes.

The *NLP* module is designed to produce automatically the phonetic transcription of any written text in Greek language and is responsible to realise text normalisation, phonetic transcription and prosody generation.

The text normalisation process performs a text pre-processing that is only a text-to-text replacement. It performs grapheme-to-grapheme conversion, sentence separation, text normalisation, and number-to-letter conversion and prosody prediction. It expands into full orthographic form acronyms, abbreviations, proper names, addresses, symbols, Arabic and Romans numerals, time expressions and date. This process is also responsible to introduce marks and labels which are very convenient for the prosody generation.

The phonetic transcription process performs the conversion from written to phonetic form. The full orthographic text is first segmented in syllables, and then stress is assigned to syllables and finally converted to phonetic units. The marks of syllabication and accentuation are very important for the correct prosody assignment. A small number of rules is sufficient to convert any text written in Modern Greek language without appear ambiguities. For each orthographic form there is a unique phonetic transcription but in some cases, it cannot be resolved with these rules and is needed a dictionary to define the transcription.

The prosody generation process is the most important factor that permits to obtain a natural sounding quality of the synthetic speech. From the two previous processes, taking

information in relation with the prosody marks and labels we can calculate the prosodic structure of the text in level of a sentence or a tonic-group. Sentences patterns describe the evolution of the prosodic parameters (pitch, duration, and amplitude) along the time axis.

The SS module generates the acoustical signal using *PCM synthesis process* or the *LPC synthesis process*. The *LPC* synthesis is used to adapt the characteristic prosody of the stored units to the values assigned by the prosodic model.

3. Speech Database Generation

From studies has been shown that phonetic behaviour of all elementary speech units is language dependent if these units are going to be used in concatenation speech synthesis. In order to clarify the speech units' behaviour, a phonetic study of the particular language is required. This study consists to describe the behaviour of the speech units in relation to their phonetic environment, develop the appropriate rules, which control their behaviour in natural speech and find how these rules must be included in the speech synthesis process.

The speech synthesis process must be consisted of two basic components:

- 1) the memory storage for the *Phonetic Units' Database (PUD)* which consists of units extracted from natural speech, and
- 2) The program which includes the above mentioned rules and applies them to the concatenation process, smoothing the parameters of units to create time trajectories where it is appropriate.

It is well known that if the required amount of memory is small because the total number of phonetic units is limited, then the complexity of the concatenation rules increases importantly. Furthermore it is known that whereas is difficult to quantify program or calculation complexity for different concatenation synthesis approaches, the amount of the needed memory for the speech units is simpler to determine. Therefore the aim of concatenation speech synthesis process is to be achieved reduction of the concatenation rules complexity lessening concurrently the required *PUD* size.

Our work concentrated on the selection of a well-defined number of elementary phonetic units able to fulfil the above specifications. We observed, from the grammatical rules of the Greek language [18], that the syllabication process offered a good framework for Greek speech synthesis method. Similar observations have been made for various other languages (e.g., English, Italian) [6] and [11]. Syllables can be used as elementary phonetic units and have the advantages that need less modelling precision of the co-articulation at the syllabic ends than at the nuclei [11]. Therefore the selection of syllables as phonetic units requires less smoothing between them during the concatenation process.

The number of *syllables* for the Greek language is about 2092 [13] almost the half of the one for the English language, which is about 4400 [11]. Further observations on the structure of the Greek syllables, which have been confirmed by experiments, permitted us to split most of them to simpler ones. Therefore the number of 2092 Greek syllables was corresponded to 113 or 210 phonetic units which, when combined appropriately by using the concatenation rules, can reproduce any of them. These phonetic units or sub-syllables that are first determined at linguistic level, constitute a set of phonetic units and are distinguished to four different types: the *C-type*, the *CV-type*, the *V-type* and the *CVV-type*. Furthermore the acoustical realisation has well-defined properties. These properties are related to the duration, pitch

and amplitude of the speech signals representing the phonetic units at acoustical level.

3.1. Definition of the Greek Speech Synthesis Units (GSSU)

The theoretical determination of the proposed phonetic units has been a guide for the construction experimentally of our *PUD*. These elementary phonetic units called "Greek Speech Synthesis Units (GSSU)," [17] "*must be phonetic elements, extracted from natural speech, that permit the reconstruction of any Greek language message with a given degree of intelligibility.*" Therefore, it is clear, that we must determine a certain number of phonetic units, which have such properties that, when they are joined together permit the reconstruction of any Greek message.

Since humans are able to recognise the phoneme of any allophone, the speech reconstruction from a large number of different allophones is not necessary. The most convenient solution to the problem of speech reconstruction is the use of a limited number of elementary phonetic units that offer no ambiguities among the phonemes, yielding at the same time intelligible synthetic messages.

The phonetic function of the Greek language and its linguistic properties are very simple, and each letter has its own phonetic sound. In the Greek language are not being observed significant phenomena of articulation reduction (shrinkage). The Greek language is comprised of 38 phonemes and allophones [18]. There are neither silent characters, nor long or short vowel sounds while the accented syllables are marked in each word. Its main characteristic is that it has only five vowels /i/, /e/, /a/, /o/, /u/ and does not present vocalic allophones as other languages do (e.g., English, French etc.).

Also, because of the considerable stability of the spectral characteristics of vowels, which are largely independent of the location of sounds in the word and the position of stress, Greek vowels are not influenced by the presence of following consonants or vowels [17]. Yet, there are a very small number of consonants that are being influenced by the presence of a following consonant or vowel. Thus we have reached to the conclusion that every word can be cut after each included vowel that corresponds to the syllabication process of all the Greek words

In the syllabication process of the Greek the following rules are applied [13] and [18]:

- ◇ any consonant between two vowels is syllabicated with the next vowel,
- ◇ two consonants between two vowels are syllabicated with the next vowel,
- ◇ three consonants between two vowels are syllabicated with the next vowel,
- ◇ two-digit consonants are not separated,
- ◇ two-digit vowels, diphthongs and the clusters of "ev" and "av" behave as vowels,
- ◇ compound words follow the same syllabication rules.

With the aid of the above syllabication rules which are valid for Greek written language we prepared a list of all the possible syllables appearing in written Greek. The resulted syllables were all the combinations of the consonant clusters having most of them one of the five vowels at the end. This list included a very large number of syllables that are grouped to vowels (V) and combinations of (VV), (CV), (CVV), (CVC), (CCV), (CCCV) etc.

By testing all the allophones of each Greek phoneme, we identified that if one replaces most of these, the replacement does not influence the intelligibility of the produced synthetic speech. It was sufficient to keep just one allophone for all that correspond to each phoneme. Consequently we achieved

significant reduction of the number of syllables adopting the following rules:

- ◊ We do not need combinations as *VC*, *VV* because the Greek vowels are little influenced by the presence of a following consonant or vowel.
- ◊ We do not need combinations as *CCV*, *CCCV*, *CVC* because the Greek consonants are little influenced by the presence of a following consonant or vowel.
- ◊ When we have the combinations *VC*, *VV*, *CCV*, *CCCV*, *CVC* we can split them after each vowel or each consonant because the existing transitions are not important.
- ◊ We need combinations of the forms *C*, *V*, *CV* only because we confirmed that in most cases for each phoneme exists one allophone.
- ◊ Pitch, duration and amplitude must be kept constant because the redundancy of Greek speech is so high that the lack of stress does not affect seriously the intelligibility.
- ◊ Intonation, stress and tempo are not needed because these prosodic (or supra-segmental) features of Greek speech are influencing only the speech quality.

Taking in account the above rules we have fixed the number of the phonetic units by using one allophone for each and splitting some of the complex syllables into simpler ones. Therefore the final set of phonetic units was composed only of simple consonants (*C*), vowels (*V*), and complex syllables (*CV*) and (*CVV*) only. This set contains the appropriate kind and number of phonetic units (sub-syllables) which guarantees at least the reconstruction of any Greek synthetic message. This set is used for the speech database construction of the UOP TTS synthesizer.

3.2. The Greek Speech Synthesis Units (GSSU) properties.

It is evident that creating new words by concatenating acoustical signals segments, which have been extracted from natural speech, are generated various problems that affect the intelligibility and quality of synthetic speech. Joining together speech segments in a new sequence, which inevitably differs from their original one, are created discontinuities and temporal development of short-time spectral envelopes of the synthetic speech signal.

These phenomena do not appear with the use of our phonetic units because they include any hazardous transitions and steady-state zones in their inner content while at their start and end have the minimum discontinuities. In the majority of Greek words the transition appears at the nuclei of every syllable. Our phonetic units are being extracted from their lexical environment in order to conserve all the acoustic features of the language including the significant transition that appears in the majority of them in their nuclei (syllables *CV*, *CVV*). These differ from those proposed until now in similar systems, such as demissyllables and diphone elements [1], [4] and [11], which have the steady-state zones at the start and the end of the units. This selection permits to minimise successfully part of discontinuities that may arise when the phonetic units are joined to generate the synthesised messages. By minimising the discontinuities we avoid to use complex and time-consuming interpolation operations between units during the concatenation process.

Because the phonetic units are extracted from natural speech, pronounced by a particular speaker, they contain characteristics, which depend on the speaker's voice. These characteristics are pitch, length, and timbre. We have ascertained that these characteristics, in Greek language, give to the speech more naturalness than intelligibility while the

emotional content or the speaker identity does not influence the degree of intelligibility. Therefore, it seems reasonable to accept a certain loss of naturalness by adopting certain common characteristics for the phonetic units in order to standardise them. For this standardisation, to be applied on the phonetic units, we must have fixed values for pitch, amplitude and duration.

We can achieve fixed pitch during the recordings of the phonetic material. This can be done by instructing the speaker to maintain constant fundamental frequency when pronounces the material to be recorded. To obtain fixed values for the amplitude and duration of phonetic units the process is more complicated and need extensive experimental work.

For each type of phonetic unit we must choose the correct duration in order to have the essential information included and to minimise prospective concatenation problems. The average duration of the basic phonetic units (phone, diphones, syllables, demissyllables etc.) was estimate some years ago [11] and [17], and in general is almost the same for all languages. The definition of duration for our phonetic units started as a theoretical assumption and was confirmed by experiments. Finally our phonetic units were chosen to have almost standard pitch and two different lengths of duration, representing the following categories of sounds:

- ◊ a standard acoustic signal segment with 128 ms duration corresponding to each consonant (*C*),
- ◊ a standard acoustic signal segment with 256 ms duration corresponding to an each vowel (*V*), or to a cluster (*CV*) and (*CVV*).

The selected duration lengths are longer than average, give the impression of slow pronunciation but increase the intelligibility. Finally for all types of phonetic units we must adjust the amplitude to the same maximum level and apply to them particular pattern shapes. For each phonetic unit amplitude and shape are changed manually to the desired level.

4. Extraction and Preparation of Phonetic Units' Database

In order to create *PUD* an extensive experimental work was done. The appropriate phonetic material was prepared and recorded. A speaker was used and a phonetic material was recorded. The recorded phonetic material was composed of Greek words, which were three or four syllables long and were carefully selected in order to contain at least twice our candidate phonetic units. We selected to have at least two source words in order to avoid inevitable mispronunciations and get the best possible quality. In special cases of phonetic units it was essential to have more than two or sometimes three source words from which the best representation of a particular phonetic unit was obtained. These words were spelled from a native Greek speaker with good pronunciation and recorded on magnetic tape. During the recording of the phonetic material the speaker took care to maintain nearly its pitch almost constant and to avoid stress phenomena. Afterwards the tape-recorded phonetic material was low-pass filtered at 4.3 kHz, digitised in 12 bits at a sampling frequency of 10 kHz and stored in digital form on the computer's hard disk.

We imposed strict recording specifications for the phonetic material because our basic aim was to acquire the most suitable one for future experimentation. In order to minimise possible distortions special care was taken for the actual recording of the phonetic material. From these source words, we extracted the phonetic units, avoiding carefully

selecting the accented syllables as far as it was possible because they contain strong prosodic phenomena.

From the recorded material a speech signal corresponding to each phonetic unit was isolated, extracted and standardised [17]. Finally all the resulted speech units were inserted to the database following a particular structure which eases their access from the speech synthesis module.

The first version (1989) of our *PUD* was comprised of 113 speech units, which are following the above specifications. The design and implementation of the first version of the *PUD* gave us very promising results. Using speech units corresponding to these 113 phonetic units we can reproduce any oral message in Greek. A flexible *TTS* was realised composed from the above *PUD* and a collection of *DOS* programs (written in Pascal language) on personal computer (IBM PC compatible) with a TMS32020 DSP board for the input/output of speech signals. This implementation presented considerable advantages as real-time speech synthesis, intelligible and good quality synthetic speech, minimum storage requirements, least possible discontinuities in the speech synthesis process, and a good integration with a user friendly and flexible *NLP* system for the Greek language.

The *PUD* is organised in Pulse Code Modulation (*PCM*) and Linear Prediction Coding (*LPC*) formats. In the *PCM* format database the speech units are coded with frequency sampling 10 kHz and 12 bits quantization and the total storage size is about 532 Kbytes. For the *LPC* format database the units were analysed with the autocorrelation method of linear prediction. Every speech unit consists of a number of frames (20, 40) each including 12 reflection coefficients, a gain factor and pitch parameter. In the *LPC* format database are stored each units' parameters for all frames and the total storage size is about 109 Kbytes [16].

The first version database was not included a number of units required in some synthesis cases with the consequence that ambiguities might result in the produced synthetic speech. These phonetic units are the diphthongs (descending or ascending), the allophones of Greek phonemes /g/, /b/, /d/, and the units corresponding in the cluster syllables /CV₁V₂/ (C is one of the Greek consonants, V₁ is the vowel /i/ and V₂ can be one of five Greek vowel).

It was shown by the listening tests that the absence of these units does not impair strongly the intelligibility of the produced synthetic speech but the quality of its. The most recent version of our *PUD* including the above phonetic units is composed of 210 speech units. The latest version was constructed adopting new specifications for the recording of the phonetic corpus, the phonetic units' extraction, standardisation, and the parametric representation.

To process the phonetic material we designed and implemented the appropriate equipment that the major part is speech signal software. The appropriate software was implemented in Pascal and in an object-oriented environment (*Delphi*), which has extraordinary flexibility, enabling the user to work simultaneously on multiple levels. It's about an interactive graphical shell that runs on Windows 9x/NT compatible computers. For the input-output of the acoustical signals is used the sound card, SoundBlaster, installed on a personal computer with a Pentium MMX/200 MHz processor running Windows 95/98. The entire system consists in two basic modules, the Speech Utilities module (*SPEEUT*) and the Speech Synthesiser module (*SPEECH*). The *SPEEUT* module includes seven distinct procedures. Each of these procedures is dedicated to perform a specific work on acoustical signals. The included procedures are:

- *Convert* to manipulate the different storage formats of the acoustical signals

- *Plot* for the graphical representation of the acoustical signals
- *Replay* for playing the acoustical signals
- *Enhance* to modify the amplitude, length and shape contour of a phonetic unit
- *Analyse* for spectrum analysis and Linear Prediction coding of an acoustical signal
- *Synthesise* for speech synthesis of an acoustical signal from its LPC representation.
- *Base_Handle* for the creation, organisation and improvement of the *PUD*.

The *SPEECH* module includes the following procedures:

- *Input_Text* for the insertion of text in written form to the computer
- *Text_Analyser* for the pre-processing, the morphological and contextual analysis of the written form text
- *Phoneme* for the translation of the grapheme representation text to the corresponding *phonetic one*
- *PCM_Synthesis* for the sequential recall of the phonetic units from the PCM phonetic database and their concatenation in order to produce the corresponding speech message
- *Prosody* for the determination of the acoustical parameters and the generation of the prosody behaviour
- *LPC_Synthesis* for the sequential recall of the phonetic units from the LPC phonetic database, their LPC synthesis and their concatenation in order to produce the corresponding speech message.

Using the *SPEEUT* module and the appropriate procedure each time we processed our phonetic material in order to select the candidate phonetic units for the creation of the *PUD*. The processing consists of the following steps, which are realised sequentially:

- 1) selection of the appropriate word from which we can isolate each phonetic unit,
- 2) location, isolation and extraction of the speech signal corresponding to the each phonetic unit,
- 3) standardisation of the raw speech signal extracted from the natural spoken word and finally,
- 4) Insertion of each speech unit in the speech database after its extraction and standardisation.

With the *Plot* and *Replay* procedures we were recalling successively the words-donors from disk trying to locate the most convenient word from which we could isolate and extract each candidate phonetic unit. The selection of the appropriate word was dictated from the environment in which each candidate phonetic unit was found in every particular word. We accomplished this control by hearing and contemporaneously watching step-by-step on the display the word's acoustical signal. A suitable environment is when the speech signal of a unit is distinguishable from their adjacent and the unit maintains an average duration at least equal or greater of the proposed standard durations (256 ms, 128ms). All the above process is realised by observing the representation and by hearing the speech signals of each word and each candidate phonetic unit. So we accepted or discarded the respective words by examining if the above specifications were valid or not.

Next we proceeded to the standardisation of the phonetic units using the *Enhance* procedure. During this process we improved the form of the raw speech units in accordance with the specifications. Our purpose was to maintain the same duration for all speech units, which for CV-type, CVV-type, V-type is 256 ms and 128 ms for C-type. After we adjusted the amplitude level and the contour amplitude of each speech unit,

so that all speech units have the same maximum level of amplitude and conform to a specific amplitude form.

When all the speech units were standardised and inserted to the speech database two more operations were performed. These operations were necessary for the creation of the speech database's final form. The first operation was the execution of some experiments, for the evaluation of the speech intelligibility in the level of phonetic units, with simple synthesised words and more complicated synthesised messages. This operation was achieved with the use of the *SPEECH* module of the system. The results of the first operation helped us to improve some speech units and create the final form of the speech database with the best-standardised speech units. Using the *Base_Handle* procedure we inserted the improved speech units to the speech database. This was the second operation, which completed the construction of *PCM* speech units' database.

For the construction of the *LPC* speech units' database the coding of each speech unit with the *LPC* analysis was required. With the *Analyse* procedure each speech unit was initially analysed by the autocorrelation method of linear prediction [10] and [17], using a 25.6 ms Hamming analysis window. Every 6.4 ms were computed 12 reflection coefficients, a gain factor, and a voicing parameter. Therefore, each speech unit of 128ms or 256 ms was represented respectively by 20 or 40 parameters' frames of fixed length. Finally with the *Synthesise*, *Plot* and *Base_Handle* procedures we constructed the *LPC* speech units' database, and with the *SPEECH* module we tested the intelligibility of the *LPC*.

5. Measurements of intelligibility assessment.

One of the most difficult problems in the area of speech synthesis and codification is the efficiency assessment of such systems. The problem exists because it is not possible to define formally the quantity of "speech quality" with a mathematical expression. Furthermore, we do not have a precise knowledge how the human ear and brain elaborate the acoustical signals while it is not clear what exactly means "good quality speech".

There are two groups of measurements for the assessment of the quality of speech coders or synthesisers: the *Objective* and the *Subjective* measurements [8], [12].

We have emphasised on Subjective measurements because the aim of this work was to develop a *TTS* system using techniques of codification (*PCM*, *LPC*) already known. Furthermore measurements regarding the distortion of the acoustical signals owed to the codification techniques did not interest us since our aim was the implementation of a *TTS* system for Greek language with standard codification techniques.

Isolated synthetic words			
	1 st listening	2 nd listening	3 rd listening
PCM	56,54%	71,60%	77,56%
LPC	47,19%	61,88%	68,75%
Synthetic sentences			
	1 st listening	2 nd listening	3 rd listening
PCM	68,75%	88,13%	90,00%
LPC	62,59%	70,97%	77,22%

Table 1. Intelligibility results of synthetic speech for Greek language

Were focused on the evaluation of the speech intelligibility of our system for each type *PUD* separately (*PCM* and *LPC*). For the intelligibility assessment we have performed a number of experimental measurements,

constructing 40 isolated synthesised words and a total number of 15 synthesised sentences for each database, which was heard by 70 listeners. Each listener was listening first the isolated synthesised words and after was proceeding with the synthesised sentences. The process for each case, words and sentences was repeated for three successive times for each of them with different announcement order each time. The listener each time should commit to paper the word or the sentence that he or she perceived.

The results of the experimental measurements given in Table 1 for the *PCM* and *LPC* speech units' databases, demonstrated that the intelligibility of the listeners increases by far the second and the third time. This abrupt increase of the intelligibility from the first listening to the third is predictable because our listeners were not trained at all. Also we observe that the intelligibility's degree in the case of the sentences is much higher and this is due to the presence of language redundancy that permits the perception even of the bad words.

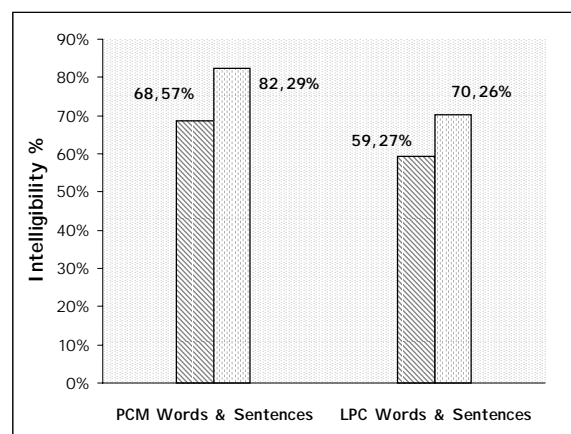


Figure 1. Comparison of the *PCM* and *LPC* intelligibility for synthetic Greek words and sentences

Also the results given in Fig. 1 demonstrated that the intelligibility's degree of the synthetic messages produced from the *LPC* synthesis process is less than of these produced from the *PCM*. The difference between the two methods for the isolated words is about 9% while for the synthesised sentences 12%.

An interpretation of this phenomenon may be that the synthetic messages produced from *PCM* contain more prosody than of these produced from *LPC*. Another interpretation may be that the *LPC* synthesis method suffers in some points. So we obliged to perform some experimental measurements using the Diagnostic Rhyme Test (*DRT*) [12] in order to locate these imperfections.

The *DRT* tests examined the elementary phonemic attributes for the distinction of the 19 Greek consonants. Most of the tests concerned to the distinction between consonants that differ to the voicing, nasality and sustention attributes. The degree of the intelligibility measured from the *DRT* tests reached 78.01%. We measured also that 91% of listeners reached intelligibility between 71% and 100% while the remaining 9% of them 48%. Therefore we conclude that our *TTS* system achieves an intelligibility degree between the typical values 75% and 95%, which is characterised as a good system [12]. During this testing we detected phonetic units that were creating intelligibility problems as well as if these problems were connected to the *LPC* method. In particular we measured that 50% of the listeners confused the 7.14% of the

words, which differed in the voicing attribute (e.g., /t/, /D/). Taking into account the results of our experiments we tried to improve the *LPC* synthesis method as well as some phonetic units.

6. Conclusions

In this paper we present a *TTS* synthesiser for the Greek language making emphasis on the design and the implementation of the phonetic units' database (*PUD*) system. We expose the reasons that guided us to the selection of specific kind phonetic units with well-defined properties and describe how the phonetic units have been constructed and standardised. These phonetic units present several advantages as, least problems during the concatenation procedure ensuring intelligible Greek speech synthesis, low storage size needs and unlimited vocabulary speech synthesis. Furthermore we present the first experimental measurements made for the intelligibility evaluation of the produced synthetic messages.

The experimental results demonstrated that the intelligibility of the produced synthetic speech is very good and justifies the selection of the particular phonetic units. The *LPC* codification of the *PUD* permits further improvements to the intelligibility of the synthetic speech and as well as the introduction of the prosody control in order to obtain more natural synthetic speech. These messages can also be improved towards more natural speech by the introduction of a prosody algorithm. Our future plans for the developed *PUD* are its further improvement, the processing of the phonetic units with different speech analysis models and the evaluation of the produced speech.

References.

- [1] Charpentier, F. J. and Moulines, E. Diphone synthesis using an overlap-add technique for speech waveforms concatenation. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2015-2018, 1986.
- [2] Delmonte, R., Mian, G.A., and Tisato, G. A Text-to-speech system for Italian", *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, San Diego, CA, pp. 2.9.1-2.9.4., 1984.
- [3] Dettweiler, H. and Hess, W. Concatenation rules for demisyllables speech synthesis. *Acoustica, Vol.57*, pp. 268-283; also *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 752-755, 1985.
- [4] Dutoit, T. *High quality Text-to-speech synthesis of French language*. PhD Dissertation, Faculte Polytechnique de Mons, Mons Belgium, 1993.
- [5] El-Iman, Y. A. An unrestricted vocabulary Arabic speech synthesis system. *IEEE Transactions Acoustics, Speech and Signal Processing*, 37: 1829-1845, 1989.
- [6] Francini, G. L., Debiasi, G.B. and Spinabelli, R. D. Study of a system of Minimal Speech-Reproducing Units for Italian Speech. *Journal of the Acoustic Society of America*, 43: 1282-1286, 1968.
- [7] Fujimura, O. and Lovins, J. Syllables as concatenative as phonetics units. in: A. Bell and J. Hooper, (Eds.) *Syllables and Segments*. North-Holland: Amsterdam, pp. 107-120, 1978.
- [8] Kitawaki, N., Honda M. and Itoh, K. Speech-Quality Assessment Methods for Speech-Coding Systems. *IEEE Communication Magazine*, 22: 26-33, 1984.
- [9] Lee, L., Tseng, C. and Ouh-Young M. The Synthesis Rules in a Chinese Text-to-Speech System. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37: 1309-21320, 1989.
- [10] Mian, G. A., Morgantini, F. and Ofelli O. An application of the Linear Prediction technique to efficient coding of speech segments. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, April 12-14, pp. 722, 1976.
- [11] O'Shaughnessy, D., Barbeau, L., Benardi, D. and Archambout, D. Diphone Speech Synthesis. *Speech Communication*, 7: 55-65, 1988.
- [12] Papamichalis, P. *Practical Approaches to Speech coding*, Prentice Hall Inc., Englewood Cliffs, New Jersey, 1987.
- [13] Setatos, M. *Phonology of Modern Greek language*, Athens, 1974.
- [14] Stathopoulou, P. and Kokkinakis, G. Synthesis of Greek speech with LPC-Coded speech segments. *MELECON 83, IEEE, Athens Greece*, May 24-26, p. C1-11, 1983.
- [15] Stathopoulou, P. and Kokkinakis, G., Measurement of the Formants of the Greek vowels. *MECO 82, IASTED*, Tunis, p. 5-1, 1982.
- [16] Stathopoulou, P., Kokkinakis, G. and Mian, G.A. Synthesis of Greek speech with elementary speech segments. *International Conference on Information Sciences and Systems*, University of Patras, Greece, July 9-14, 1: pp. 506-510, 1979.
- [17] Stathopoulou, P., (1989). *Speech Synthesis System in Greek language with Unlimited vocabulary* PhD Dissertation, Electrical & Computer Engineering Dept., University of Patras, Patras Greece, 1989.
- [18] Triadafillidis, M. *Modern Greek Grammar*. Athens, 1980.