1. INTRODUCTION
This study is a modeling of the Egyptian Arabic phonetics. It is based on gradient constraints of the verbal root concerning sequences. OCP-place (the obligatory contour principle of place of articulation) approach gave a large description of the Arabic triliteral root and the similarity constraints of its consonants. Since the Egyptian Arabic (EA) Dialect has been considered as the most widely understood dialect throughout the Arabic world, creating a model for spoken word segmentation of EA became an essential task. The main goals are (i) to formalize a model for the human phonetic constraints detecting based on probabilistic measurements of the verbal root consonants, (ii) to improve a segmentation performance and (iii) to evaluate the ability of the model to detect the morphological tokens of a speech by evaluating the phonetic constraints. The goals have been achieved by designing a phonetic model measuring the maximum likelihood for the verbal root consonants occurrence based on the Hidden Markov Model states, and use the estimated values to detect the tokens boundaries. Although the model is still in progress its performance produced a high accuracy for the required assignments, and gave a high coverage for the morphological forms based on the phonetic constraints.

2. ARABIC NON-CONGATATIVE STRUCTURE
Arabic language is a member of the Semitic family language which is distinguished by its root-pattern system, where verbal root consists of a set of three or four consonants. Three approaches describe the Arabic word structure:

In Templatic approach (McCarthy, 1990) the Arabic word is represented in separate levels which are called Autotopes:

- The root tier consists the consonants that constitute the root
- The pattern (skelatal tiers) the word in term of consonants C and the vowels V which are inserted together in a certain patterns or templates
- The vocalic (vocalic melody tiers) represents the short vowels that specify each template or pattern.

For example, in the root (k, b, t) it has among its word forms: ‘kabbara’ to write, ‘kabbabi’ to be written and ‘kabbaba’ ‘book’

Affixational approach:
McCarthy and Prince (1995) claimed that templates is not the main issue in analyzing Arabic but measures are derived by affixes usage, where they have called the "prosodic circumflexion" Prosodic circumflexion defines the domain of the morphological operation this domain is the grammatical category resulting by affixation.

Description of Egyptian Arabic Phonology:
Egyptian Arabic (EGY) as MSA has a phonological profile includes 28 consonants, three short vowels, and 3 long vowels in addition to two diphthongs: /aː/ and /oː/.

Egyptian Arabic followed the same phonological system of MSA but differs in the following:
- MSA v(aw) is realized in EGY as /æ/ or /æ:/
- MSA allophonic affiliate (a) is realized in EGY as /v/ or /v/
- MSA coronal Obruants (b, t, and k) are realized as EGY as /b, t, and k/ respectively
- MSA diphthongs /aː/ and /oː/ have mostly become /a/ and /o/ respectively

Syllabification:
The phonological word in the EA consists of a stem and affixes, such as definite articles, consonants of subject and object pronouns.

Where morphemes cannot be syllabified according to the syllabification algorithm one of the following recurring processes occurs:

Euphonic:
When 3 consonants are juxtaposed within the utterance, an effusion of [j] or [w] occur between the second and the sound

[αι]ta ta ... [αι]ta

Consonant prothesis:
Although each syllable in EA requires an Onset, and the majority of morphemes which may occur in utterance-initial position have an underlying initial consonant (Watson, 2007), there are some vowel-initial morphemes such as the derivates article /il/ that doesn’t have to be realized through processes of a glottal stop which is replaced in the continuous speech by the cola of the previous syllable.

Closed syllable shortening:
When a consonantal initial syllable CVVC is consonantated with a consonant-initial morpheme, the long vowel is shortened (Watson, 2007)

3. PHONOTACTIC MODELING OF ARABIC LANGUAGE
Arabic root is a perfect example of OCP (Obligatory Contour Principle of Place of Articulation), while Penhuent (1997) found that there is a relation between the consonant pairs within the root and the acceptability of the verbal root, this relation is more obvious depends on the homorganic similarity of these consonants, a highly similar homorganic consonants are less frequent than dissimilar homorganic consonants pairs.

OCP-place constraints are filters which prohibits a root repeated homorganic consonants, but these constraints are non-categorical but a gradient constraints which have fuzzy constraint boundaries.

OCP-place model (Penhuent, 1997) is a stochastic constraint model that can be parameterized to account for gradient constraints within successive pairs of the Arabic verbal roots using the OCP ratio. The OCP ratio is the ratio of the observed consonant pairs to the number of expected consonants to occur by chance (Penhuent, 1999).

Where X can be any segment following Y, X and Y can be any segment preceding Y.
The ratio of less than 1 indicates underrepresentation of the diacrit as seen in fig. (1)

4. METHODOLOGY

4.1 Phonotactic modeling
Concept: Triliteral roots of 915 Egyptian words have been extracted from ARZ-ATB Arabic Textbook corpus, the corpus presents Egyptian written conversations

Transcription: Coda/Ae Convention (Oriental for Digraphic) is used for transcription, CODA describes the Egyptian Arabic phoneme, it doesn’t concern with the phonetic variations between speakers.

Hidden Markov Model (HMM): It is a statistical tool for modeling generative sequences that can be characterized by an underlying process generating an observable sequence (Blunsom, 2004)

The Maximum Likelihood Estimate (MLE) was calculated for the most probable hidden state sequence of the adjacent root consonants: C1C2C3 by the matrix:

\[
P(C1/C2/C3) = \frac{1}{n} \sum_{i=1}^{n} \delta(i)
\]

5. RESULTS
Three different thresholds have been evaluated and the best results have been shown in the threshold 0.03.

The model performed good recall of precision, recall, and F-score of some of the elements, these elements are (true positives, false positives, and false negatives).

The phonetics of the root consonant the observed word stem and its prefixes with an acceptable percentage.

The highest accuracy was for the data set three due to the observed ability to detect the prefixes.

The model covered 92.22% of the data set two and 44.74% of the suffixes. (fig. 3)

6. REFERENCES

Adorno, T., Dopp, W., 1941. The philosophical foundations of the German Democratic Republic: A Collection of Original Sources.


5. CONCLUSION AND DISCUSSION
The study shows that the similarity constraints of the Arabic verbal root have significant consequences for phonotactic modeling and speech segmentation.

The phonological segmentation based phonotactics was accomplished by the phonotactics modeling.

The model succeeded to cover the majority of its required tasks, it needs to be modified to cover extra tasks and to manage words with successive affixes, and words that are not based on verbal root.

Future work
The experimenter is working on some shortening rules for more accurate phonetic estimation

6. ACKNOWLEDGEMENTS
The Department of Arabic.

6. REFERENCES

Adorno, T., Dopp, W., 1941. The philosophical foundations of the German Democratic Republic: A Collection of Original Sources.