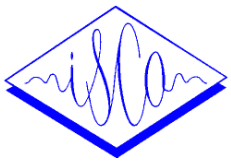


First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

DISCRIMINATION OF WORDS IN A LARGE VOCABULARY SPEECH RECOGNITION SYSTEM

S. Datta* & M. Al-Zabibi**

*Dept. of Electronic & Electrical Engineering
Loughborough University of Technology, Leics, LE11 3TU, U.K
**Scientific Studies & Research Centre, Damascus P.O. Box 4470, Syria

ABSTRACT

Word discrimination according to broad phonetic classes is presented in this paper. Different Phonetic classification strategies are used to describe large vocabulary lexicons. The phonetic description in these strategies varies from 2 to 17 phonetic classes. The statistical results show that about 83% of the 10,000 test Arabic words can be uniquely represented by using 7 broad phonetic classes for consonants and six classes for vowels. In this case, the maximum number of words having the same phonetic labelling is 6. The paper summarises the results of ten different phonetic classification schemes and discusses their implication for a large vocabulary speech recognition system. Distributions of vowels and consonant classes are also presented.

I. INTRODUCTION

Isolated word speech recognition can be achieved with very good accuracy for small vocabularies by using pattern recognition approaches [1], or stochastic modelling using hidden Markov models [2]. An unknown word is matched against all vocabulary reference patterns or models (using all word templates or models). These mathematical approaches utilise little or no speech-specific knowledge. The performance of such systems would surely deteriorate for large vocabulary systems, and also for acoustically similar words .

For large vocabulary systems (including continuous speech recognition systems) the acoustic-phonetic approach has been used, where recognition is achieved by mapping the acoustic signal into a sequence of linguistic units, such as phonemes, diphones, demisyllables or syllables. Words in the lexicon are represented by a concatenation of the chosen linguistic units according to their considered pronunciation. The major problem with this method is our inability to extract phonetic information reliably from the speech signal, because of the variability in the acoustic realisation of utterances. The variability comes from diverse sources such as the talking environment, speaking rates, and differences across speakers.

Broad phonetic classification techniques can be used in large vocabulary systems, instead of detailed acoustic-phonetic analysis, to overcome the above mentioned variability. In these techniques a coarse reliable acoustic analysis in terms of broad phonetic classes (**BPCs**) is performed. In such systems, lexical access is performed in a bottom-up phase on the basis of broad phonetic information extracted from the test word. The lexicon is structured into sets of words sharing the same broad phonetic labelling called 'cohort'. The most likely word-candidate which matches the test word is chosen by a top-down phase (verification). In this phase, the constraints imposed by the phonemic structure of the chosen set of words determines the most appropriate verification analysis.

In order to reduce the number of words to be verified in the top-down phase, a suitable broad phonetic representation should be used. This representation should be broad enough to maintain minimum detailed acoustic-phonetic analysis, and narrow enough for effective lexical access.

Shipman and Zue [3] have used six broad phonetic classes (i.e., vowel, plosive, weak fricative, strong fricative, glide and semi-vowel). They reported that the maximum cohort size corresponds to about 1% of the size of the lexicon (using 20,000 words of American English) and that almost one third of the words are uniquely represented at this broad phonetic level.

Applying this approach to Arabic words in a large vocabulary Arabic speech recognition system, taking into consideration some features of the Arabic language, has led to powerful lexical access, as will be explained in the following sections. The Arabic language has an unlimited vocabulary size because of its derivative property. Two lexicons have been considered in this study. The first one comprises the most frequent 3,000 words in the language [4]. The second lexicon contains 10,000 words, which includes words of the first lexicon along with the derivatives of some of them, and other randomly chosen words. The phonetic descriptions of these words have been obtained from the orthographic form by means of a set of translation rules according to the standard Arabic pronunciation (which is used throughout the Arab world).

In this paper, a brief description of the Arabic phonetic system, the syllabic types, and the morphological structures is given in section II. Ten different phonetic classification schemes are presented in section III. Section IV reports the results of word discrimination according to the proposed phonetic classification schemes. Discussion and conclusion are presented in sections V and VI.

II. ARABIC PHONETIC SYSTEM AND MORPHOLOGY

Standard Arabic language has basically 35 phonemes, of which there are 29 consonants and six vowels [5] . The vowels consist of two groups, namely three short /a/, /i/, /u/, and three long /aa/, /ii/, /uu/. The short vowels are written as diacritic marks below or above a consonant, while the long vowels are written as separate letters. The consonants are described in Table (1), which shows a tentative chart of the standard Arabic consonantal system [6, 7].

In this table, consonants are categorised according to their place of articulation, manner of articulation, voiced or unvoiced, and pharyngealised or non-pharyngealised. The Arabic phonetic system differs from the Latin one primarily by the presence of pharyngealised (emphatic), uvular, pharyngeal, and glottal phonemes. The pharyngealised consonants in the table are underlined (i.e., /d̲/, /t̲/, /s̲/, /k̲/,/ð̲/,

and /l̲/) to distinguish them from their counterparts, the non-pharyngealised (plain) consonants (i.e., /d/, /t/, /s/, /k/, /ð/, and /l/). In general phonetic terms, pharyngealisation or emphasis has been described as a rearward movement of the back of the tongue towards the back wall of the pharynx. The result of this movement is a vocal tract shape with an increased oral cavity (between the surface of the tongue and hard palate), and a reduced pharyngeal cavity above the epiglottis compared to non-pharyngealised counterparts. The consonant /l̲/ is used only in one word '?allaah' (God) (according to standard Arabic). By avoiding this word, we can confine the pharyngealised consonants to the five consonants (/d̲/, /t̲/, /s̲/, /k̲/, and /ð̲/).

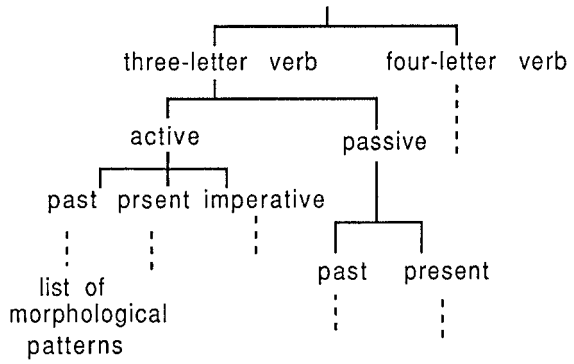
			Bilabial	Labiodental	Interdental	Alveodental	Alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	v	ph				d̲						
			b			d						
	uv	ph				t̲				k̲		
						t			k			ʔ
Fricative	v	ph			ð̲							
					ð	z	ʒ		ɣ	ɛ		
	uv	ph				s̲						
				f	θ	s	ʃ	x			ħ	h
Nasal	v		m			n						
Liquid	v	ph				l̲						
						l,r						
Semivowel	v		w				j					

Table (1) The Arabic Phonetic System (consonants)
('v': voiced, 'uv': unvoiced , 'ph': pharyngealised)

The Arabic language is characterised by well-defined syllabic types and structures. It uses three main syllabic types, namely /CV/, /CVC/, and /CVCC/ [7]. The third type /CVCC/ is less frequent than others, and occurs only at the word-final position or in isolation (monosyllabic words). This means that a consonant cluster in any word contains two consonants at most.

The Arabic language, like all other semitic languages, has a very systematic morphological structure compared with Latin languages. There exist strict morphological rules which control the vocabulary structure. Arabic words are morphologically derived from a shorter list of generative roots. Arabic has 11347 roots [8], composed of two, three, four and five consonants (letters). The three-letter roots (triradical) represent 63% of all roots, and are more frequent than other roots.

Words are classified into three main categories, namely verb, noun and tool (tools are pronouns, preposition, affixes and others). Each category (except tool) has its own sub-categories. For example the sub-categories of ' verbs' are:



At each final branch we have a list of morphological patterns which define the syllabic structure and the actual vowels used in the word. For instance, the word "kataba" (he wrote) has a pattern /CaCaCa/. In this pattern /C/ could be any consonant out of the 29 consonants, but the sequence of the three consonants is subject to phonological rules. In general, Arabic words (except non-inflectional words such as pronouns, prefixes, and suffixes) are already categorised according to their morphological patterns. These patterns lead to grammatical, syntactic, and semantic information about words. Such information is very useful in large vocabulary speech recognition systems.

III. THE PROPOSED CLASSIFICATION SCHEMES

Ten different classification schemes are considered in this study. These schemes can be divided into two groups. In the first group, vowels are replaced with the symbol /V/, while in the second group, vowels retain their phonetic symbols (i.e., /a/,/u/,/i/, /aa/, /uu/, and /ii/). These schemes are as follows:

In the first group, vowels are replaced by the symbol /V/, and consonants are classified as follows:

1) **C/V**, Consonants are replaced by the symbol /C/.
2) **4BPC/V**, Consonants are classified according to four BPCs: voiced plosive, unvoiced plosive, unvoiced fricative, and other voiced consonants.
3) **5BPC/V**, Consonants are classified according to five BPCs: plosive, fricative, nasal, liquid, and semivowel.
4) **7BPC/V**, Consonants are classified according to seven BPCs: voiced plosive, unvoiced plosive, voiced fricative, unvoiced fricative, nasal, liquid, and semivowel.
5) **11BPC/V**, Consonants are classified as in the 7BPC/V scheme, but the plosive and fricative classes are divided into pharyngealised and non-pharyngealised classes yielding 11 BPCs (see Table (1)).

In the second group, vowels are classified according to their phonemic form, and consonants are classified as in the first group, giving the following schemes:

6) **C/6V** 7) **4BPC/6V** 8) **5BPC/6V**
8) **7BPC/6V** 10) **11BPC/6V**

Because the morphological pattern of a word gives the syllabic structure and the actual vowels of that word, the classification scheme C/6V represents a classification according to morphological patterns.

IV. Statistical Results

The distributions of vowels and consonant classes in the two lexicons are given in Table (2). In the 10,000-word lexicon, the vowels represent about 43% of the total number of phonemes (75875 phonemes), while the consonants represent about 57%.

		10,000 words	3,000 words
Plosives	v	0.66 %	0.76 %
	v -ph	4.80	5.77
	uv	3.16	3.54
	uv-ph	10.41	8.99
Fricatives	v	0.35	0.39
	v -ph	5.04	5.95
	uv	0.92	1.16
	uv-ph	9.57	11.85
Nasals	v	10.84	9.24
Liguids	v	7.43	8.74
Semivowels	v	3.89	3.13
/a/	v	20.32	21.65
/u/		6.25	4.14
/i/		7.63	5.34
/aa/		5.82	6.76
/uu/		1.39	0.93
/ii/		1.51	1.66

Table (2) Distribution of vowels and consonant classes for the two lexicons

The results of using the ten classification schemes are summarised in tables (3) and (4) for the 3,000-word and 10,000-word lexicons respectively.

Table (3) shows that on one hand, the number of unique word cohorts (i.e., cohorts having just one word) increases with the number of BPCs, while on the other hand, the maximum cohort size (maximum number of words in a cohort) decreases as the number of BPCs increases. For the classification scheme C/V, the 3,000 words are grouped in just 31 cohorts, while for the C/6V scheme they are grouped in 286 cohorts (morphological patterns). The percentage of uniquely represented words rises from about 55% for the 11BPC/V scheme to about 82% for the 11BPC/6V scheme. In the latter case, the maximum cohort size is 5 and the average cohort size is just 1.11.

The classification results of the second lexicon (10,000 words) given in Table (4), are quite similar to those of the first lexicon. However, in this table, there is a rise in the percentage of unique word cohorts for all the classification schemes compared to that of Table (3). This is mainly due to the increase in the number of polysyllabic words in the second lexicon, where the number of syllables varies from 1 to 5 for the first set and 1 to 7 for the second.

	Vowel					6 Vowels				
	C	4BPC	5BPC	7BPC	11BPC	C	4BPC	5BPC	7BPC	11BPC
no. of cohorts	31	868	1079	1835	2156	286	1762	1972	2558	2714
no. of unique word cohorts	3	420	527	1244	1651	134	1210	1432	2226	2479
maximum cohort size	599	66	39	19	15	174	16	13	6	5
average cohort size	96.77	3.46	2.78	1.63	1.39	10.49	1.70	1.52	1.17	1.11
% of uniquely represented words	0.1	14.	17.56	41.46	55.03	4.46	40.33	47.73	74.20	82.63

Table (3) Classification results for the 3,000-word lexicon

	Vowel					6 Vowels				
	C	4BPC	5BPC	7BPC	11BPC	C	4BPC	5BPC	7BPC	11BPC
no. of cohorts	72	2981	3722	5810	6654	1437	6922	7579	9043	9384
no. of unique word cohorts	9	1518	2048	3862	4785	683	5365	6180	8317	8876
maximum cohort size	1022	89	54	28	22	197	17	15	6	5
average cohort size	138.8	3.35	2.69	1.72	1.50	6.96	1.44	1.32	1.11	1.07
% of uniquely represented words	0.09	15.18	20.48	38.62	47.85	6.83	53.65	61.80	83.17	88.76

Table (4) Classification results for the 10,000-word lexicon

The percentage of uniquely represented words has also risen here from about 48% when using the 11BPC/V scheme, to about 89% when employing the 11BPC/6V scheme. Even for a simple classification scheme such as the 4BPC/6V, the percentage of uniquely represented words (about 54%) is high, the maximum cohort size is 17 words, and the average cohort size is 1.44 word.

In general, the detailed vowel classification has almost doubled the number of uniquely represented words (e.g., from 38% for the 7BPC/V scheme to 83% for the 7BPC/6V scheme), and has led also to specifying the morphological pattern of a word.

V. DISCUSSION

A model for a large vocabulary isolated word speech recognition system is given in Figure (1). In this model the speech signal is first transformed into acoustic parameters through the feature measurement stage. The parameter complexity depends on the employed set of BPCs in the broad phonetic classification stage. These parameters are used in the vowel recognition stage and in the broad phonetic classification stage. The output of the broad phonetic classification stage is a string of phonetic labels which is used for lexical access (bottom-up phase). The result of the lexical access is a small set of word candidates (or more likely a single word candidate), sharing the same phonetic labelling. In this model, detailed acoustic knowledge is applied in a top-down verification mode (just when it is needed), to select the most likely word candidate.

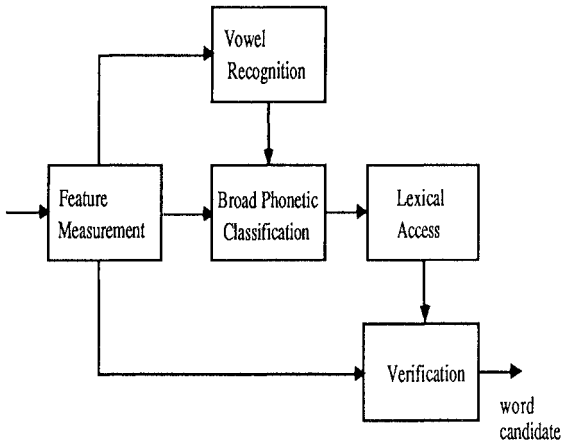


Fig. (1) Speech recognition model

A recognition system based on a hierarchical classification can start with detailed vowel recognition and a simple set of broad phonetic classes. The resultant sequence of labels is used for lexical access. If the number of word candidates is more than one, it goes back to the classification stage and widens the set of BPCs until it reaches the minimum possible number of word candidates. The verification stage is then activated if the number of word candidates exceeds one.

VI. CONCLUSION

The use of some phonetic classification schemes leads to drastic cuts in the number of word-candidates at the lexical level for a specific pattern. Using broad phonetic classification for consonants and detailed vowel classification has led to a powerful lexical access for the Arabic language. It has also given at the same time some information about the morphological pattern of a word, which is very important for higher level sources of knowledge, especially in continuous speech recognition or speech understanding systems. The results show that phonological constraints imposed by the language have important implications in speech recognition. They suggest that a complete and detailed phonetic analysis of the speech signal may be unnecessary.

We have seen that the maximum cohort size is 5 words for the 11BPC/6V scheme (using a lexicon of 10,000 words). Prosodic information such as stress position, duration of different phonetic segments could also serve as cues for reducing the cohort size.

ACKNOWLEDGMENT

The authors are very grateful to Dr. M. Mrayati, head of the research group working on Arabic computational linguistics and Arabic speech processing at SSRC Damascus, Syria, where this research is being done, and to Mr. M. Bawwab for his comments on the orthographic-to-phonetic conversion procedure.

REFERENCES

[1] L. R. Rabiner, and S. E. Levinson, "Isolated and Connected Word Recognition, Theory and Selected Application", IEEE Trans. on Communications, vol. COM-29, pp. 621-659, May 1981.

[2] L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On The Application of Vector Quantisation and Hidden Markov Model to Speaker-Independent Isolated Word Recognition", Bell Sys. Tech. J., vol. 62, pp. 1075-1105, April 1983.

[3] D. W. Shipman, and V. Zue, "Properties of Large Lexicons: Implications for Advanced Isolated Word Recognition System", IEEE Proc. ICASSP-82, pp 546-549, 1982.

[4] D. A. Abduh ,"The Common Words in the Arabic Language," Publication of Riyadh University, Suadi Arabia, 1979 (in Arabic).

[5] S. H. Alani, Arabic Phonology: "An Acoustic and Physiological Investigation," Ph.D. Thesis, Indiana University, USA, 1970 .

[6] M. Mrayati, "Speech Processing Application to the Arabic language," Workshop on Computer Processing of the Arabic Language, Kuwait, April 1985.

[7] A. S. Shaheen "Phonetic Method for Arabic Structure", Alrisalah Establishment, 1985 (in Arabic).

[8] H. Tayyan, Y. Meer Alam, and M. Mrayati, "Data Base for Arabic Roots," Second Conf. on Arabic Computational Linguistics, Kuwait, Nov. 1989 (in Arabic).