ISSN 2277-3061
Volume 15 Number 7
International Journal of Computers and Technology

An English -Arabic Real Time System (EARS)

Ali A. Sakr
Computer Engineering Department,
Faculty of Engineering, KFS University, Egypt
ali_asakr@yahoo.com

ABSTRACT

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 (AE-EA) via a real time interactive system. This necessitates standard references for Speech to Text (ST) , Text to Speech (TS) and Text to Text Translators (TTT). The paper test statistically, the accuracy of transformation, it gave encourageous results.

Keywords: Speech to Text (ST) , Text to Speech (TS) , Text to Text Translators (TTT). Interactive systems, voice to text (VT), and real time systems.

1. INTRODUCTION

The secure interactive committees necessitates on line 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 systems), text to text dictionary (intelligent dictionary) and trained text to voice (reader systems). The system model is shown in fig1.

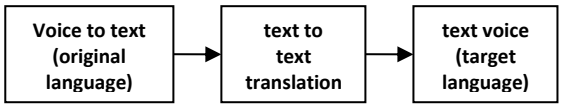

Fig1-a. The Interactive System Model

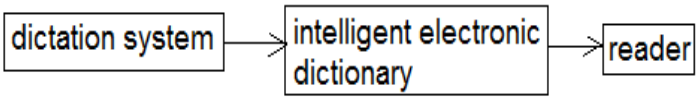

Fig1-b. The simplified Interactive Model

EARS is a speech interface software that translate between Arabic and English languages. EARS 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 (TS), could be trained to read Arabic text in Arabic version. ST 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 and uncertainty 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.

2. REVIEW for the WORK DONE

Machines that convert TS and ST 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 [1]. SR systems can recognize the voices of the individuals, and express them in the text format[2]. 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 to standardize the accents and tones. Acronyms must be considered. Language Software (LS) support time tense, prepositions, learning symbolic languages, and other filling words [3]. LS tests the sentence compatibility, and grammar of sentences to 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 LS[4]. Speech Production System (SPS) simulates the vibration of vocal folds to produce sounds, and resonance that characterize the vocal tracts[5]. Precision of English dictation machines are affected by some elements like: non-standard error rate, sex voice (male/ female), word complexity, pauses, vowels, hesitations, lip smacks, filling fragments, and non-speech noise. About 88% of errors are detectable and correctable, about 7% are detectable but un-correctable and about 5% are undetectable. About 60% of the detectable errors are syntactic, 50% are semantics, 14% are both syntactic and semantic[6]. It is hard to translate filling, slang and un-recognized words in standard languages. Fragments are about 2%

of words, in average. A fragment takes about 280ms, while vowels have shorter periods[7]. Synthesized speech (SS) necessitates AI training tools, smoothing techniques, Bayesian analyzer, and Natural Language Processing (NLP) to produce a voice subsystem. The conversation system includes an on-line Automatic SR (ASR), Speech Verifier (SV), ST, TTT, and TS converters. TTT proceeds the textual acronyms, and ambiguities to enhance the EARS. TTT 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[8]. Video Services (VS) can apply EARS for immediate translations. Multimodality services that deal the XML videos in semantic web, provide audio/video and media mixing records[9,10]. 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). Fig2 explores the elements affecting QoSR. The NLU is necessary to get a reduced error text, specially, the filling pronunciations, which result in a better Natural Language Generation (NLG). The Semantic Dialogue (SD) and Semantic Query (SQ) result in a better QoSR.

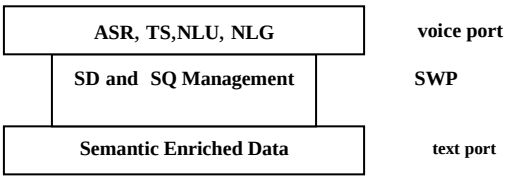

Fig2. The Model for QoSR Semantic Web Processing (SWP)

AURORA was explored at [11], as a SR system used for teleconferences. AURORA has a good word accuracy, and a reduces error rate. It could be installed on the server, and all clients can browse it. AURORA allows rich speech interaction, and exchanges data between clients and server. It handles multiple-languages like English, German, French; 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 Automated Target Recognition (ATR). FR, FPR and ATR are out of our scope. Pattern Recognition (PR) systems use Linear Discriminated Analysis (LDA) and statistical Bayesian decision approaches to supervise and improve QoSR [12].

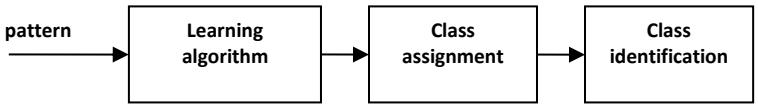

Fig3. Pattern Recognition Model

Speech Systems (SS) produces the speech in standard understandable form [12]. The sound source may include noisy sounds, due to the improper microphones and speakers. The vocal tract can be modeled as an acoustic tube, with different frequencies. Aspects of the generated signal depend on its sampling rates, and frequencies of utters. Fig4 indicates the process of digitizing the speech signal. These samples are stored in memory. These stored samples are quantized to produce the speech in the target language.

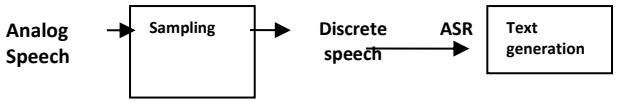

Fig4. Digitizing the Speech Signal

The conversion of speech to text must consider the ASR, dialogue manager, NLU, acronym , filling words, and speech synthesizer (SS). SS applies conversion evaluation to evaluate the QoSR metric. Speaker Recognition tasks necessitates: defining Speaker, mode of speaker, and multiparty conversation. Mode of speaker affects the utterance and speed of word sequence. [13]. Standard pronunciations helps to produce the perfect letters, which lead a perfect sentence. Deviation from standard form lead to mal- recognition for characters, which lead to faulty words. This gives faulty translation , which lead to faulty sentences in the target language. Errors may be positive or negative, which affect the resultant product in the target langue. Most errors occur during the translation process, where synonyms may give a different meaning. ASR incises sentences into words, each word is incised into characters, which are incised into tones. These tones are stored and used to generate the corresponding text. The greatest problem is the words that have multiple- synonyms, and words that have closed pronunciation, that give many meanings.

3. SYSTEM IMPLEMENTATION

The proposed system is explored as in fig 5. Speech Understanding Systems (SUS) merge technologies from pattern recognition, NL, DSP, and statistical QoSR to understand the speech vocabularies. The phones are analyzed , there are about 32 phonemes in English and 42 in Arabic[11]. English phonemes are like vowels, semivowels, pauses, and consonants. There are three branches of phonetics: articulator phonetics that are produced by the vocal system; acoustic phonetics that sounds through the analysis of the waveform. 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 are difficult to represent[4]. Next sections explores the three cycles: voice to text, text to text and text to voice.

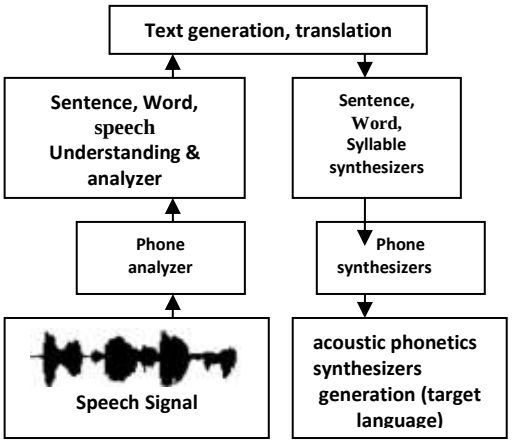

Fig5. Detailed Auto Speech Translator System

3.1. DESIGN OF VOICE TO TEXT INTERFACE SYSTEM (VT)

Accent, voice quality, speakers and microphone quality, have the 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 model of words. Words vocabularies must be defined. The ASR is analyzed as shown in fig6:

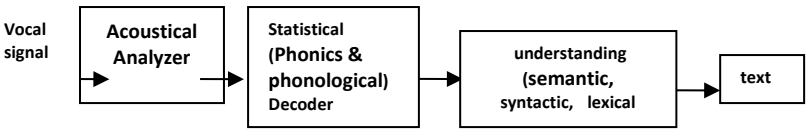

Fig6. A Scenario Model for VT System

The VT model necessitates interpreters, grammar analyzer, Fast Fourier transform (FFT) analyzer, learning and prediction programs, and acoustic phonetic analyzer. VT 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 and standardized. Increasing number of referees reduces the error. VT model detects the silences, fragments, spoken numbers, boundaries of clauses, and phrases. The main interruptions (uh ,um, etc.) are used to announce delays and must be suppressed. They have no Arabic translation. About 8% of English phrases use such interruption words in editing [15]. 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. Mixtures 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 Baum-Welch approach[13]. Viterbi is applied in dialogue systems, where desired semantic output is more clear. VT systems must analyze the spectra and intensity of sound waves. Speech is digitized by ADC. 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 sampling rate is 16,000 samples/sec for most microphones. The digital signal is represented as an integer value, 16-bit for resolution (can express values between -32768 to 32767). Not all vowels have regular resonances, speech frequencies, or the same power of speech[16]. Conversational systems must include VT dictation system, TTT, and TV phonic system as shown in fig7. This fig, indicates the state diagram for a word recognition process.

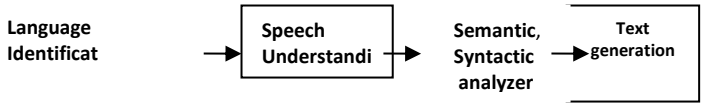

Fig7.a. Dictation System (sentence generation)

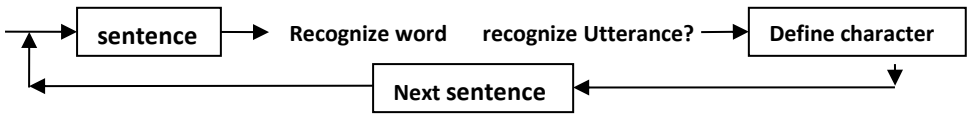

Fig7. b. Training for a Sentence Generation

The recognizer recognizes the utterance during training for building the sentence. VT system applies the Baum-Welch estimation algorithm, to emulate the statistics of the training database[11]. This minimizes the error rate. 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 character. Average values and standard deviations are used to smooth the phonics and help to get the perfect text.

Romany, Arabic and English numbers, are dictated in character form, to be defined well in databases to be retrieved as numbers in the other language. The integrity of generated text depends on the microphone quality, Specifications for SR Grammar (SRG), Semantic Interpretation (SI), Speech Synthesis Markup Language (SSML), and Specifications of Pronunciation Lexicon (PL). The basic emotions such as: scream, surprise, happiness, anger, fear, disgust, and sadness; could not expressed with the same manner in the other language. There are other elements like 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, about 9% was angry talk, 5% sad talk, and 8% happy talk. These could not be translated in other language. Many Arabic dictation systems like Viavoice , secretary, and Arabic-dictation-pro are efficient in translating VT. The more trained dataset can result in a proper text. The training system concerns on the data reference, tools for decisions, and dialects. Conversations vary during: speech, vowels, syllables, and utterances. 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 at least double of frequency. Human speech is less than 4KHz, so need about 8K 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). Discrete Fourier Transform (DFT), Hidden Markov Models (HMM), Isolated Word Recognition (IWR), are applied to analyze and recognize the utters to produce the proper words. The recognizers find the optimal start/stop of utterance with respect to the acoustic model. The system can recognize the sequences of words or non-speech events. Pattern detection is accomplished by parsing the phonic sequence and generate the relevant character sequence. This requires all possible combinations of symbols generated from the speech patterns. Several algorithms like Viterbi, Baum-Welch, hybrid schemes and Artificial Neural Networks (ANN) are applied for the supervised learning process[12]. ANN network simulates the input and estimates the weights to adjust automatically and decides the best corresponding output. The VT 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 affected by 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 TT translating problems.

3.2. DESIGN OF TEXT - TEXT INTERFACE SYSTEM (TTIS)

Dictionaries map Arabic and English words. AI tools are necessary to search for acronyms, regarding sentence situations. The syntactic and semantic 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. But they could not generated perfectly by VT system. 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 define verbal and noun clause, adjectives, adverbs, pronouns, propositions, conjunctive, interrogative expressions, direct and indirect speech, joint and complex sentences, vocabularies, the five nouns, masculine plural, feminine plural, irregular plural, verb tenses, the appositions, exceptions, and many grammatical rules that 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 start 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 ef''al. Gerunds may be three, four, five, or six letters. 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, but Arabic define them. 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-qamar or ash shams). 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 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 word in the sentence. Conditional statement in Arabic has many articles with many rules, but in English, there is only if- statements. Exclamation verbs and query statements in Arabic have their own articles and rules, while in English just use 'how good or bad'. Swearing nouns are not exist in English. Negation and Prohibition verbs have many situations in Arabic, while 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 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 sampled within the next structures:

<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> <predicate>

<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> & <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 read/ written from right to left. The parser proceeds 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 and synthesizer are applied to construct the destination sentence. Then sentence structure is scanned by semantic analyzers. These steps are shown in fig 8.

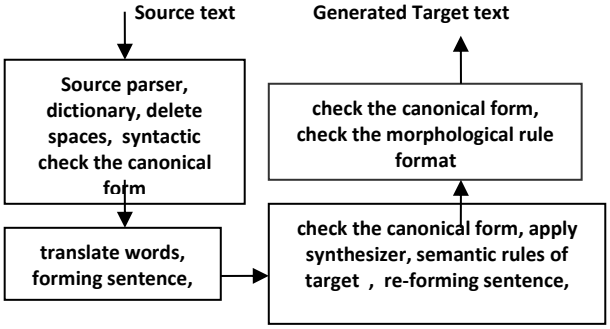

Fig 8. Components of the Proposed TTT System

The Arabic To English To Arabic (A-E-A) translator must analyze the source text, translate the source words, check the canonical form of sentences, and generate the destination text. The acronyms are selected and syntactic sentence is generated using Structured Language. 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 translator to produce the perfect sentences. Many sentences must be reformed when being translated semantically. Declarative, interrogative, imperative, and exclamation sentences have variable structures in Arabic and English. Many English words have no relevant in Arabic like ummm, and other filling words. Roman numbers must be read as numbers not letters, e.g. IV is read as four not I, V. So, when translated by conventional software programs, they give the correct means. Therefore, cognitive training are applied to these software to produce the proper translations. Many slang Arab words have no relevant in English. Some upper and lower signs are used in Arabic to adjust the meaning of sentence have no relevant in English . Some signs are used to extend the pronunciation of some characters, these vowels and consonants change the meaning of sentences. Some Arabic word may be translated into a sentence in English, e.g. "anolzemkmoha" that is translated into "shall we enforce you to do it", "asqainakmoh" which means "we made you drink it", and many other words must be clustered syntactically before translation. Other English words must be translated into many Arabic words, like "internetworking" which means "dealing between networks". These attributes change the morphology , and phonology of Arabic and English words. Arabic language has 29 letters with 92 different sounds, while English has 22 letters with 42 different sounds[13,17]. Translator dictionaries must use intelligent oriented search tools. Selecting the acronyms, synonyms, and vocabularies necessitates one to many word-pair assignment, to generate a well trained sentence[17]. The proposed dictionary use look ahead tools, and semantic synthesizer to produce an efficient TT translation. The translated Arabic by this system gives a better quality than the translation by Google by 6%. This test was done for 250 sentence from different subjects.

3.3. Text to Speech (TS) Systems,

Evolution of TS quality has three generations: character synthesizers, formant synthesizers and concatenative synthesizers. Synthesizing emotions, impressions, and temper; gives more powerful value for TS. 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 fig 9. The pronunciation rules include information about the contextual speak, and acoustic frequencies. Natural reader and Oddcast are efficient software that transfer TS.

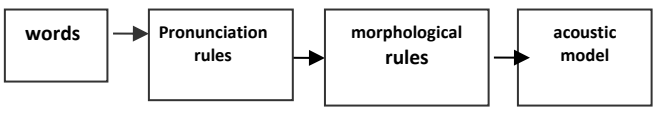

Fig 9. A Pronunciation Model

Fig10 indicates a scenario for generating speech. This necessitates trained DBs 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.

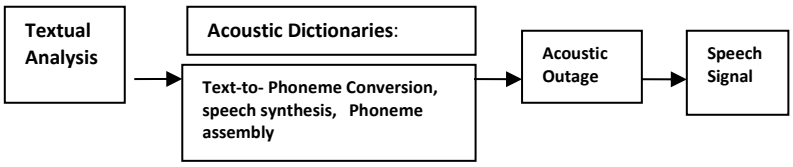

Fig10. Model for Generating Speech (TS)

Enhancing TS system is accomplished by voice morphing [6]. Speaker verification is done via a biometric measure, where every person has a unique voiceprint [7]. 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 morphologic 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 include the envelope information, dynamic warping, and signal re-estimation. Speech morphing can be achieved by transforming the signal's 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. To do this match, the signals are 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 phone likelihood tests. Generation of utterance need no lexical interference, but needs bigger DB and knowledge base (KB). While using natural words is easier to pronounce and needs smaller KB. After generating DBs , it must make recordings consistent, this is done by using constant pitch (monotone). It should avoid pronunciation problems, and keeps speaker consistent. Text must be normalized in both languages, numbers and abbreviations must be defined in dataset. Intelligent rule based system is used for proper translation. The more DBs lead to proper selection for the relevant words, this may increase uncertainty and ambiguity to select the proper words. Large DBs lead to more delay time to search for the proper translation of some word. Binary search tools are applied to shorten the delay time for searching. Synthesizing software use polyphone, and concatenating synthesizers to produce the speech word. The sentence is compounded during the translation phase using rule based system.

4. RESULTS and CONCLUSION

The analyzer and synthesizer concerned in dealing punctuation and spaces of natural language. They deal affixes, lexical ,semantics and syntactics, of the different sentences. Many automated measures have been proposed to test the speed and accuracy of the translated-language process. These evaluation tools must be fast and cheap[18,19]. 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 (semantic of a sentence). These metrics are applied for EARS system, for 1000 Arabic sentences, with average 5.1 words per sentence and standard deviation= 3.2 words. A comparison between Google translation and EARS translation is shown in fig 11.

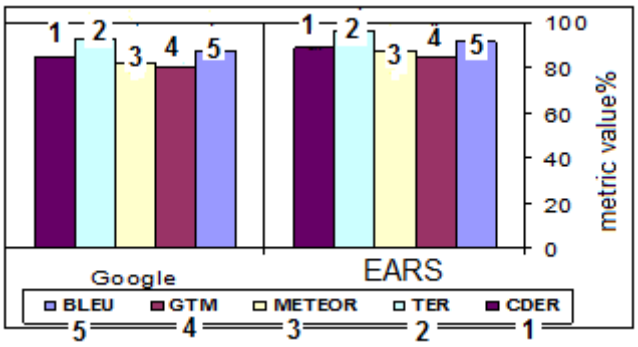

Fig.11. A comparison between the intelligent EARS and Google translation system

The EARS fulfill a fluent words translation using databases English and Arabic dictionary. These datasets are drawn from NIST (Free Online Databases), and internet documents. 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 EARS 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 pre-processing was applied to all datasets. These pre-processing include different punctuations with different degrees. It is observed that the translated part of EARS results in a better score than

Google. These results are statistically significant and confident. EARS sustains a delay between spoken and heard translation less than 1.4sec.. This response delay is the sum of delays due to ST, reading text, TTT, and TS. This cumulative delay is due to Dictation software, translation and reading. Analyzing the EARS performance, indicate that more than 96% of the translated tasks are acceptable. EARS introduces some features in typical Arabic style, that are not captured by other MT (Machine Translators) 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 is composed of verb, subject, object, and interrogation expression, and it means " can we enforce you to do it?". So, Arabic language is a powerful language as shown in fig12.

The word "Anolzemkomoha? "

The relevant statement "can we enforce you to do it ?"

Fig.12-ST –TS (Arabic and English Sampled statement

ANN and Support Vector Machines (SVM) are applied for machine learning to enhance the TTT product. Fuzzy rough neural is a mixed technique that is used to reduce the ambiguous words[20]. Comparing the results of EARS system and Google translation, it is observed that Goggle results in the lower score, EARS shows higher metrics values between 2% to 9 % over Google. This implies that EARS 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 for EARS showed that it is more formed, and structured than Google dataset. Evaluation of speech system considers: quality of microphones, utters, percentage of formality, efficiency of dataset dictionary, vocabularies, efficiency of speech synthesizer, elapsed time for transformation, percentage of ambiguous words, quality of ASR, percentage of inappropriateness, ability to understand, and ability to sustain user satisfaction. Most systems use regression to train weights to get the perfect outage. EARS 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 TS. Matlab is used for analyzing and synthesizing signals and constructing DB for TS, TTT translation demo translates Arabic to English and vice versa. Systems like systransnet, Natlang are used to translate A-E. 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 TTT, and turned into speech using TS using readers or talkme software. Microphones and standard references have a main factors in quality of generating the text. EARS 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 EARS 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 and political conferences. 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 of EARS (ST, TTT and TS), to get the present situation. It still need many datasets to extend the database to all domains in both languages. It needs fund to make it industrial. The team-work used language experts, computer programmers, computer engineers for speech analysis and synthesizing. EARS may be splitted into ST ,TTT, TS and SR systems and could be attached to telephone systems. Where each of caller and called persons, can speak and hear using his mother language. Some errors still un-dealt like: hesitations, mouth noises, Laughter, Unintelligible, Spoonerism, pauses, vocalized noise, and coinage.

REFERENCES

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

2. Ellis Mandel and Wendy Holmes, "Speech Synthesis and Recognition", CRC, April 2009. ISBN 0748408576.

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

3. Ellen Eide and alii Narayanan, voice morphing , CRC, April 2009.

4. Ostendorf Bulyko, Cross-fertilization between ASR and TTS areas, Prentice Hall, 2010

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

6. D. Jurafsky, J.H. Martin, “Speech and Language Processing”; Prentice Hall; 2nd edition (2006).

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

8. Jim Larson, “VoiceXML Introduction to Developing Speech Applications”, Prentice-Hall, 2008.

9. Stuart Russell, Natural Language Processing and Applications", AI Journal, v2, #5, 2007

10. Douglas A., "machine Translation", 2007, ISBN 1855542-17x.

11. Schlesinger Hlaváč, statistical and structural pattern recognition. Journal of Pattern Recognition Society, v1, 2007.

12. F. AlAnzi, "Automatic English/ Arabic Translation Tool", Journal of KSU, 2005

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

14. Ahmed Altanni, " A Direct English –Arabic Machine Translation System", IT Journal, 4 (3), 2006

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

16. Alansary, S., Nagi, M., Adly, N.: 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. (2006)

17. Agrawal, A., Lavie, A.: 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, June (2008)

18. Banerjee, S., Lavie, A.: 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, (2015)

19-H. Shi and Y. Liu, Naïve Bayes versus Support Vector Machine, Resillience to Missing Data, Proceedings on International Conference of Artificial Intelligence, Springer, pp 680-687, 2011.