6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

PROSODY AS A DISTINCTIVE FEATURE FOR THE DISCRIMINATION OF ARABIC DIALECTS

Melissa BARKAT
Laboratoire DDL-ISH
14, avenue Berthelot
69363 Lyon Cedex 07
France
Melissa.Barkat@ish-cnrs.fr

John OHALA
Phonology Laboratory
1203, Dwinell Hall
University of Berkeley,
94720-2650 California
ohala@cogsci.Berkeley.edu

François PELLEGRINO
Laboratoire DDL-ISH
14, avenue Berthelot
69363 Lyon Cedex 07
France
Francois.Pellegrino@ish-cnrs.fr

ABSTRACT

The aim of the work to be reported here is to explore the utility of prosodic information in language identification and discrimination tasks. The purpose of this study is to see whether prosodic patterns can be considered as reliable acoustic cues for the discrimination of Arabic dialects by investigating, via a perceptual experiment, if listeners are successful in identifying the Arabic dialect used by a speaker when they only have access to fundamental frequency, amplitude and some rhythmic characteristics of the original voice signal. Results show that prosodic cues alone can distinguish between dialect pairs, since native Arabic listeners are significantly more successful in identifying the Arabic dialectal varieties both in their natural and synthesized forms and that listeners' identification rate are higher for the discrimination of their own dialectal variety when presented under its processed form. This perceptual study must be regarded as a first step towards the determination of a set of reliable cues for the Automatic Identification of Arabic Dialects.

Keywords : Language Identification, Prosody, Perceptual experiment, Arabic dialects

1. INTRODUCTION

It is a common observation that languages "sound" different not only because they use different inventories of segments but also because of their typical prosodic configurations. The study of prosodic patterns could then have a considerable impact in many practical areas including speech synthesis and Automatic Language Identification (A.L.I.). Unfortunately very few studies have considered the role of prosody as a reliable cue for language and dialect discrimination [1] ; [2] ; [3] and [4] none of them dealing with the Arabic language. Nevertheless, impressionistic evidence suggest that Eastern and Western dialectal varieties of Arabic can be differentiated by their prosodic patterns [5].

Prosodic parameters include stress, rhythm and intonation. Each cue is a complex, language dependant perceptual entity expressed primarily as the combination of three acoustic cues : pitch (i.e. fundamental frequency or Fo), amplitude (i.e. energy or intensity) and duration. As far as Arabic language is concerned, recent experimental studies have determined the importance of Fo and amplitude variations to implement prosodic information over the sentence [6] [7]. These studies tend to show that each dialectal variety develop a peculiar prosodic pattern by enhancing one particular acoustic parameter. If so, Western Arabic (corresponding to the dialectal varieties spoken in the Maghreb, i.e. Northern Africa) should probably be perceived as a distinct 'language' from Eastern Arabic, as spoken in the Middle-East.

In this study, we assume, following, among others, Di Cristo [8], that intonation is encoded as a sequence of key-points distributed throughout the 'Intonation Unit' and that an intonation contour is perceived as an interpolation between these points as suggested by Thorsen [9] : "we anchor our perception of intonational phenomena on certain points in the time varying course of pitch and disregard what lies between such fixed points".

Since prosody is a supra-syllabic phenomenon, it is not necessary to specify the pitch-point for each syllable, but we can assume that the key-points will vary from one language to another and furthermore, since Arabic dialects exhibit drastically different stress patterns, from one areal dialectal variety to the other. Indeed, the loss of short vowels in Western Arabic and its consequences on the prosodic level constitute a major difference between Western and Eastern Arabic dialects [12] ; [13] which, we assume, could be by itself perceptually significant for the identification of Arabic speakers' dialectal origin (i.e Western vs Eastern).

The purpose of the study reported here is thus, to see whether prosody can be considered as a reliable acoustic cue for the discrimination of Arabic dialects by investigating, via a perceptual experiment, if listeners are successful in identifying the dialect used by a speaker when they only have access to the fundamental frequency, the amplitude and some rhythmic characteristics of the original voice signal.

2. MATERIAL AND METHODS

Two perceptual experiments were performed : a baseline experiment based on natural speech was meant to evaluate the subjects' knowledge and perception of dialectal areal characteristics. A second experiment based on speech synthesis was meant to evaluate the reliability of prosodic information for the identification of Arabic dialects in terms of zone. Recordings of unrehearsed but

elicited story-telling[1] were obtained from four adult male native speakers of Arabic coming from the two major dialectal areas of the Arab World and from four different countries (i.e. Morocco and Algeria, accounting for the Western zone ; Syria and Jordan representing the Eastern one). For each speakers, we selected six samples of speech, yielding a total of twenty-four passages that were to be presented twice to the subjects so as to evaluate their answers' coherence. The stimuli were presented to a group of 38 adult listeners divided into two populations. The first subject-population was composed of nineteen people whom knowledge of Arabic is limited or nil (i.e. no *a priori* consciousness of Arabic dialectal varieties), the second one of nineteen native speakers of Western Arabic for whom dialectal characteristics should be perceptually significant.

To produce a signal in which segmental information has been removed, we extracted from the original voice signal the values of Fo and energy every 20ms and used Matlab to generate sinusoidal signals having the same frequency and amplitude (i.e. prosody) as the original speech signal. When there was no fundamental frequency signal (i.e. no voicing) it resulted into silence. In this way the original speech signal was converted to a "buzz" having the same amplitude, frequency and timing (i.e. relative timing of voice-on vs. voice off).

In addition we constructed two training sessions which included speech samples from 4 extra speakers of Arabic Western and Eastern speakers), both in their original unprocessed form and in the processed synthesised version (synthetic stimuli). The training passage plus instructions followed by the 48 test samples in "natural" then "buzz" form, randomised, were dubbed onto master stimulus CDs for presentation to listeners. Each test item was followed by 2 seconds of silence during which the dialectal variety was to be identified on a formatted scoring sheet. For both experiments, subjects were asked to identify the stimuli in terms of zone (i.e. Western variety vs. Eastern variety).

3. RESULTS

The results are given in Figures 1 through 4. Listed in percentage on the vertical axis are the correct identification rates obtained by the subjects.

Figure 1 shows the overall score obtained by the two subjects-populations (i.e. Arabic vs. non-Arabic listeners) for the first task (i.e. identification of Arabic dialects per zone in natural speech). We observe 97 % of correct identification for the Arabic subjects and 56% for the non-Arabic ones. These results confirm the assumption that the linguistic classification of the Arabic dialect area adopted for the study is significant for native speakers on the basis of perceptual cues. Statistical analysis (ANOVA, Fisher's PLSD and one-tailed t-test) reveal :

- that the scoring differences existing between the two populations is highly significant :

  Population effect : ($F_{(1.36)}$ = 259,838, $p < .0001$) S.

- that the correct identification rate obtained by the non- Arabic subjects (56%) is higher than the ratio 1:2 which would be expected by chance ($p < ,05$) S

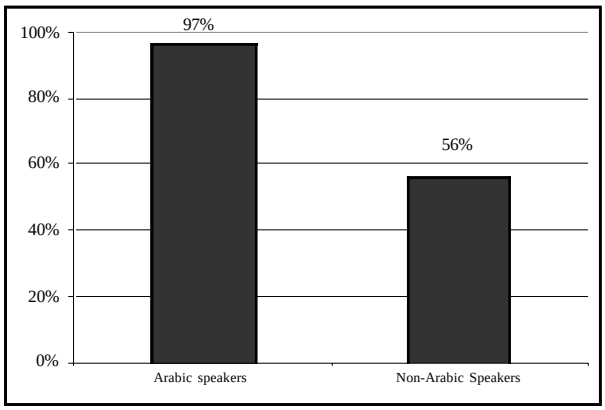

Fig.1 : % of Correct Identification for the discrimination of Arabic dialects in terms of geographical zone by the two-subject-populations in Natural Speech.

Figure 2 below shows the identification rate obtained by the two subjects-populations (i.e. Arabic vs. non-Arabic speakers / listeners) for the second task (i.e. identification of Arabic dialects per zone in synthesised speech). We observe 58 % of correct identification for the Arabic subjects and 49% for the non-Arabic ones. Statistical analysis (ANOVA, Fisher's PLSD and one-tailed t-test) show :

- that the population effect is still significant to explain the score differences (i.e. Arabic subjects' correct identification rate are higher than non-Arabic subjects).

  Population effect : ($F_{(1.36)}$ = 4,470, $p < .0415$) S.

- that the correct identification rate obtained by the Arabic subjects (58%) is higher than the ratio 1:2 which would be expected by chance ($p < ,0020$) S.

- that the prosodic criterion does not seem to be 'relevant' for the non-Arabic subjects whom 49% of correct answers are not significant.

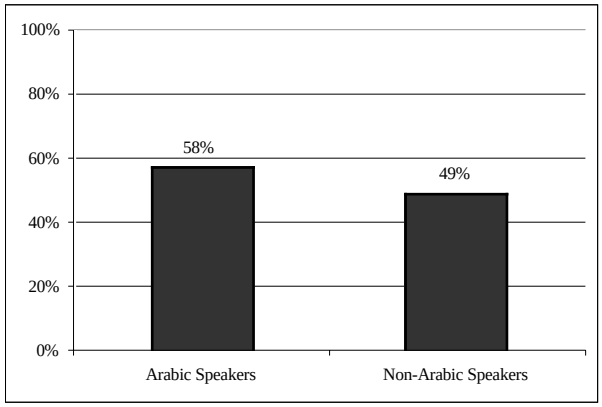

Fig.2 : % of Correct Identification for the discrimination of Arabic dialects in terms of geographical zone by the two-subject-populations in Synthetic Speech

[1] Mayer, M., 1969, Frog, where are you ? Sequel to a Boy, a Dog and a Frog, Book for Young Readers, New-York, 15 pp.

These scores seem to confirm the hypothesis that prosodic patterns alone help for the discrimination of Arabic dialectal varieties in terms of geographical zones.

Examining the rate of correct identification for each group of stimuli (western vs. eastern) as obtained by each one of the two subject populations reveals that there is no significant difference between the score reached for the recognition of Western and Eastern varieties in natural speech for both Arabic and no-Arabic subjects (Fig. 3).

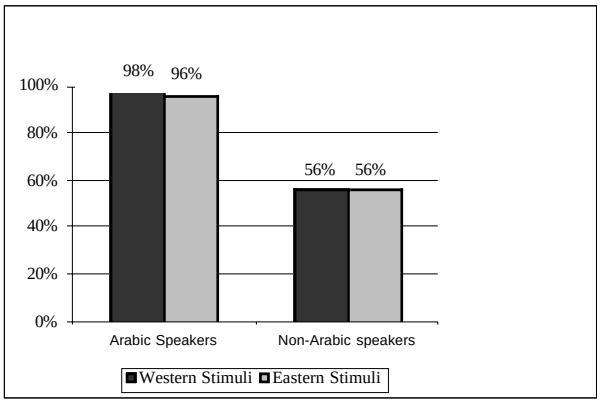

Fig.3 : % of Correct Identification function of stimuli origin by the two-subject-populations in Natural Speech.

Nevertheless, as shown in Figure 4 below, Western Arabic listeners' rate of identification for their own language area (Western stimuli) under its processed form is significantly higher than identification of the other dialectal variety ($p < 0,005$).

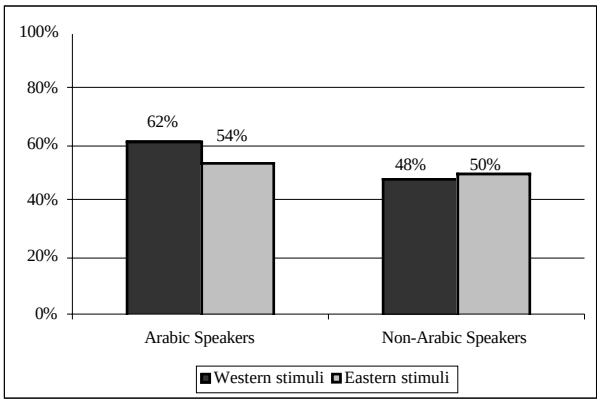

Fig.3 : % of Correct Identification function of stimuli origin by the two-subject-populations in Synthetic Speech.

4. CONCLUSION

This study on the perceived areal prosodic characteristics for vernacular Arabic highlighted the following points : Speakers / listeners of Arabic are aware of prosodic dialectal differences and are significantly more successful in identifying the stimuli corresponding to their dialectal area even when they only have access to the prosodic pattern. These results show that they perceive a particular pattern related to Western Arabic. Using a group of middle-eastern subjects should confirm the idea that both Western and Eastern Arabic dialects exhibit different prosodic patterns.

In this study we have tried to focus attention on the role of prosodic information for the discrimination of Arabic dialects. In the course of our research we shall attempt to evaluate the diagnostic potential of this cue for the automatic identification of Arabic dialects by developing a recognition model based on prosodic information. The combination of prosodic parameters with our already existing phonetic model [14] should increase its performances.

5. REFERENCES

[1] Ohala, J. & Gilbert J. (1979) , *'Listeners' ability to identify languages by their prosody'*, in Problème de prosodie Vol. 2, pp. 123-131, P. Léon & M. Rossi (eds).

[2] Maidment J. (1983), *'Language recognition and prosody : further evidence'*, in Speech, Hearing and Language : Work in progress, U.C.L. n°1, pp. 133-141.

[3] Ramus, F., & Mehler, J. (1999*), 'Language identification with suprasegmental cues: A study based on speech resynthesis'*. J.A.S.A. 105(1), pp. 512-521.

[4] Thyme-Gobbel, A. & Hutchins, S. (1999), *'On using prosodic cues in Automatic Language Identification'*, in ICSLP 96.

[5] Barkat, M. (1999), 'Identification of arabic dialects and experimental determination of distinctive cues', in Proceedings of the 19th International Congress of Phonetic Sciences, San Francisco, USA.

[6] Nejmi H., & Bouziri, R. (1988), 'L'accent de l'arabe parlé à Casablanca et à Tunis : étude phonétique et phonologique', in Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en Provence, Vol.5, pp. 134-137.

[7] Benikrane, T., (1997*), 'Intonation in Western Arabic (Morocco)*' in Intonation Systems : a Survey of Twenty Languages. Cambridge University Press, Cambridge, Hirst, D. & Di Cristo, A., (Eds.) pp. 349-363.

[8] Hirst, D. **&** Di Cristo, A. (Eds.), (1997), Intonation Systems : a Survey of Twenty Languages. Cambridge University Press, Cambridge.

[9] Thorsen, N., (1977) 'On the interpretation of row fundamental frequency tracings", in Actes des 8èmes Journées d'Etudes sur la Parole, Aix-en-Provence, pp. 175-182.

[9] Haydar, Y. & Mrayati, M., (1985*) 'Etude de l'intonation, la courbe mélodique de phrases de l'arabe standard'*, Travaux de l'Institut de Phonétique de Strasbourg, n°17, pp. 75-113.

[12] Marçais, Ph., (1977), Esquisse grammaticale d'arabe maghrébin, Librairie d'Amérique et d'Orient, Maisonneuve (ed), Paris, pp. 24-25.

[13] Barkat, M., (1999), "Détermination d'indices acoustiques robustes pour l'identification des parlers arabes : la dispersion vocalique comme critère pertinent de discrimination dialectale", in Actes des Premières Journées d'Etudes sur l'Identification Automatique des Langues, Lyon, F. Pellegrino (ed), [to be published].

[14] Pellegrino, F. & André-Obrecht, R. (1999), 'An Unsupervised Approach to Language Identification', in Proceedings Of ICASSP'99, March 1999, Phoenix.