



Ain Shams University

Faculty of Engineering

Electronics and communication Engineering Department

Synthesis of Arabic Speech Signals

A Thesis submitted in partial fulfillment
of the requirements of the M.Sc. degree in the Electrical Engineering
(Communications)

621.3828
S.M

By

621.3828

Eng. Sammar Mohamed Elsayed Soliman

Eng. General Organization of Educational Buildings

Supervised by:

Prof. Dr. Salwa H. El-Ramly

Ain Shams University

Dr. Nemat Sayed Abd El-Kader

Cairo University



Cairo-1997

ACKNOWLEDGMENT

The author wishes to acknowledge the sincere effort and continual guidance of her supervisor **Prof. Dr.: Salwa Hussein El-Ramly**. Without her scientific assistance and precious suggestions the thesis would not have that reach.

The author also wishes to express her sincere gratitude to **Dr. : Nemat Sayed Abd El-kader** for helping in deciding the subject of the research, defining the problem, useful suggestions, technical support, and continual follow-up during the research.

The author wishes to thank the **Department of Electronics and Communications Engineering, Faculty of Engineering, Ain Shams University**.

Furthermore, the author wishes to thank the **General Organization of Educational Buildings, Nasr City**.



Statement

This dissertation is submitted to Ain Shams University for the M.Sc. degree in Electrical Engineering (Communications Engineering).

The work included in this thesis was carried out by the author at the Electronics and Communications Engineering Department, Ain Shams University.

No part of this thesis has been submitted for a degree or qualification at any other university or institute.

Date:

Signature

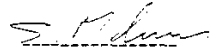
Name: Sammar Mohamed Elsayed Soliman

EXAMINERS COMMITTEE

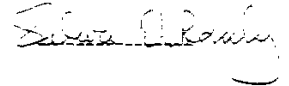
Name, Title & Affiliation:

Signature

1- Prof. Dr. Safwat Mahrous Mahmoud
Electronics & communication Eng. Dept.
Ain Shams University.



2- Prof. Dr. Salwa Hussein El- Ramly
Electronics & communication Eng. Dept.
Ain Shams University.



3- Prof. Dr. Mohamed Yones Abd El-Samee Elhamalawy -----
Computers and system Engineering Dept.
Elazhar University.

Date : 1997

ABSTRACT

Sammar Mohamed Elsayed Soliman. Synthesis of Arabic Speech Signals. Master Degree. Electronics and Communications Engineering Department. Faculty of Engineering. Ain Shams University. 1997.

Using computers in many fields of our daily life requires direct and easy method for dealing with it. Direct communication with computers by text or speech is still the greatest aim of researchers in this field.

During the last twenty years, researchers have developed many techniques to produce computerised speech, requiring relatively small data storage requirements and offering excellent capability for concatenation of phrases.

Electronic speech synthesis aims to cover concisely but completely all fundamental subject matter which is useful either for research or for the design of speaking systems.

Although many text-to-speech systems are commercially available today for a number of languages, (English, French, German, Japanese, etc.), in Arabic language few research works have been done. Recently, there is a growing interest among computer scientists to develop Arabic text-to-speech systems and to synthesize naturally sounding Arabic.

This thesis aims to build a software formant synthesizer that is suitable for the nature of the Arabic language (in spite of its simulation complexity) and using this synthesizer to construct an inventory of the basic Arabic speech units with all phonetic variants of these synthesis units that can be used as a first step of an Arabic text-to-speech system. A comparative study between different basic units is made and then the allophones are chosen as the basic units.

The implemented synthesizer consists of two main steps, extraction of the analysis parameters for each allophone (analysis phase), and the speech reproduction (synthesis phase).

In the analysis phase each allophone is extracted from a set of natural Arabic words containing this allophone in the same place or in various places and then analysed to extract its features (pitch, first four formants, first four bandwidths, voiced/unvoiced classification). A new set of parameters will be stored every 20 ms on the PC hard disk.

Storing these parameters instead of the speech signal itself reduces the required memory storage for the constructed data base.

In the synthesis phase, the files which contain the desired parameters of a new word or utterance, are called and concatenated in a new file that activates the synthesizer to produce the output speech.

Experimental results indicate good output speech quality due to good implementation of the synthesizer and careful choice of the phonetic units. Using the constructed data base with a suitable interpolation method the output speech will be improved.

Contents

| | |
|---|-----|
| ABSTRACT | V |
| LIST OF FIGURES | XII |
| LIST OF TABLES | XIV |
| LIST OF ABBREVIATIONS | XV |
| CHAPTER 1: SPEECH SIGNAL ANALYSIS | 1 |
| 1.1 Introduction | 1 |
| 1.2 Mechanism of speech production | 3 |
| 1.3 Characteristics of speech waveform | 6 |
| 1.3.1 Probability density functions (pdf) | 8 |
| 1.3.2 Nonstationarity of speech signals | 10 |
| 1.3.3 Autocorrelation functions (acf) | 10 |
| 1.3.4 Power spectral density functions (psd) | 12 |
| 1.4 Speech signal analysis | 13 |
| 1.4.1 Time-domain processing for speech signal | 16 |
| 1.4.1.1 Short-time energy and average magnitude | 17 |
| 1.4.1.2 Short-time average zero crossing rate (ZCR) | 18 |
| 1.4.1.3 Voiced-Unvoiced classification | 19 |
| 1.4.1.4 Short-time Autocorrelation function (acf) | 21 |
| 1.4.2 Frequency-domain processing for speech signal | 21 |

| | |
|--|-----------|
| 1.4.2.1 Linear predictive coding (LPC) | 22 |
| 1.5 Fundamental frequency estimation. | 27 |
| 1.5.1 Time-domain pitch detection algorithms (PDAs). | 28 |
| 1.5.2 Spectral domain techniques | 29 |
| 1.5.2.1 The cepstrum technique | 29 |
| 1.5.2.2 Methods using modified autocorrelation | 31 |
| 1.6 Formant estimation | 36 |
| 1.6.1 Formant estimation from cepstrum | 37 |
| 1.6.2 LPC-Based estimation | 37 |
| 1.6.2.1 Peak picking technique | 38 |
| 1.6.2.2 Root finding technique | 39 |
| 1.7 Conclusion | 43 |
| CHAPTER 2: SPEECH SYNTHESIZERS | 44 |
| 2.1 Introduction | 44 |
| 2.2 Speech synthesis techniques | 45 |
| 2.2.1 Synthesis by direct concatenation | 45 |
| 2.2.2 Synthesis through a model | 47 |
| 2.2.2.1 Formant synthesizer | 47 |
| 2.2.2.2 LPC synthesizer | 51 |
| 2.2.2.3 CELP synthesizer | 52 |
| 2.2.3 Vocal tract simulation (Articulatory synthesis) | 54 |
| 2.3 Comparison between synthesis techniques | 56 |

| | |
|---|----|
| 2.4 Conclusion | 59 |
| CHAPTER 3: KLATT SYNTHESIZER | 60 |
| 3.1 Introduction | 60 |
| 3.2 Cascade/parallel formant synthesizer | 61 |
| 3.2.1 Cascade versus parallel | 61 |
| 3.2.2 Waveform sampling rate | 66 |
| 3.2.3 Parameter update rate | 66 |
| 3.3 Digital resonator | 66 |
| 3.4 Digital antiresonator | 69 |
| 3.5 Low-pass filter | 69 |
| 3.6 Synthesizer block diagram | 70 |
| 3.6.1 Sources of sound | 70 |
| 3.6.1.1 Voicing source | 71 |
| 3.6.1.2 Normal voicing | 73 |
| 3.6.1.3 Quasi-sinusoidal voicing | 73 |
| 3.6.1.4 Frication source | 75 |
| 3.6.1.5 Aspiration source | 76 |
| 3.6.2 Vocal tract transfer function | 76 |
| 3.6.2.1 Cascade vocal tract model | 77 |
| 3.6.2.2 Parallel vocal tract model | 80 |
| 3.6.3 Radiation characteristic | 81 |
| 3.7 Synthesis strategy | 81 |
| 3.8 Conclusion | 84 |
| CHAPTER 4: IMPLEMENTED SYNTHESIZER | 85 |
| 4.1 Introduction | 85 |

| | |
|---|-----|
| 4.2 Analysis stage | 85 |
| 4.3 Synthesis stage | 86 |
| 4.3.1 Voicing source | 90 |
| 4.3.2 Noise source | 91 |
| 4.3.3 Cascade synthesizer | 92 |
| 4.3.4 Parallel /cascade synthesizer | 93 |
| 4.3.5 Output stage | 94 |
| 4.4 Evaluation of the synthesizer quality | 94 |
| 4.4.1 Objective measurements | 94 |
| 4.4.1.1 Signal-to-noise ratio, SNR | 95 |
| 4.4.1.2 Segmental signal-to-noise ratio, SNRSEG | 95 |
| 4.4.1.3 SNRMAX, SNRMIN | 96 |
| 4.4.1.4 Signal-to-uncorrelated noise ratio, SNRu | 97 |
| 4.4.2 Subjective measurement | 98 |
| 4.4.2.1 Quality measurements | 99 |
| 4.4.2.2 Intelligibility tests | 101 |
| 4.5 Experimental results | 103 |
| 4.6 Comparison between the implemented synthesizer and the Klatt synthesizer | 108 |
| 4.7 Conclusion | 109 |
| CHAPTER 5: CONSTRUCTION OF AN ARABIC DATA BASE | 110 |
| 5.1 Introduction | 110 |

| | |
|---|------------|
| 5.2 Arabic Allophones | 110 |
| 5.3 Choice of phonetic units | 111 |
| 5.3.1 Phonemes | 113 |
| 5.3.2 Allophones | 114 |
| 5.3.3 Diphones | 116 |
| 5.3.4 Syllables | 117 |
| 5.3.5 Demisyllables | 118 |
| 5.4 Why using allophones | 120 |
| 5.5 Construction of the Arabic speech data | |
| base | 121 |
| 5.6 Conclusion | 131 |
| CHAPTER 6:CONCLUSION AND FUTURE | |
| WORK | 133 |
| REFERENCES | 136 |

List of Figures

| Figure | page |
|--|------|
| 1.1 Speech production model | 5 |
| 1.2 Amplitude waveforms of speech | 7 |
| 1.3 Typical long-time averaged probability density functions for speech amplitudes | 8 |
| 1.4 The dependence of short-term speech variance | 11 |
| 1.5 Long-time averaged auto-correlation function | 14 |
| 1.6 Short-time power spectral density for voiced and unvoiced speech segments | 15 |
| 1.7 Block diagram of the SIFT algorithm | 33 |
| 2.1 Speech synthesis model | 45 |
| 2.2 Formant synthesizer | 50 |
| 2.3 Parallel cascade formant synthesizer | 51 |
| 2.4 Block diagram of LPC synthesizer | 53 |
| 3.1 Relation of the software synthesizer to the hardware | 62 |
| 3.2 The transfer function of the vocal tract simulated by cascade or parallel digital resonators | 64 |
| 3.3 The different configurations of the formant synthesizer | 65 |
| 3.4 Digital resonator | 68 |
| 3.5 Block diagram of the cascade parallel formant synthesizer | 72 |
| 3.6 Voicing source amplitude spectra of normal and quasi sinusoidal voicing | 74 |
| 4.1 Analysis stage flow chart | 78 |

| | |
|---|-----|
| 4.2 Flow chart of the cascade synthesis stage | 88 |
| 4.3 Flow chart of the parallel/cascade synthesis stage | 89 |
| 4.4 Speech intelligibility versus articulation index | 102 |
| 4.5 Waveforms of the natural and the synthetic word “كب” | 106 |
| 4.6 Waveforms of the natural and the synthetic sentence “بسم الله الرحمن الرحيم” | 107 |
| 5.1 Waveform of allphone (ك) | 124 |
| 5.2 Waveforms of the natural and the synthetic word “كب” | |
| 5.3 Pitch contours of some natural and the synthetic Arabic utterances | 129 |
| | 130 |

List of Tables

| Table | page |
|---|------|
| 2.1 Principle features of speech synthesis techniques | 58 |
| 3.1 List of control parameters for the software formant synthesizer | 83 |
| 4.1 List of control parameters for the implemented software formant synthesizer | 91 |
| 4.2 Five-point adjectival scales for quality and impairment, and associated number scores | 100 |
| 4.3 MOS for the the utterance "أنا رجُلٌ رَجِيحٌ" | 104 |
| 4.4 MOS for the the utterance "أنا رجُلٌ رَجِيحٌ رَجِيحٌ" | 104 |
| 4.5 MOS for the the utterance "أنا رجُلٌ" | 104 |
| 4.6 MOS for the the utterance "أنا رجُلٌ" | 105 |
| 4.7 Parameter values of the selected English vowels | 108 |
| 5.1 List of the Arabic consonant phonemes | 112 |
| 5.2 Comparison between the different speech synthesis units | 119 |
| 5.3 Control parameters of allophone ٤ | 125 |
| 5.4 Phonetic symbols of the implemented database | 127 |