# Optimization of Arabic Database and an Implementation for Arabic Speech Synthesis System using HMM: HTS_ARAB_TALK

Krichi Mohamed khalil
Signal Processing Laboratory, Science
Science Faculty of Tunis
SFT 1060, Tunisia

Cherif Adnan
Signal Processing Laboratory, Science
Science Faculty of Tunis
SFT 1060, Tunisia

## ABSTRACT
This paper presents an optimization of the Arabic database and a prototype for real-time speech synthesis. Statistical parametric speech synthesis is a relatively new approach to speech synthesis. Hidden Markov model based speech synthesis, the techniques in this approach, has been demonstrated to be very effective in synthesizing high quality, natural and expressive speech. This work modified the publicly available HTS to establish a complete architecture system, called HTS_ARAB_TALK, which provides us with a basis for further research for a future fully real-time speech synthesis system and we give an overview of the Arabic speech synthesis system using HMM. A brief description of the HTS_ARAB_TALK is presented with some emphasis on the feature that is relevant to the Arabic language. Finally, a mean opinion score for the synthesized speech is presented. These results were supported by subjective evaluation.

## General Terms
Signal processing, analysis and synthesis speech.

## Keywords
HMM, Speech Synthesis, Text to Speech, Arabic Language, Statistical Parametric Speech Synthesis, Hidden Markov Model

## 1. INTRODUCTION
Speech synthesis is defined as the process of generating speech signal by machine. This target can be accomplished using many ways. The traditional way is waveform concatenation such us PSOLA [1]. This technique referenced above has proved to be of a high quality synthesis, typically more natural sounding speech, now the RealSpeak from Nuance and AT&T Labs Text-to-Speech (TTS) are famous concatenative commercial speech technology systems for TTS [2]. But concatenative systems have the drawbacks of limited number of voices and large size memory used to store speech waveforms. Another way for speech synthesis is through software using linguistic rules and features based on analyzing human speech. So this method sometimes called rule-based synthesis, formant speech synthesis or parametric synthesis since it generates small compact parameters from human speech and uses them to synthesize speech signal. DEC talk is still the best commercial formant synthesizer [2]. The advantage of formant synthesizers consist in using small memory since the size of the extracted parameters is less than the size of the speech signal in waveform and the easy customization of synthesized voices. But they have the disadvantage in the generated sound that it is more mechanical sounding so it results less quality than concatenative ones. Statistical parametric speech synthesis system based on hidden Markov models (HMMs). HMM is a new approach that has grown in the last few years which has been proved as a powerful tool in speech recognition since the models produced from the training process contain statistical data that models; the input speech signal and these models have small size. Arabic HMM-based Speech Synthesis is the state-of-the-art high quality natural TTS systems. *HTS_ARAB_TALK* is one of these systems, which is developed specially for Arabic language. This paper describes the overall architecture, several components of the system, and linguistic concepts for Arabic language. This paper is structured as follows. Section 2 describes the HMM-based Speech Synthesis. The optimization of Arabic database with extra information prosody is presented in section 3. The preparation of the environment (HMM training) is presented briefly in section 4. Section 5, describes the development of HTS_ARAB_TALK. Section 6, describes the evaluation result and tests. Finally, section 7 summaries our conclusions and an expected future work.

## 2. HMM-BASED SPEECH SYNTHESIS
As computer technology evolves and the power and resources of computer increases, the task of building more and more natural and intelligible synthetic voices progresses rapidly. In fact; the need for smaller synthesizer footprints and more flexible control of synthesis has brought researchers to envision other ways of using the knowledge contained in the database. The real speech samples at synthesis runtime are not used by the so-called statistical parametric speech synthesis. The pre-recorded database instead is analyzed and various production features are extracted, e.g. spectral envelopes, fundamental frequency, duration of the phonemes, as well as first and second time derivatives of these features [3]. Then the extracted features are used to train a statistical model. In HTS, for instance, each phoneme is typically modeled by 5-states HMM, and the multidimensional Gaussian pdf associated to each state is made by the concurrent training of a context-dependent decision tree. At synthesis time, the HMM models of each phoneme in the target sentence are concatenated, and synthetic speech spectral, pitch, and duration trajectories are generated from HMMs themselves, based on a maximum likelihood criterion [4]. Finally the statistical model is used for generating trajectories, regarding

a given target. These trajectories are used to control a voice production model in order to synthesize the speech [5]. As a result, the overall footprint of the system is fairly small (around 1MB per voice), as just the trained statistical model is needed to run the system. This makes HMM speech synthesis a perfect candidate for portable applications such as those targeted in digital lutherie. On the other side, the real waveforms substituted by the use of production model introduce a small loss in segmental quality, but definitely not as much as in the early rule-based synthesis systems.

## 2.1 Training Part

Figure 1 is a block diagram of Training stage of HMM speech synthesis. The basic architecture of an HTS system consists of two parts, the training and the synthesis, that will be discussed.

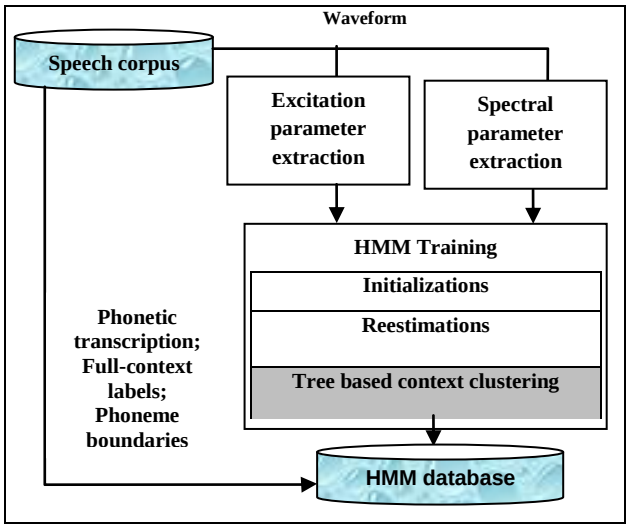


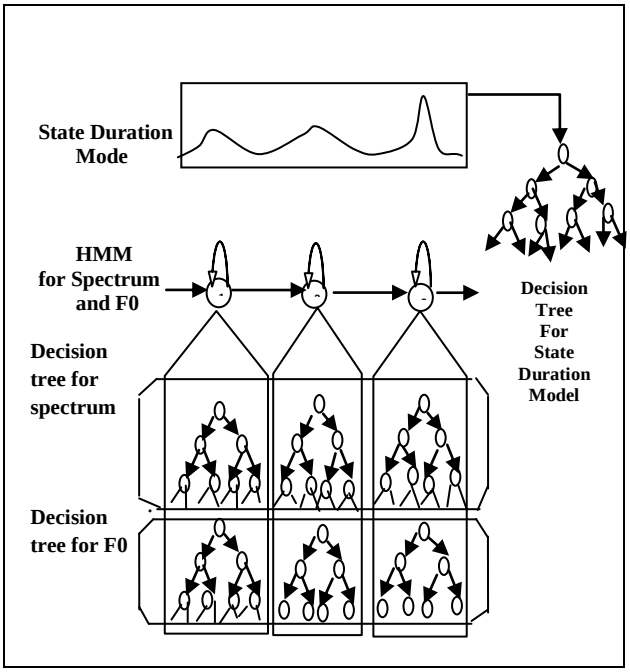

**Fig.1 Training stage of HMM speech synthesis. [5]**

In the training part of HTS, the given database of natural speech where both spectrum (mel-cepstral coefficients and their dynamic features) and excitation (log F0 and its dynamic features) parameters are extracted and then are modeled by a Figure 2. Decision trees for context clustering [7] set of context-dependent HMMs. Notice that not only phonetic, but also linguistic and prosodic context is taken into account. More specifically, the output vector of HMM consists of the mel-cepstral coefficient vector, including the zeroth coefficients, their delta and delta-delta coefficients and of the log fundamental frequency vector, its delta and delta-delta coefficients. In order to model speech in time, HMMs model the state duration densities by using a multivariate Gaussian distribution [8]. As it was described above, in order to handle the contextual factors, such as phone identity factors, stress or accent related factors that affect the targeted synthetic speech output, we use context-dependent HMMs. more the number of these factors soon, the greater variety of prosodies, intonations, emotions and tones of voice become available, as well as different speaker individualities and speaking styles, leading to higher degrees of naturalness. On the other hand, increasing the number of factors increases the number of possible combinations exponentially. Thus, the model parameter estimation is not precise enough when the training data are sparse, i.e. not covering the entire contextual space. In HTS, this problem is solved by applying decision-tree based context clustering techniques [9].As mentioned, magnitude spectrum, fundamental frequency and duration are modeled independently; therefore there is a different phonetic decision tree for each one, as illustrated in Figure 2.

## 2.2 Synthesis Part

In the synthesis part of HTS, the initial input is the target text to be transformed into synthetic speech. This text is parsed

**Fig.2 Decision trees for context clustering, [24].**




and mapped into a context-dependent phonetic label sequence, which is then used to construct an HMM sentence by concatenating the context-dependent HMMs according to this label sequence. When the sentence HMM is constructed, the sequences of spectrum and excitation parameters are generated, [10]. Finally, by using these generated parameters and a synthesis filter module, in this case an MLSA filter [11], a speech waveform is created. Based on the nature of statistical parametric speech synthesis, by modifying the HMM parameters we can obtain different voice qualities of synthesized speech. It has been proved that by using speaker adaptation [13], speaker interpolation [15], or eigenvoice techniques [12] the voice characteristics can be modified.

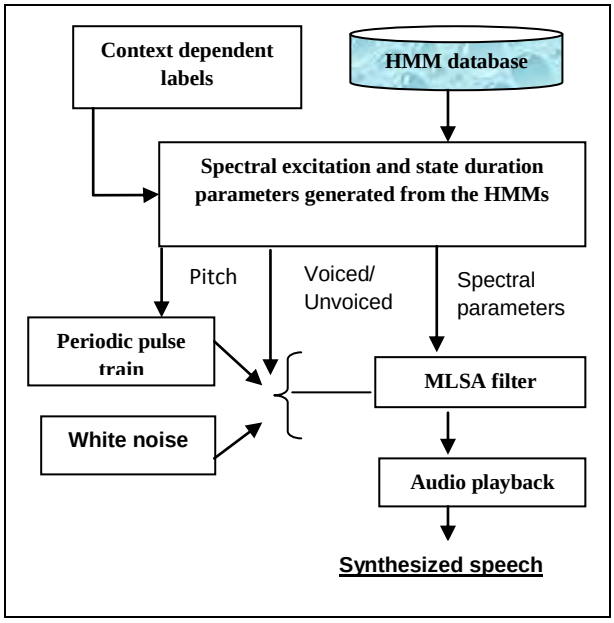



**Fig.3 Synthesis stage of HMM speech synthesis.[5]**

## 3. SPEECH DATABASE CONSTRUCTION

Standard Arabic is the language used by media and the language of Qur'an. Modern Standard Arabic is generally adopted as the common medium of communication through the Arab world today. Standard Arabic has 34 basic phonemes, of which six are vowels, and 28 are consonants [17]. Arabic vowels are affected as well by the adjacent phonemes. The ideal is to have a database sufficiently provided with several examples phonemes [15]. The database used in [16] is without prosodic information. Our target is to improve the database in [16].The audio portion of the database is the only one that interested .wav format is PCM coded 16-bits at a sampling frequency of 16 kHz. To adapt the Arabic phonemes with the HTS system, we use a new presentation of phonemes, since for example the HTS system reject "?". In this work, we added in each labels file a lot of prosody information.

### 3.1 Add prosodic information

The term prosody refers to certain properties of the speech signal which are related to audible changes in pitch, loudness, and syllable length. Prosodic features have specific functions in speech communication (see Fig. 4). The contexts used in the HTS Arabic recipes:

1. phoneme:
- current phoneme
- preceding and succeeding two phonemes
- position of current phoneme within current syllable

2. syllable:
- numbers of phonemes within preceding, current, and succeeding syllables
- stress and accent of preceding, current, and succeeding syllables
- positions of current syllable within current word and phrase
- numbers of preceding and succeeding stressed syllables within current phrase
- numbers of preceding and succeeding accented syllables within current phrase
- number of syllables from previous stressed syllable
- number of syllables to next stressed syllable
- number of syllables from previous accented syllable
- number of syllables to next accented syllable
- vowel identity within current syllable

3. word:
- guess at part of speech of preceding, current, and succeeding words
- numbers of syllables within preceding, current, and succeeding words
- position of current word within current phrase
- numbers of preceding and succeeding content words within current phrase
- number of words from previous content word
- number of words to next content word

4. Phrase:
- numbers of syllables within preceding, current, and succeeding phrases
- position of current phrase in major phrases
- ToBI endtone of current phrase

5. utterance:
- numbers of syllables, words, and phrases in utterance

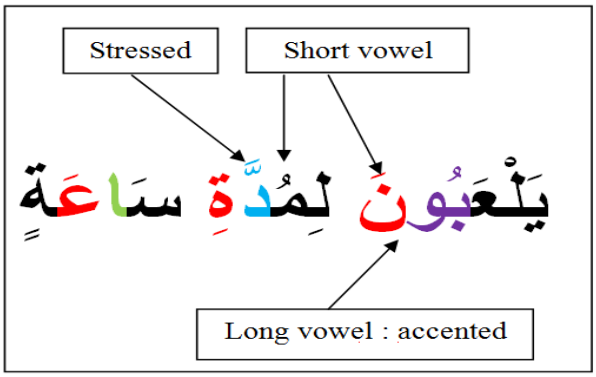



**Fig.4 Example of a sentence / jalAIabuuna limuddati saAItin / ("They play for an hour")**

Technically, the Arabic letters are written from right to left and most of them are linked to one another. Most Arabic words can be reduced to a root which often made up of three letters. Modifying this root by adding prefixes and/or suffixes and changing the vowels results in many word patterns. The target is to obtain a label file for each sentences, in each file, there are once sentences. In the following next figure, we explain the information involved in each utterance.

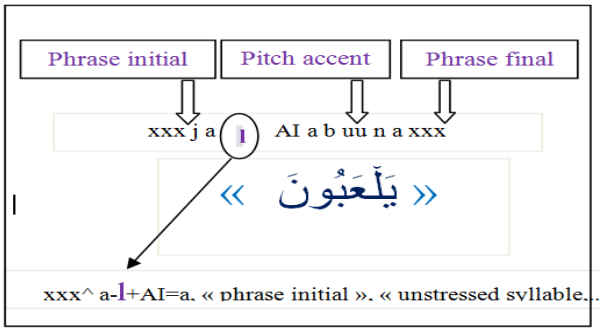



**Fig.5 sentences Arabic with prosodic information**

As far as the Arabic language is concerned, word–stress and its placement are predictable because if we take the structural patterns of the word, then rules can be formulated so as to highlight, the syllable on which stress falls. Word–stress, therefore, is non–phonemic in Arabic [23] stress does not produce a distinction in meaning. Most linguists and orient lists, nevertheless, have distinguished three degrees of non–phonemic stress: primary, secondary and weak. [24] In fact instance, stating a general rule of word–stress placement in Arabic, stipulates that: stress falls on the long syllable nearest to the end of the word. In the absence of a long syllable, the stress falls on the first syllable and on the third syllable from the end in words of three or more syllables[25] extensively discusses accentuation and other phonological phenomena related to syllable structure in classical Arabic. His approach is prosodic. To him, a final syllable of the word is stressed if it is long, i.e. VVC(C) or VCC. He does not consider VV# as a long vowel, e.g.
/darabt/ I hit
/aAImaal/ jobs
If the pre–final syllable is closed, that is of CVC, or CVVC it will be stressed, e.g.
/katabta/ you wrote
/AiTnaani/ these two
But if the pre–final syllable is open, i.e. of the form CV, then either that syllable or the syllable preceding it is stressed, e.g.
/ katabta / You wrote
/ kaataba / he corresponded with
/ qattalat / she murdered
/ katabataa / they (feminine) wrote

## 4. HMM-TRAINING

In this work, we use a first step in HTS that a training part. All of parameters are obtained by a HTS system. After this training, we send all of parameters in HTS-Engine but this step need other file.

### 4.1 Question file

Questions file are text files that define the questions to the nodes of decision trees for HMM clustering. The questions that are asked for phonemes are directly related to background information provided in the label files (labels). The following figure 6 is a question file extract.

```
QS "R-Word_GPOS==0"          {*/F:0_*}
QS "R-Word_GPOS==aux"        {*/F:aux_*}
QS "R-Word_GPOS==cc"         {*/F:cc_*}
QS "R-Word_GPOS==content"    {*/F:content_*
QS "R-Word_GPOS==det"        {*/F:det_*}
```

**Fig.6 Extract a file questions**

## 5. DEVELOPMENT OF HTS_ARAB_TALK

Figure 7 shows the architecture of the current system. It is composed of three major components: a HTS-training, a HTS-engine, an Arabic keyboard. In the HTS-training component, we prepare a prosodic Arabic database and construction of the statistical parametric speech. After training part, we send this parameter to HTS-engine. Text is the input of the system.

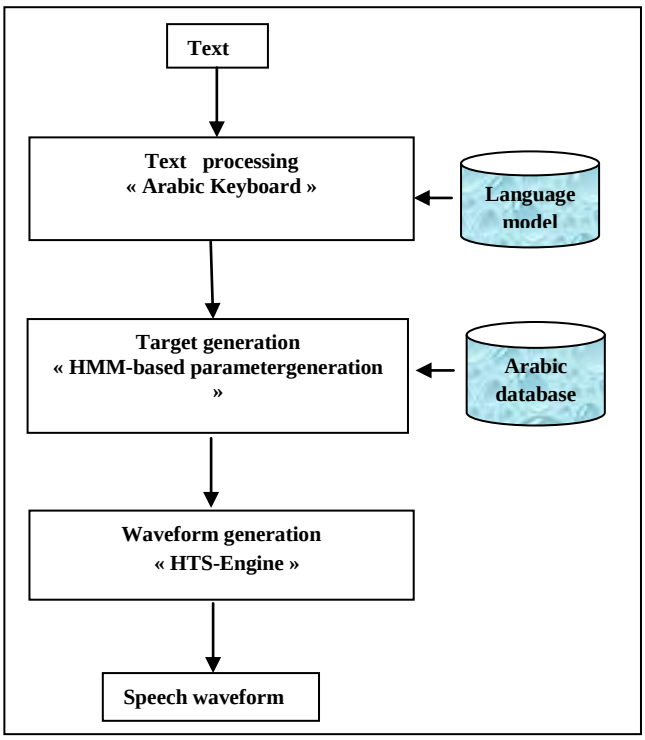



**Fig.7 Block diagram of HTS-ARAB-TALK**

### 5.1 Text segmentation

Syllable Parser will segment the normalized text to syllable unit according to Arabic rules. The architecture is based on Input, Processing and Output Schematic. This module will convert the symbols input into readable text. Input text may be in the form of paragraphs, sentences, or words. Thus, it is necessary to segment the text in a hierarchal order: higher level structures to paragraphs, paragraphs to sentences, sentences to words and words to syllables and syllable to phonemes. In this research, we limited the input text to paragraph form. A paragraph was segmented into sentences by finding the sentence punctuation marks such as '.', '!' and '?'. To segment sentences into words, blank spaces were located in the text that has been classified as a sentence. From the text that has been identified as words, the phonemic representations equivalent to the set of letters of the retrieved word were generated.

### 5.2 Waveform generation

HTS-engine-API: Since version 1.1, a small stand-alone run-time synthesis engine named HTS-engine has been included in the HTS release. It works without the HTK libraries, and it is released under the new and simplified BSD license; Users can develop their own open or proprietary software based on the run-time synthesis engine and redistribute these source, object, and executable codes without any restriction. In fact, a part of HTS-engine has been integrated into several pieces of software, such as ATR XIMERA [20], Festival [21], and Open MARY [22]. The spectrum and prosody prediction modules of ATR XIMERA are based on HTS-engine. Festival includes HTS-engine as one of its waveform synthesis modules. The upcoming version of Open MARY uses the JAVA version of HTS-engine. The stable version, HTS-engine API version 1.0, was released with HTS version 2.1. It is written in C and provides various functions required to setup and drive the synthesis engine. In this step, we used a HTS-Engine (1.07). The following figure 8 represents the general appearance of the HTS_ARAB_TALK.

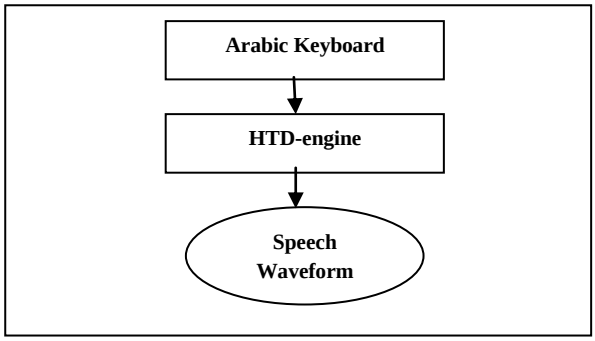



**Fig.8 HTS_ARAB_TALK**

## 6. RESULTS OF SUBJECTIVE TESTING

The main interest concern when choosing the test group is that they should be non-speakers of Arabic language. In order to decide what a good command is, it was decided that the participants need to have the Arabic language as their second language. The group is made up of 36 people. The vast majority of the participants are students at Bourguiba Institute of Languages University Elmanar, Tunisia at the Department of Arabic Linguistics. The level of fluency is varying among the participant, some of them are somehow fluent while the other are not. The main goal of this evaluation test is to determine how much of the spoken output one can understand is. The test is divided into three parts. The first part is to evaluate the system with respect to naturalness. The second part is to evaluate the sound quality. The last part is to evaluate the pronunciation. The participant is asked a few questions about these aspects and is asked to mark how well the voice performs. These simple exercises will asses the overall assessment of this TTS Synthesizer System. Now that the test is done, a summary of the results is presented in this section. The results are presented in diagrams and tables with percentage values.

### 6.1 Naturalness

Concerning the question whether the voice is nice to listen to,**38 %** (14 respondents out of 36) considered the voice natural, **50 %** (18 respondents out of 36) thought that the naturalness of the voice was acceptable and **12 %** (4 respondents out of 36) considered the voice unnatural. Table 1 below shows the outcomes of the questionnaire in detail.

**Table. 1 Naturalness**

| | Very Natural | Natural | OK | Unnatural | Very unnatural | Total |
|---|---|---|---|---|---|---|
| N. of Respondents | 0 | 14 | 18 | 4 | 0 | 36 |
| % of Respondents | 0 | 38 | 50 | 12 | 0 | 100 |

The result are shown in fig.9

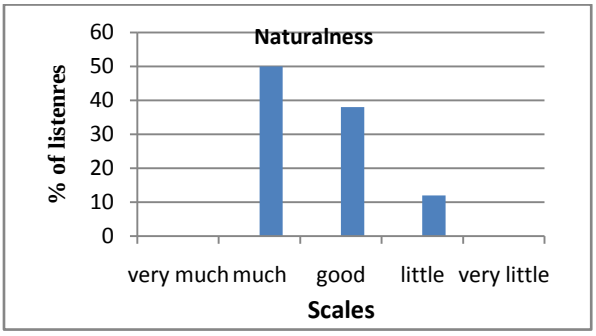



**Fig.9 Naturalness of the voice**

### 6.2 Sound quality

The question for this part is "Do you consider the system to be of good sound quality?" After listening to the sound, **22.22 %** (8 respondents out of 36) considered the voice has a very good sound quality. **41.6 %** (15 respondents out of 36) considered the voice has a good sound quality. **27.7 %** (10 respondents out of 36) thought the sound quality of the voice is neither bad nor good and the remaining **8.33 %** (3 respondents out of 36) considered that the sound quality of the system bad. Table 2 below shows the outcomes of the questionnaire in detail.

**Table. 2 Sound quality**

| | Very Good | Good | OK | Bad | Very Bad | Total |
|---|---|---|---|---|---|---|
| N. of Respondents | 8 | 15 | 10 | 3 | 0 | 36 |
| % of Respondents | 22.22 | 41.6 | 27.7 | 8.33 | 0 | 100 |

The result are shown in fig.10

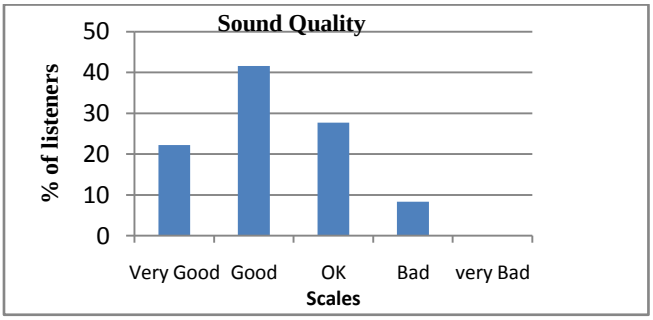



**Fig.10 The sound quality of the voice**

### 6.3 Pronunciation

The pronunciation part consists of two questions addressed to the participants. This is chiefly done to be able to get an idea of how difficult the speech uttered by the system, is to grab/get and to be able to decide what sounds are the most difficult to catch and process in order to be improved them. The first question in this category is if the listeners found it was very hard to grab/get some of the words. I hoped to get some information about what words were considered hard to grab/get and what sounds these words contained. **10.06 %** (6 respondents out of 36) of the listeners thought it is very hard

to grab/get some of the words. **41.6 %** (15 respondents out of 36) of the listeners thought it was easy to grab/get, while **22.8 %** (8 respondents out of 36) thought it is neither hard nor easy, and **25 %** (9 respondents out of 36) thought it is hard to grab/get some of the words. Table 3 below shows the outcomes of the questionnaire in detail.

**Table.3 Pronunciation Question 1**

| | **Very Hard** | **Hard** | **OK** | **Easy** | **Very Easy** | **Total** |
|---|---|---|---|---|---|---|
| N. of Respondents | 6 | 15 | 8 | 9 | 0 | 36 |
| % of Respondents | 10.06 | 41.6 | 22.8 | 25 | 0 | 100 |

The result are shown in fig.11

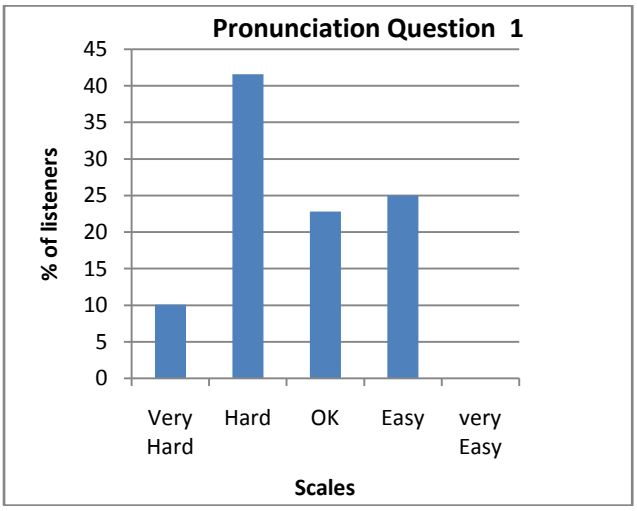



**Fig.11 The sound quality of the voice**

The second question, in the pronunciation part, is intended to investigate if the participants had to concentrate hard to be able to grab/get the speech uttered by the system. This question can give information about how difficult the voice is to grab/get and how much the participants had to concentrate to grab/get the voice. The results are summarized according to the subjects' own estimations. The results after listening to the sound show that **19.4 %** (7 respondents out of 36) of the participants do not have to concentrate on the sound. While, **36.11 %** (13 respondents out of 36) of the participants consider the system requires normal concentration. **27.7 %** (10 respondents out of 36) of the participants have to concentrate a little. For **13.8 %** (5 respondents out of 36) some concentration is needed for specific sounds. The remaining **2.7 %** (1 respondent out of 36) had to concentrate a lot. Table 7.6 below shows the outcomes of the questionnaire in detail.

**Table. 4 Pronunciation Question 2**

| | **A lot of concentration** | **Some of concentration** | **Normal concentration** | **Little concentration** | **No concentration** | **Total** |
|---|---|---|---|---|---|---|
| N. of Respondents | 7 | 13 | 10 | 5 | 1 | 36 |
| % of Respondents | 19.4 | 36.11 | 27.7 | 13.8 | 2.7 | 100 |

The result are shown in fig.12

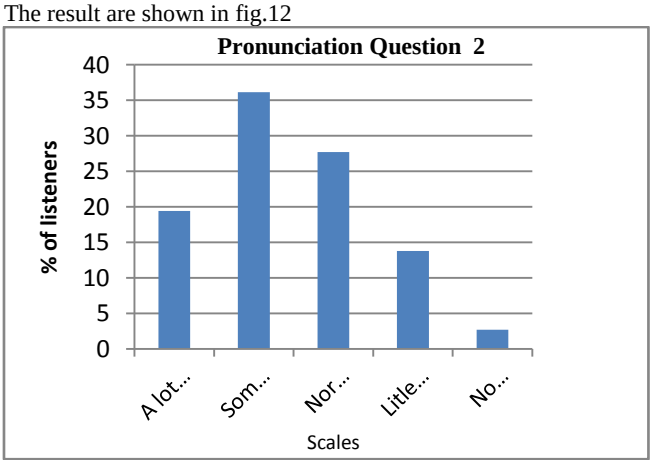



**Fig.12 the concentration needed to hear the pronunciation**

## 7. DISCUSSION AND FUTURE WORKS

Text-To-Speech Synthesizer has been developed gradually over the last few decades and it has been integrated into several new applications. For most applications, the intelligibility and comprehensibility of TTS Synthesizer have reached the acceptable level. Nevertheless, in prosodic, text preprocessing, and pronunciation fields there is still much work and improvements to be done to achieve more natural sounding speech. Natural speech has so many dynamic changes that perfect naturalness may be impossible to achieve. However, since the markets of TTS Synthesizer related applications are increasing gradually, the attention for giving more efforts and funds into this research area is increasing as well. Current TTS Synthesizer Systems are so complicated that one researcher cannot handle the whole system. With good modularity it is likely to divide the system into a number of individual modules whose developing process can be done alone if the communication between the modules is made carefully. Some of the possible improvements that can be made are: Record more sounds in the sound database. More sounds can be recorded to have better performance and more vocabularies. Users can learn more words without much limitation. Build more user friendly interfaces, such as a command to select different voices, for example, voice of a man and voice of a woman. As well as an interface, this will allow users to click on the Arabic words rather than typing them – applicable for users who do not have Arabic keyboard. Adding an animation character (Agent). To attract user to continue using this software, one can include an agent or mount utterance character. Humans are more attracted to animated and attractive interfaces which can create interest and fun in learning. The characters are able to speak the input text, along with the output sound with mouth utterances and gestures.

## 8. REFERENCES

[1] Cheng-Yuan,L.and Jang,J."A two-phase pitch marking method for TD-PSOLA synthesis" ICSLP,2004.

[2] Yorozu, Y. Hirano, M. Oka, K. Tagawa, and Y. "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[3] Fukada, T. Tokuda, K. Kobayashi T. and Imai, S. "An adaptive algorithm for mel-cepstral analysis of speech," Proc. of ICASSP'92, vol.1, pp.137–140, 1992.

[4] http://kt-lab.ics.nitech.ac.jp/˜tokuda/SPTK/.

[5] Tokuda, K. Masuko, T. Miyazaki N. and Kobayashi, T. "Hidden Markov Models Based on Multi-Space Probability Distribution for Pitch Pattern Modeling," Proc. of ICASSP, 1999.

[6] Toth, B. andNemethG. "Optimizing HMM Speech Synthesis for Low-Resource Devices", November 15, 2011.

[7] Tokuda, K. Zen, H. andBlack, A.W. "An HMM-based speech synthesis system applied to English", in IEEE Speech Synthesis Workshop, 2002.

[8] Assaf, M. "A Prototype of an Arabic Diphone Speech Synthesizer in Festival," Master Thesis,Department of Linguistics and Philology, Uppsala University, 2005.

[9] Zen, H. Tokuda, K. and Kitamura, T. "Decision tree based simultaneous clustering of phonetic contexts, dimensions, and state positions for acoustic modeling", in Proc. Eurospeech,2003b, pp. 3189-3192.

[10] Tokuda, K.Yoshimura, T. Masuko, T. Kobayashi, T.Kitamura, T. "Speech parameter generation algorithms for HMM-based speech synthesis", in Proceedings of International Conference on Acoustics, Speech, and Signal Processing,2000, Vol. 3, pp. 1315-1318.

[11] Fukada, T., Tokuda, K., Kobayashi, T., Imai, S., "An adaptive algorithm for mel-cepstral analysis of speech", in Proc. Of ICASSP92, 1992, vol.1, pp.137-140.

[12] Shichiri, K., Sawabe, A., Tokuda, K., Masuko, T.,Kobayashi, T., Kitamura, T., "Eigenvoices for HMM-based speech synthesis", in Proceedings of International Conference on Spoken Language Processing,2002, pp. 1269–1272.

[13] Tamura, M., Masuko, T., Tokuda, K., Kobayashi, T., "Adaptation of pitch and spectrum for HMM-based speech synthesis using mllr", in Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 2001, Vol. 2, pp. 805808.

[14] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T., "Speaker interpolation in HMM-based speech synthesis system", in Proceedings of European Conference on Speech Communication and Technology97, 1997, Vol. 5, pp.2523-2526.

[15] M.Boudraa, B.Boudraa, B. Guerin, ''Elaboration d'une base de données arabe phonétiquement équilibrée'', Actes du colloque Langue Arabe et Technologies Informatiques Avancées, pp 171-187, Casablanca, Décembre 1993.

[16] K. Mohamed Khalil, C.Adnan," Arabic HMM-based Speech Synthesis", in International Conference on Electrical Engineering and Software Applications ICEESA 2013.

[17] M. Assaf, "A Prototype of an Arabic Diphone Speech Synthesizer in Festival," Master Thesis,Department of Linguistics and Philology, Uppsala University, 2005.

[18] Eriwn, W. M. (1963) A Short Reference Grammar of Iraqi Arabic.Washington: Georgetown University Press.

[19] Mitchell, T F (1975) *Principles of Firthian Linguistics*. London: Longman.

[20] Kawai, H. Toda, T. Yamagishi, J. Hirai, T. J. Ni, Nishizawa,T., Tsuzaki, M. and Tokuda. K. XIMERA: A concatenative speech synthesis system with large scale corpora. IEICE Trans. Inf. Syst. (Japanese Edition), J89-D(12):2688–2698, Dec. 2006.

[21] Black, A.W. Taylor, P. and Caley, R. The festival speech synthesis system.http://www.festvox.org/festival/. Young M., The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[22] Schroder M. and Trouvain J. The German text-to-speech synthesis ¨ system MARY: A tool for research, development and teaching. International Journal of Speech Technology, 6:365–377, 2003.

[23] Omar, A. (1985) Dirasat Al–Swat Al–Lugawi.Cairo: Alam Al– Kutub.

[24] Eriwn, W. M. (1963) A Short Reference Grammar of Iraqi Arabic.Washington: Georgetown University Press.

[25] Mitchell, T F (1975) Principles of Firthian Linguistics. London: Longman.Tokuda, K. Zen, H. and Black A., An HMM-Based Speech Synthesis System Applied to English. IEEE TTS Workshop 2002. Santa Monica. California, USA. 2002.