International Research Journal of Engineering Science, Technology and Innovation (IRJESTI) Vol. 1(2) pp. 47-56, May 2012
Available online http://www.interesjournals.org/IRJESTI

Full Length Research Paper

An Arabic- English interactive system (Aeis)

Ali A. Sakr

Computer Engineering Department, Faculty of Engineering, KFS University, Egypt
E-mail: ali_asakr@yahoo.com

Accepted 27 April, 2012

Researchers, international traders, and politicians necessitate a common interactive language to deal and keep their work secure. Using a direct language translator enables customers to deal with others; each one uses his own mother languages. This machine fulfills security, and confidence. The translation machines include in their memories, databases for synonyms, and vocabularies for the bi-languages. This system may be used in teleconferences and international political committees. This paper presents a modular system that translates Arabic to English and English to Arabic (ATE-ETA). This necessitates standard references for Speech to Text (STT), Text to Speech (TTS) and Text to Text Translators (TTTT, 4T's). The paper test statistically, the accuracy of transformation, it gave encourageous results.

Keywords: Interactive Language, Translators, International traders, Arabic to English.

INTRODUCTION

The secure interactive committees necessitates on line, live talks and shared thoughts. The interactive bi-languages system fulfills these requirements, and enables the consultations to interact together. This paper presents a modular translator that translates English to Arabic and vice versa. This necessitates trained voice to text (dictation) and trained text to voice (reader) systems. The system model is shown in figure 1.

AEIS is speech interface software that translates between Arabic and English languages. AEIS necessitates help software package for Speech Recognition (SR), Speech Synthesizer (SS), Semantic Interpretation (SI), and Pronunciation Lexicon System (PLS). NaturalReader and Readplease plus, are software that transfer Text to Speech (TTS), could be trained to read Arabic text, Arabic version. STT secretary and Dictation software could be trained to type the Arabic text. Dialogue behavior, and the linguistic knowledge (phonetics, pronunciation, intonations, dialect, thesaurus and emotions) are needed to develop the system. Programming experience and familiarity with probability of the augmented utters are also needed for the system.

The paper starts in chapter II by a review for the work done, chapter III explores the design process for the speech to text dictation process, text to text translation and text to speech reading systems, and chapter IV explores the results of statistical tests , conclusion, and hint for the future related work.

Review for the work done

Machines that convert TTS and STT help in learning the system. SR Algorithms must consider the noisy channel model, Hidden Markov Models (HMM), and GMM (Gaussian Mixture Model) to understand speech (Martin 2008). SR systems can recognize the voices of the individuals, and express them in the text format (Ellis and Wendy 2009). This is done by analyzing the sound frequencies to understand them. SS proceeds the text and synthesizes the sound frequencies to deliver the perfect pitch. The standard voice is applied with AI tools to train the system that converts the accents to their standard tones. Acronyms must be considered. Language Acquisition Software (LAS) support time tense, prepositions, learning symbolic languages, and other filling words (Ellen and Alii 2009). LAS are software, based on the acquiring language, and the gradual parameter-setting model. The acquisition process enables LAS to represent signals, textual language, and sentences' structure. LAS test the sentence compatibility and grammar of sentences and support the inductive linguistic data. Recognizing speech is based on analyzing the spectrum of characters, and words. Speech emotions, utterance, vowels, consonants, and speaker identity are characterized in LAS (Ostendorf 2010). Frequency space of speech is between 300 Hz and 4 kHz. Speech Production System (SPS) simulates the

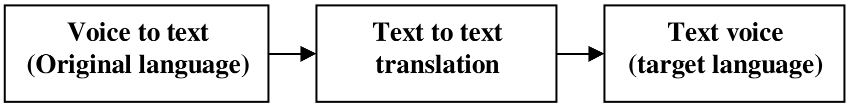

Figure 1. The Interactive System Model

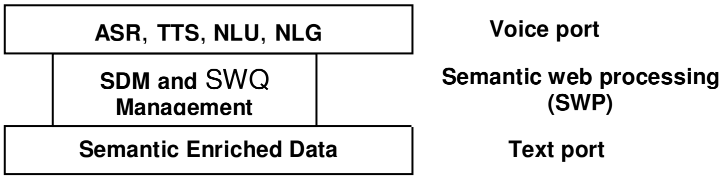

Figure 2. The Model for QoSR Semantic Web

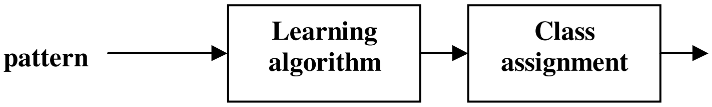

Figure 3. Pattern Recognition System

vibration of vocal folds to produce sounds, and resonance that characterize the vocal tracts (Narayanan and Alwan 2005). Precision of English dictation machines are affected by some elements like: error rate, sex voice (male/ female), word complexity, pauses, vowels, hesitations, lip smacks, filling fragments, and non-speech noise. About 82% of errors are detectable and correctable, about 7% are detectable but un-correctable and about 11% are undetectable. About 60% of the detectable errors are syntactic, 50% are semantics, 14% are both syntactic and semantic (Jurafsky and Martin 2006). It is hard to translate vowels or filling fragments, since they are slang and un-recognized in standard languages. Fragments are about 2% of words, in average. A fragment takes about 280ms, while vowels have shorter periods (Massimino and Pacchiotti 2005). SS necessitates AI training tools, smoothing techniques, Bayesian analyzer, and Natural Language Processing (NLP) to produce TTV subsystem. The conversation system includes an on-line Automatic SR (ASR), Speech Verifier (SV), STT, TTTT, and TTS converters. TTTT proceeds the textual substitutions (acronyms, and ambiguities) to enhance the AEIS. TTTT and SR Grammar (SRG) inter-operate to produce a well built sentence. The Speech Analyzers (SA) checks out the utters after being recognized. SS manage phonetics, pronunciation, and voice files. SA may fail to manage emphasized pauses, or emotive expressions (Jim 2008). Video Services (VS) can apply AEIS for immediate translations, with multiple pronunciations (Stuart 2007). Multimodality services can deal the XML videos for semantic web (Douglas 2007). The multimodality web services provide audio/video and media mixing records. The interactive modality investigates the hand-written text and transient audios. Quality of Speech Recognition (QoSR) is based on NLP and Natural Language Understanding (NLU). Figure 2 explores the elements affecting QoSR. The NLU is necessary to get an error free text, specially, the filling pronunciations, which result in a better Natural Language Generation (NLG). The Semantic Dialogue Management (SDM) and Semantic Web Query (SWQ) result in a better QoSR, as shown in figure2.

AURORA was explored at (Schlesinger 2007), as a SR system used for teleconferences. AURORA improves the word accuracy, and reduces error rate. It is installed on the server, all clients can browse it. AURORA allows rich speech interaction, exchanges data between clients and server, and synchronizes the multimodality services. The interactive modality investigates the Optical Character Recognition (OCR) systems, and biometrics applications. OCR concerns with recognizing the handwritten letters, and convert them into standard typed letters. Biometrics concern with Face Recognition (FR), Finger Prints Recognition (FPR) and military applications like the Automated Target Recognition (ATR). These FR, FPR and ATR are out of our scope. The Pattern Recognition (PR) systems include learning algorithms (LA) and pattern classifiers (PC) as shown in figure 3. Both PC and LA use the supervised methods that apply Linear Discriminated Analysis (LDA) and statistical Bayesian decision approach to improve QoSR (AlAnzi 2005).

Speech Production Systems (SPS) produces properly the output of SS after filtering and smoothing (AlAnzi 2005). The sound source may include noisy sounds. The vocal tract can be modeled as an acoustic tube with resonances and anti resonances. Changing the frequency, results in a different speech signal. Aspects of

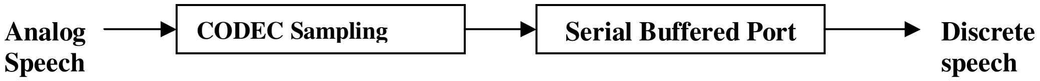

Figure 4. Digitizing the Speech Signal

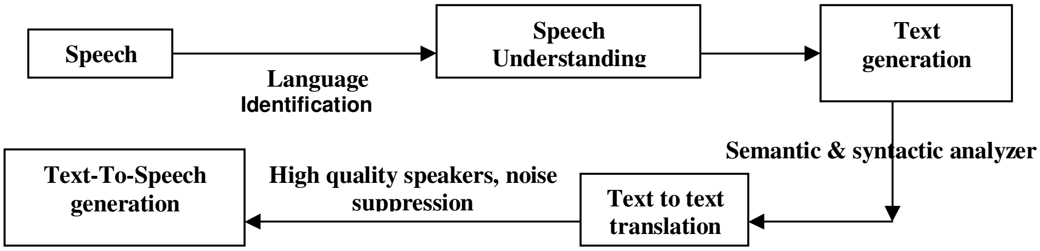
Speech
Language Identification
Speech Understanding
Text generation
Semantic & syntactic analyzer
Text to text translation
High quality speakers, noise suppression
Text-To-Speech generation

Figure 5a. Diagram for the Auto Speech Translator System

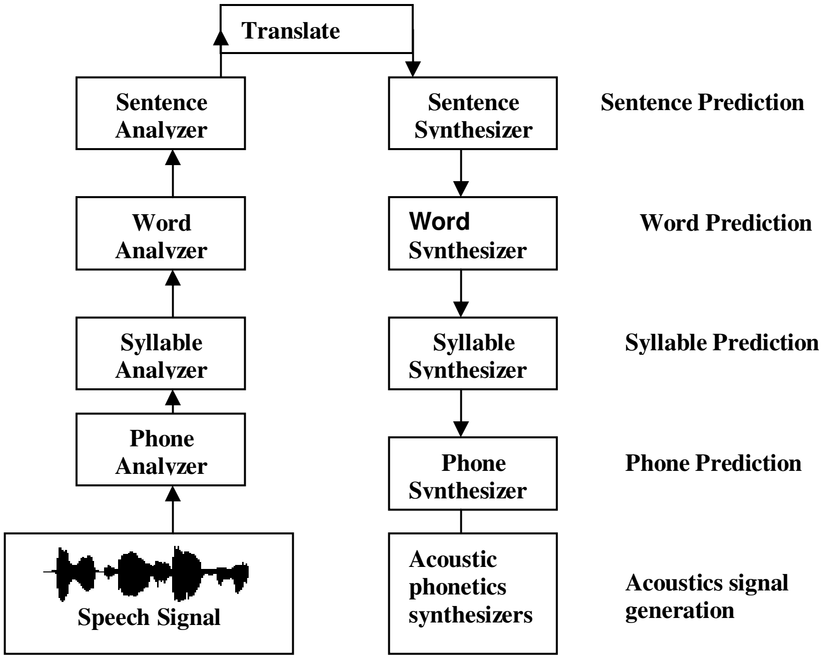

Figure 5b. Detailed Auto Speech Translator System

a signal depend on its sequence numbers and sampling rates. Figure 4 indicates the process of digitizing the speech signal.

The conversation must consider the ASR, dialogue manager, NLU, and dialogue generation. SPS applies conversation evaluation utility to evaluate the QoSR. Speaker Recognition tasks necessitates: defining Speaker Identifier, Gender ID, mode of speaker, and multiparty conversation. Mode of speaker affects the utterance and speed of word sequence speech. (Beesley 2006)

System implementation

The proposed system is explored as in figure 5. Speech Understanding Systems (SUS) merge technologies from pattern recognition, NL, DSP, and statistical QoSR to understand the speech vocabularies. The phones are analyzed due to the frequency. For English, there are about 32 phonemes and 42 in Arabic (Schlesinger 2007). English phonemes are vowels, semivowels, pauses, and consonants. There are three branches of phonetics: articulator phonetics in which the

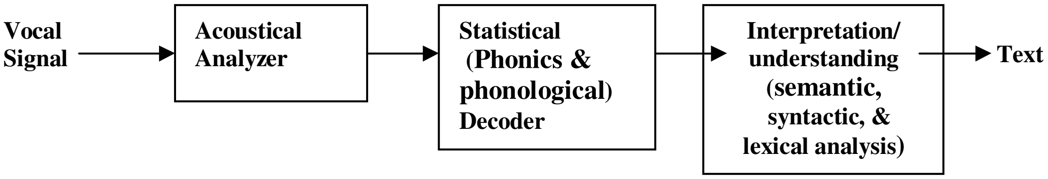
Vocal Signal
Acoustical Analyzer
Statistical (Phonics & phonological) Decoder
Interpretation/ understanding (semantic, syntactic, & lexical analysis)
Text

Figure 6. A Scenario Model for VTTIS

Start	**Data Length**	**Sampling rate**	**Compression type**		**data**

Figure 7. Format of the Digital Wave

sounds are produced by the vocal system; acoustic phonetics that sounds through the analysis of the waveform and spectrum; and auditory phonetics that studies the perceptual response to sounds as reflected trials. The representation of acoustic signal considers the effects of time variation, vocal tracts, losses due to heat conduction and viscous friction at the vocal tract walls, softness of the vocal tract walls, emission of sound at the lips, and excitation of sound in the vocal tracts (Ahmed 2006). The next sections explores the three cycles: voice to text, text to text and text to voice.

Design of voice to text interface system (vttis)

Accents, voice quality, speakers and microphone quality, have a main effect on the QoSR. Voice quality is affected by noise level, noise frequency domain, reverberation, warping, recording microphone and tongue of the standard reference. The speech is analyzed using frequency spectrogram, and pseudo model of words. Words vocabularies must be defined. The ASR is analyzed as shown in figure 6.

The VTT model necessitates interpreters, grammar analyzer, Fast Fourier transform (FFT) analyzer, learning and prediction program, and acoustic phonetic analyzer. VTT systems apply speech recognizers, to transform signals into text. They must consider dissimilarities between utters. Female and male spectra are different. Different speakers give different spectra, which must be smoothed. Increasing number of referees reduces the error. VTT model detects the silences, fragments, numbers, boundaries of clauses, and phrases. The main interruptions (uh, um, etc.) are used to announce delays. They have no Arabic translation. About 8% of English phrases use such interruption words in editing (Pereira 2005). Voice quality is affected also by Jitters spectra, and vibration of the vocal folds. Modes of phonation are either: voiceless, normal voice, whisper, breathy voice, or creaky voice. A mixture of multivariate Gaussians (MMG) acoustic model fulfills the least error rate regarding the means, covariances and variances. It trains the Viterbi training that takes the most likely solution. Viterbi training is much faster than others like Baum-Welch approach (Beesley 2006). Viterbi is applied in dialogue systems, where desired semantic output is clearer. VTT systems must analyze the spectra and intensity of sound waves. Speech is digitized by A-D conversion. Samples define the amplitude of the signal at period "t", there must be at least two samples per cycle. Less than two samples per cycle will debase the expressing of signals. The frequency for a given sampling rate is 16,000 samples/sec for microphones. The digital signal is represented as an integer value; 16-bit for resolution (can express values between -32768 to 32767). The digital signal has the format as shown in figure 7.

The power of a speech is proportional to the square of amplitude. Not all vowels have regular resonances, speech frequencies, Multimodality or the power of speech (Alansary et al, 2006). Conversational systems must include VTT dictation system, TTTT, and TTV phonic system as shown in figure 8. This fig indicates the state diagram for a word recognition process.

The recognizer is to recognize the utterance during training for building the sentence grammar. The recognizer must learn the context variations. VTT system applies the Baum-Welch estimation algorithm, to emulate the statistics of the training database. This minimizes the error rate. Often, the test dataset contains data never met in the training database. Training aims to find the proper character for the given utter. After few iterations of training, the items represent the training data with a percentage of error. Increasing the training samples results in reducing the error rate and improves the certainty to identify a specific character. Estimated, items are based on the most likelihood from the language model. Average values and standard deviations are used to smooth the phonics and help to get the perfect text.

Romany numbers are dictated in character form, it must be defined well in database to be retrieved as numbers in the other language. The text integrity depends on the microphone quality, SR Grammar Specification (SRGS), Semantic Interpretation for Speech

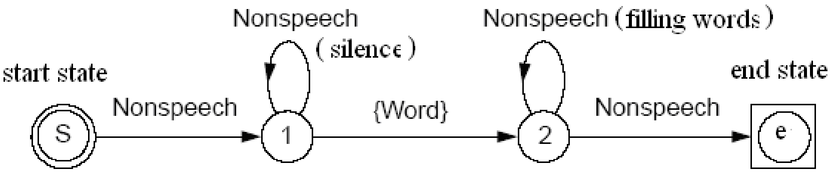

Figure 8a. State Diagram for Word Detection

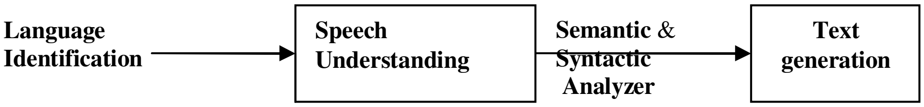

Figure 8.b. Dictation System

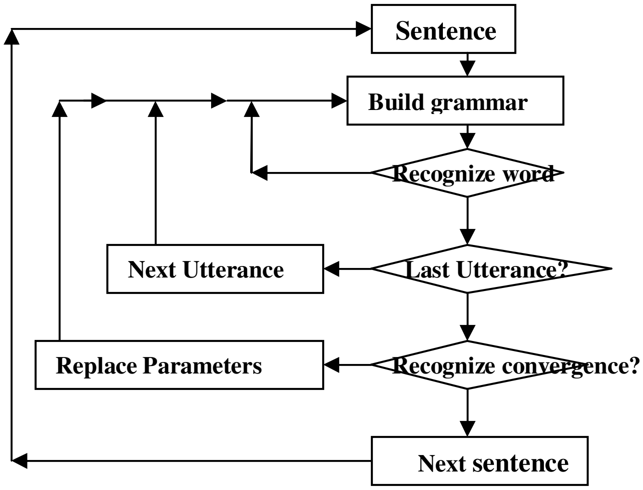

Figure8c. Training for a Sentence Generation

Recognition (SISR), Speech Synthesis Markup Language (SSML), and Pronunciation Lexicon Specification (PLS). The basic emotions such as: scream, surprise, happiness, anger, fear, disgust, and sadness; could not expressed. There are other elements like confusion, certainty, frustration, annoyance, anxious, boring, courageous, fatigue, disappointments, amuse, surprise, noise, and lying, These factors could be sensed from speech, so when they are transformed into text give no sense. Speech could be characterized with: pitch, loudness and speaking rate. Jurafsky [7] could classify these emotions for about 4000 conversations, as in Table 1. Many Arabic dictation systems like via voice, secretary, and Arabic-dictation-pro is efficient in translating VTT. The more trained dataset can result in a proper certain text. The training system concerns on the data reference, tools for decisions, and dialects.

Conversations vary regarding: duration of speech, vowels, syllables, utterances, and speaking rate. Phones are not always homogeneous. Each phone has 3 sub-phones: Start, mid, and End. Digitizing the voice signal necessitates: sampling at periodic times, and measuring amplitudes of the samples. The sampling rate is double of frequency. Human speech is less than 6 KHz, so need about 12K sample/sec. Frequency analysis of speech is based on Gaussian mixture model (GMM), and weighted sum of the multiple Gaussian distributions regarding (mean, variance and covariance). Z transform of a signal, Discrete Fourier Transform (DFT), Hidden Markov Models (HMM), Isolated Word Recognition (IWR), are applied to analyze and recognize the utters and yield the proper words. The recognizers find the optimal start/stop of utterance with respect to the acoustic model. The system can recognize sequences of words or non-speech events. Pattern detection is accomplished by parsing the phonic sequence and generates the relevant character sequence. This requires all possible combinations of symbols generated from the speech patterns. The overall best hypothesis is assigned. Several algorithms are applied for time synchronous (e.g. Viterbi), state synchronous (e.g. Baum-Welch), stack decoding (e.g. the Best - First), and hybrid schemes (e.g. Fast

Table1. Frequencies of Voice Regarding Emotions

Aspect	**Mean frequency (HZ)**	**Standard deviation (HZ)**
happy	330	109
sad	222	54
angry	350	84

Matching). Artificial Neural Networks (ANN) are applied for the supervised learning process (AlAnzi 2005), where network simulates the input and estimates the weights to adjusts itself automatically to decide the best corresponding output. The VTT system is based on transcription of words, and is evaluated by word error rate. Speech understanding is estimated by the measured number of words successfully transcripted. Dictation faults result in erroneous sentence which let to erroneous translated sentence. ASR system is based on dialogue interpreter, ambiguities, filling words, direct and indirect speech, explicit and implicit expressions, and emotions. Most filling words and emotions are not interpretable, which results in incomplete mapping. The next subsection explores the TTTT problems.

Design of text - text interface system (ttis)

Dictionaries map Arabic and English words. AI tools are necessary to search for acronyms. The syntactic and semantic analyzers / developers are applied to form the sentence structure. The English sentence may include verbs with present, past, future, and pp tenses, while Arabic verbs have no pp tense. Nouns, adjectives, adverbs, pronouns, prepositions, conjunctions, objects, prepositions, articles, abbreviations, fragments, and possessives, all exist in both Arabic and English. Sentences types may express direct or indirect speech, clauses or phrases, joint sentences, complex sentences, conditional sentences, passive or active sentences, and interrogative sentences. These sentences are defined grammatically and semantically, to give the right meaning. Punctuations, brackets, commas, dashes, exclamation marks, hyphens, parenthesis, periods, quotation marks, and semicolons, all have the same meaning in both Arabic and English. Numbers in Arabic are read in special manner that differ from that in English, but when being translated they are mapped well. Repeated sentences must also define their tenses. Auxiliary verbs, definite and indefinite verbs, and irregular verbs must be defined in the DB of the translating machine. Arabic grammar must consider many terms like verbal and noun clause, adjectives, adverbs, pronouns, propositions, conjunctive, interrogative expressions, direct and indirect speech, joint and complex sentences, vocabularies, the special five nouns, masculine plural, feminine plural, irregular plural, definite and indefinite nouns, verb tenses, the appositions, exceptions, and many grammatical rules that have their particular affect the meaning of sentences. There are in Arabic some verbs with one or two objects, some with no object and other with no subject. The nominal sentence starts with nouns that may be gerunds or delivered. Verb may be in future, present or past tense. The gerunds may be abstract nouns or not. These nouns have standard reference et al., Gerunds may be three, four, five, or six letters verb. Each of them has its own performance. The noun may point to a tool, time, place, object, preference, a matter of time, or a matter of place. The imperfect names, elongated names, regular or irregular names, masculine or feminine plurals, these names have no relevant in English. English language does not distinguish between two or many persons in pronouns or verbs. English language does not distinguish also between masculine and feminine plurals. In English verbs have only one object, while in Arabic some verbs have more than an object, and some have no object. In Arabic, pronouns may be connected, hidden or disconnected from verbs, while in English it is disconnected only. Pointing names in English (this, that and those), while in Arabic they differ regarding number of persons, and gender. As well, the connecting articles in Arabic differ due to number of persons, and gender. The definite English article 'the' is used for all nouns, while in Arabic the article differ regarding whether the name starts with silent letter or non-silent one (al-lunar or al solar). In Arabic, a noun may be defined by adding it to a definite noun. Noun may also defined by a calling article. Discrimination nouns with numbers have some rules in Arabic that are not in English. Compound numbers have a special deal regarding masculinity or femininity. Accusative and absolute Accusative are related to verbs in Arabic, but they are not in English. The rules for recompense, exclusion, Adjective, and Assertion nouns have no relevant in English. Arabic enjoy with the grammatical signs upon or under the letter, which are not in English. These signs differ regarding the position of noun in the sentence. Conditional statement in Arabic has many articles with many rules, but in English, there is just 'if' statements. Exclamation verbs and query statements in Arabic have their own articles and rules, while in English just use 'how good or bad'. Swear nouns are not exist in English. Negation and Prohibition verbs have many situations, in English just use 'never'. Metonymy and implicit expressions are not found in English. Arabic doesn't start with capital litters. Gerunds , interjections, blame , praise, emphatic, preference and

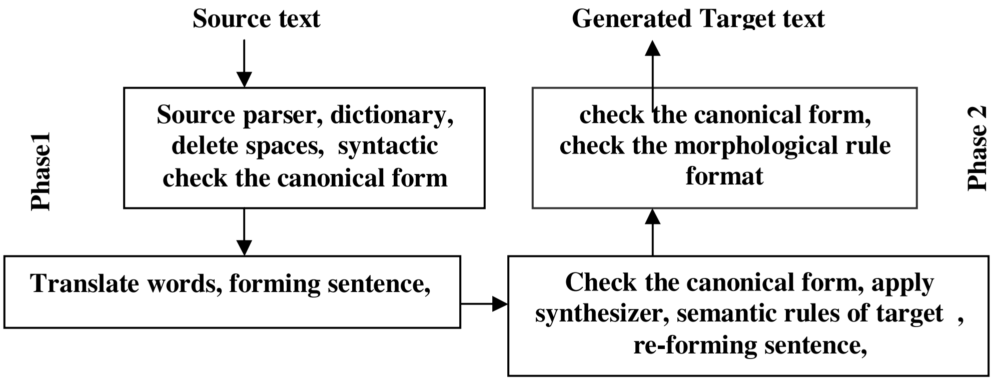
Source text
Generated Target text
Phase1
Source parser, dictionary, delete spaces, syntactic check the canonical form
check the canonical form, check the morphological rule format
Phase 2
Translate words, forming sentence,
Check the canonical form, apply synthesizer, semantic rules of target , re-forming sentence,

Figure 9. Components of TTT Proposed System

slander expressions are more definite in Arabic. All English statements are nominal sentences, but in Arabic, there are verbal, nominal and pseudo sentences. Therefore the Arabic grammatical system is much complicated than that for English. The English grammar form can be summarized within the next structure:

<English Sentence> = <Simple Sentence> | <Compound Sentence>

<Simple Sentence> = <Declarative Sentence> | <Interrogative Sentence> | <Imperative Sentence> | <Conditional Sentence>

<Compound Sentence> = <Simple Sentence> <conjunction> <Simple Sentence> | "Either" <Declarative Sentence> "or" <Declarative Sentence> | "Either" <Imperative Sentence> "or" <Imperative Sentence> | "Neither" <simple Sentence > "nor" <simple Sentence >

<Declarative Sentence> = <subject>

<Subject> = <simple Sentence > | <compound Sentence > | <noun >

<Predicate>= <verb>| <object> | <prep. Phrase>

<Simple Sentence > = <noun phrase> | <nominative personal pronoun>

<noun phrase> = "the" < noun> | <proper noun> | <non-personal pronoun> | <article> [<adverb>* <adjective>] <noun> | [<adverb>* <adjective>] <noun-plural> | <proper noun-possessive> [<adverb>* <adjective>] <noun>| <personal possessive adjective> [<adverb>* <adjective>] <noun> | <article> <common noun-possessive> [<adverb>* <adjective>] <noun>

<Noun> = <noun> [<prep phrase>*] | < adjectives> │< pronouns> │< number > │ <prepositions > │<adverbs

<adjective> = <adjective> ("and" | "or") <adjective>

<prep phrase> = <preposition> <object

< Article> = < the> | < a> | <an>

In Arabic, the grammatical structure of sentence can be defined as:

<Arabic Sentence>:= <noun phrase>|< verbal phrase>;

<Verbal phrase>= <verb>[< noun phrase>*< preposition>] | <verb>< subject> < object>

<Noun sentence> =: <noun>[<noun>* <noun>] | < preposition> | < adjective><noun phrase> |
< noun ><article>< noun>

< Article> = < al > | <not verb> and <not noun> │<definite article> │< vocative article

< **Informative** phrase> = <starting noun > <informative noun phrase>.

< **Verbal** sentences> =< verb>< Subject>.

<**Number** sentence>= <Number >< discriminators>

<**Verb**>= [< present Verb>|<past verb> |<future verb> |< Imperative verb>] [<Object>| [< Pronouns>| <Prepositions >]]

<**Prepositions**>= <Feminine Prepositions >| < Masculine Prepositions > |<Plurals Prepositions>| <two person Prepositions> |<single Prepositions>

< **Sick** Verb> = <verb> [<subject>*]

<**Vocations**> = < Masculine> | <Feminine >

< **Apposition**> = < all = pronounced as kol> |<gamee'> ,

<**Subjunctive** if>= if <condition> <action>

<**Connection** nouns> = <sentence><article> <sentence>

<**Additives**> = <Prepositions> <definite noun>

<**Derivation** of Active Participle> = < Verbal Noun> | <connective Pronouns> | <Cognate Accusative> | < Emphasis> | <five nouns >

< **Derivation** of Passive Participle> = < Accusative of Distinction.> | <Exception> | <Accusative of Purpose>,

<**Word**>= <noun> │ < verb >│ <particle >

<**noun**> = <subject>| <object>| < adjectives> │< pronouns> │< number >│ <prepositions > │<adverbs>

< **Pronouns** >= < masculine> │<feminine >

< **Pronouns** >= < possessive pronoun> |< talking pronouns>| <absent pronouns>

The Arabic sentence is written from right to left. The parser precedes the source sentence, then syntactic analyzer checks the integrity of sentence structure, then the bilingual dictionary rules are applied to translate words. The destination syntactic synthesizer is applied to construct the destination sentence. Then sentence structure is scanned by semantic analyzers. These steps are shown in figure 9.

The Arabic to English to Arabic (ATETA) translator must analyze the source text, translate the source words, check the canonical form and generate the destination text. The search within DB acronyms is developed using

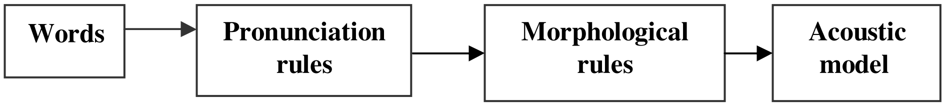

Figure 10. A Pronunciation Model

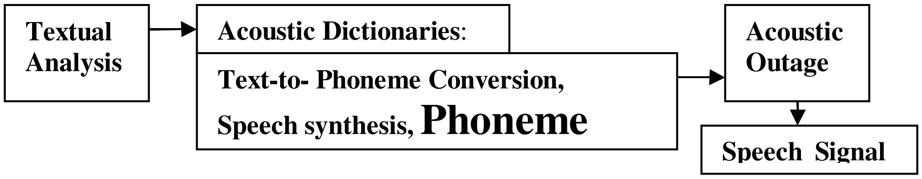

Figure 11. Model for Generating Speech (TTS)

c# and SQL. Many standard programs like Natlang, sentence-analyzer, real-time analyzer are used. Semantic rules are applied to enhance the sentence' structure. The sentence "meat eats cat" is correct syntactically, but incorrect semantically. If words of a sentence match the canonical form, this sentence is translated into the relevant phrase. Considering vocabularies and canonical rules enforces the translated sentences. Many sentences must be reformed when being translated. Declarative, interrogative, imperative, and exclamation sentences have variable structures in Arabic and English. The words like adjectives, adverbs, nouns, pronouns, conjunctions, and prepositions have different actions in both Arabic and English. Many English words have no relevant in Arabic like ummm, and other filling words. Roman numbers are read as letters not numbers, e.g. IV is read as I, V not the number four. So, when translated by conventional software programs, they do not give the correct means. Therefore, cognitive training is applied to this software to produce the proper translations. Many slang Arab words have no relevant in English, as well. Some upper and lower signs are used in Arabic to adjust the meaning of sentence. Some signs are used to extend the pronunciation of some characters, these vowels and consonants change the meaning of sentences. These attributes change the morphology, and logyphono of Arabic words. Arabic language has 29 letters with 92 different sounds, while English has 22 letters with 42 different sounds (Beesley 2006;Agrawal and Lavie 2008). Translator dictionaries use intelligent oriented search. Selecting the acronyms, synonyms, and vocabularies necessitates one to many word-pair assignment, to generate a well trained sentence (Agrawal and Lavie 2008). The proposed dictionary uses look ahead tools, and semantic synthesizer to produce an efficient TTTT. Using the TTTT model enhances the translation process. The translated Arabic by our system gives a better quality than the translation by Google by 6%. This test was done for 250 sentences from different subjects.

Text to speech (tts) systems

Evolution of TTS quality has three generations: character synthesizers, formant synthesizers and concatenative synthesizers. Synthesizing emotions, impressions, and temper; gives a more powerful value for TTS. Emotions, extroversion, and passion could be expressed in speech but not text. The phonic library must be linked to the dictionary outage and phonic training program. A pronunciation system is shown in figure 10. The pronunciation rules include information about the contextual speaking rate, and acoustic frequencies. Natural reader and Oddcast are efficient software that transfers TTS.

Figure 11 indicates a scenario for generating speech. This necessitates trained databases for standard phonics that map sentences to their relevant speech. The system need at least 10 thousands utters, to get the mean vocal utter for each character and concatenate these utters to form words.

Enhancing TTS system is accomplished by voice morphing (Jurafsky and Martin 2006). Speaker verification is done via a biometric measure, where every person has a unique voiceprint (Massimino and Pacchiotti 2005). The morphing process aims to make the smooth transition from speech signal to another. The implementation of the morphed speech signal would have the duration of the signal. There are three stages for morphing: the envelope; the pitch; and the pitch peak. Speech morphing aims to preserve the characteristics of the start and end of signals, and to smooth the transition between them. Arabic morphological rules must consider the effect of female / male sounds. In English, the verb does not vary due to gender. The process to obtain the morphed speech signal includes the envelope information, dynamic time warping, and signal re-estimation. Speech morphing can be achieved by transforming the signals from the acoustic waveform to splited frames. This describes the average spectra at each frequency band. If two signals with two distinct pitches are crossfaded, this will result in two distinct sounds will be heard. To do this match, the signals are

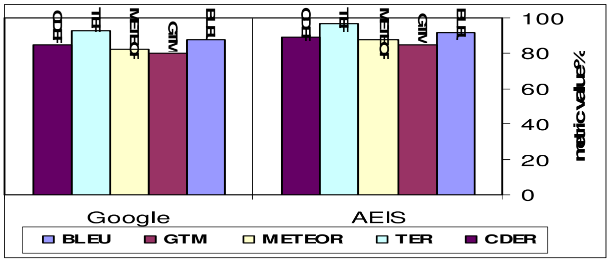

Figure12. A comparison between the intelligent AEIS and Google translation system

stretched and compressed so that sections of each signal match in time. The interpolation of the two sounds can be performed to create the intermediate sounds. The morphing algorithms contain a number of fundamental signal processing like sampling, discrete Fourier transform and its inverse, signal acquisition, interpolation and signal re-estimation. Speaker verification is based on the likelihood ratio, and hypothesis tests. Generation of utterance need no lexical interference, but needs bigger database, while using natural words is easier to pronounce and needs smaller database. After generating databases, it must make recordings consistent; this is done by using constant pitch (monotone). It should avoid pronunciation problems, and keeps speaker consistent.

RESULTS AND CONCLUSION

The analyzer and synthesizer concerned in dealing punctuation and spaces of natural language. They deal affixes, lexical, semantics and syntaxes, of the different sentences. Many automated measures have been proposed to test the speed and accuracy of the trans-language process. These evaluation tools must be fast and cheap (Banerjee and Lavie 2005). Most efforts focus on measuring the closeness of the output to human translation. The **BLEU** metric demonstrated a high correlation with system adequacy and fluency. It reflects the entire content of the reference translation. **GTM** used the mean of precision. **METEOR** use a weighted mean, based on synonyms (uniform weight = 1/N). **TER** metric is applied on the word level. It measures the number of edit operations needed to fix a candidate translation. **CDER** metric is based on reordering of word blocks. These metrics are applied for AEIS system, for 1000 Arabic sentences, with average 5.1 words per sentence and standard deviation= 3.2 words. A comparison between Google translation and AEIS translation is shown in figure 12.

The AEIS fulfill an enhanced fluency for words. The available datasets for English and Arabic are drawn randomly from NIST, and internet documents News. The test dataset contains around 1000 sentences, composed of 5143 words. Three professional -referees translators, were requested to generate the translated texts, and review the grammatical product. They were requested to fix errors resulting from the lack of grammar or semantic rules represented in a AEIS system. The post editing was conducted by two experts in the Arabic language, to ensure that the sentences are written and spoken in a typical Arabic style. Basic preprocessing was applied to all datasets. These preprocessing include different punctuations with different degrees. It is observed that the translated part of AEIS results in a better score than Google. These results are statistically significant and confident. AEIS sustains an improvement, it is observed that the delay between spoken and heard translation is 1.4sec., in average. This is the sum of delays due to read any text, TTTT, and Dictation software. Analyzing the AEIS performance, indicate that more than 96% of the translated tasks are acceptable. AEIS introduces improvements over Google translation. AEIS considers some features in typical Arabic style, which are not captured by other MT such as Google and Sakhr. The test dataset has been categorized into two groups according to the difficulty of the sentences. Difficulty is judged by linguists based on the complexity of the structure of the sentences as well as its length. It should be mentioned that the number of words per sentence, in the Arabic language is shorter than it , in English. One Arabic word may be translated into an English statement, e.g. Anolzemkomoha, which composes verb, subject, object, and interrogation expression, and it means “can we restrict you to do it?". So, Arabic language is an agglutinative language. Comparing the results of AEIS system and Google translation, it is observed that Goggle results in the lower score, AEIS shows higher values for metrics reach 2-9 % over Google. This implies that AEIS outperforms Google in generating sentences. It is noteable that English sentences are nominal only, while Arabic sentences are nominal, verbal or pseudo sentence. The applied syntactic grammar of AEIS showed that its results more formed, and structured than Google dataset. Evaluation of transformation system considers: quality of microphones, utters, percentage of formality, efficiency of dictionary and its vocabularies, efficiency of speech synthesizer, elapsed time for transformation, percentage of inconsistent or ambiguous words, quality of ASR, percentage of failed and inappropriateness, ability to understand, ability to sustain user and team satisfaction. Most systems use regression to train weights to get the perfect outage. AEIS uses adaptive ANN to estimate weights. Parameters are used to adapt the synthesizing process are: mean recognition accuracy (MRA), elapsed time (ET), requests for ambiguity (RA). Oddcast and readanytext are used for TTS. Matlab is used for analyzing and synthesizing signals and constructing DB for TTS, TTTT translation demo translates Arabic to English (ATE) and vice versa. Many systems like systransnet, Natlang can translate ATE. Interfaces for the proposed system are carried out using c#.net, with SQL developer. The input speech is documented using dictation software, which dictates the software and produce text, text is translated using TTTT, and turned into speech using TTS using readers or talkme software. Microphones and standard references have a main factor in quality of generating the text. AEIS is evaluated using 1000 sentence (500 Arabic and 500 English) in multiple fields. Arabic and English morphological synthesis rules are trained using Matlab tools. The efficiency of the AEIS system is better, concerning the technical fields. The performance of system is function of number of words defined in databases, acoustic similarity, number of synonyms words at each point, number of possible combinations of vocabularies per word, speed of processor and memory size used. A statistic was recorded for 10000 utterances, 4% errors were detected, about 12% pauses and fragments, 11% hesitations, 9% lip smack, 13% breath and non-speech noise, 42% operator utters, 6% echoed Prompt, and 3% filling words (uh, um, you know, so, I mean, etc.).

This paper presented an effective and valuable project, which is useful in trading, technical conferences and politics. Up to now, no complete system is generated for interactive-speech translation between Arabic and English. Millions of syllabus and expressions in standard and slang language include slander, swards, blames, praises, and distress must be added. Many expressions vary due to mode of conversation. Many training phones are needed to give the proper utter. The system still need many efforts to be in practical form, in spite of, the author had worked for more three years in the three phases to get the present situation. It still needs many datasets to extend the database to all domains in both languages. It needs fund to make it industrial. The team-work may conclude 3 language experts, 3 computer programmers, 2 computer engineers for speech analysis and synthesizing, the work can be prototyped in less than 6 months, the total cost needed may be 1M$, including devices and software.

REFERENCES

Agrawal A, Lavie A (2008). METEOR, M-BLEU and M-TER: Evaluation Metrics For High Correlation with Human Rankings of Machine Translation Output. In Proc. of the 3rd Workshop on Statistical Machine Translation, pp. 115-118, Ohio.

Ahmed A (2006). A Direct English –Arabic Machine Translation System, IT J., 4 (3).

Alansary S, Nagi, M, Adly, N (2006). Generating Arabic text: The Decoding Component of an Inter-lingual System for Man-Machine Communication in Natural Language. In the 6th International Conference on Language Engineering, 6-7 December, Cairo, Egypt.

AlAnzi F (2005). Automatic English/ Arabic Translation Tool, J. KSU,

Banerjee S, Lavie, A (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In Proc of ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and Summarization, Ann Arbor.

Beesley K (2006). Arabic Morphological Analysis ", Xerox Research Center, Melan, France.

Douglas A (2007). Machine Translation, , ISBN 1855542-17x.

Ellen E, Alii N (2009) Voice morphing , CRC..

Ellis M, Wendy H (2009). "Speech Synthesis and Recognition", CRC, ISBN 0748408576.

http://www.nist.gov/speech/tools/

Jim L (2008). VoiceXML Introduction to Developing Speech Applications, Prentice-Hall..

Jurafsky D, Martin JH (2006). Speech and Language Processing; Prentice Hall; 2nd edition

Jurafsky Martin (2008). Speech and Language Processing, (2nd edition). Prentice-Hall.

Massimino P, Pacchiotti A (2005). An automaton-based machine learning technique for automatic phonetic transcription, INTERSPEECH-

Narayanan S, Alwan A (eds.) (2005).Text to speech synthesis, Prentice Hall.

Ostendorf B (2010). Cross-fertilization between ASR and TTS areas, Prentice Hall.

Pereira F. (2005). Warren, "Definite clause grammar for language analysis", Artificial Intelligence, Vol. 13, pp. 231 - 278,

Schlesinger H (2007). Statistical and structural pattern recognition. J. Pattern Recognition Soc. v1,.

Stuart R (2007). Natural Language Processing and Applications, AI J.v2, #5.