RESEARCH ARTICLE OPEN ACCESS

Creating a High-Quality Syrian Audio Database for Analysis of Speaker Personality

Raneem Knaj [1], Jafar Al- kheir [2]
Department of Computer and Control Engineering
University of Tishreen
Latakia - Syria

ABSTRACT
The majority of this study is examining the variations in the acoustic characteristics of vowels across 36 Syrian native speakers from nine Syrian dialects. For created speaker database containing 18 males and 18 females, speakers have been recorded 24 syllables in optimal recording conditions. Long and short vowel polygons were generated and analyzed in F1-F2 formant plane. The main motivation of this research is finding the most suitable Syrian vowel polygon for dialect distinction. Dialects' varieties have been observed by the polygons' area and the length of vector between the polygons' centroids. Long vowel polygons for females reached the best results in both criteria. Presented work can be useful for studying the effect of gender and age on vowel polygons not only Syrian dialects.

Keywords :— Vowel polygons, formant, speech database, dialects, Syrian.

I. INTRODUCTION

The intelligibility of generated speech considered by any uttering automatic device is the measurement for its efficiency [1]. Although, most of the devices used nowadays for reading texts may produce understandable speech but it is automated, standard and far from natural speech properties, so the great need to a high quality automatic generator for the tone emerged. There are many researches and modern projects in the field of generating speech systems [2], which are all based on the analysis of phonetically labeled speech corpora. According to the shortage of achieved actions in the domain in the Syrian language, this research offered the following stages to achieve a speech database for the Syrian.

Speech database represents the main pillar in constructing communicating systems with the computer, to contribute in the expansion of the computer users to include blind, illiterate and those with special needs in addition to the possibility. One of the most important challenges facing the speech recognition systems is the ignorance of acoustic vowel spaces in any dialect. This research allows developing a suitable tool to understand the Syrian speech in all different public dialects automatically. That is through knowing the acoustic vowel space, which is really the area of the vowel polygon [3]. The main motivation of calculating the areas of the vowel polygons is to present a novel metric for the study of speech production deficits and reductions in intelligibility, in addition to the traditional study of vowel distinctiveness.

Recently in this field, various researches are done for generation and analysis of vowel polygons in many languages, e.g. set polygons for Arabic speakers by Nabil and Hesham [4]. Another paper presented developed software for generating and analyzing vowel polygons. It was tested by Czech native speakers [5]. Our study presents a helping tool to analyze the Syrian vowels by generating the Syrian vowel polygons developed in MATLAB environment.

First, we make records of each vowel for the possible highest amount of speakers. From these records, the means of Mel Frequency Cepstral Coefficients (MFCC) and its standard deviations are analyzed for each single vowel. In addition, delta/velocity coefficients means and their standard deviations as well as, delta-delta/acceleration coefficients means and their standard deviations are calculated. A covarriation matrix is done from all means values. The mined MFCC means is employed as inputs to a Feed-Forward Neural Network (FFNN). In this work, two-layered FFNN with eight hidden neurons was used for six Syrian vowels. It is necessary to do few-times training on the designed neural network for reaching the best results. Formant frequencies are estimated using Linear Predictive Coding (LPC) spectra [6], [7].

II. SPEAKERS

When taking a random sample for representing a huge group of people it is important that the selecting is accurate as much as possible. The best representing is to choose the total numbers of the group – in the case of this research the selecting is for all the Syrian dialects – but it is difficult to do this. This research aims that the sample involves the difference in utterance. The correct method is to find a sample of each group representing a particular region. A linguistic map for the Syrian Arab Republic has never been done before. Consequently, it is hard to make sure of the number of dialects and the geographical area, which they cover. Therefore, the alternative is to choose samples from different regions whose inhabitants represent different dialects [8], [9]. They cover

main different dialects in Syria and the result is the following information database about the elements of the created speech database, see Tab. 1.

TABLE I
SPEAKERS INVOLVED IN CREATION OF THE SPEECH DATABASE.

Speaker no.	Birthday year	Gender	Region
1	1993	Male	Aleppo
2	1993	Female	Aleppo
3	1994	Male	Aleppo
4	1995	Female	Aleppo
5	1995	Male	Damascus
6	1992	Female	Damascus
7	1990	Male	Damascus
8	1990	Female	Damascus
9	1990	Male	Al-Hasakah
10	1990	Female	Al-Hasakah
11	1986	Male	Al-Hasakah
12	1990	Female	Al-Hasakah
13	1991	Male	Idlib
14	1995	Female	Idlib
15	1991	Male	Idlib
16	1991	Female	Idlib
17	1992	Male	Al-Suwayda
18	1989	Female	Al-Suwayda
19	1990	Male	Al-Suwayda
20	1994	Female	Al-Suwayda
21	1988	Male	Daraa
22	1988	Female	Daraa
23	1987	Male	Daraa
24	1991	Female	Daraa
25	1995	Male	Latakia
26	1991	Female	Latakia
27	1989	Male	Latakia
28	1992	Female	Latakia
29	1994	Male	Tartous
30	1993	Female	Tartous
31	1992	Male	Tartous
32	1990	Female	Tartous
33	1991	Male	Hama
34	1993	Female	Hama
35	1992	Male	Hama
36	1990	Female	Hama

People in the database speak the above-mentioned dialects as native dialects. They and their parents are local population. They spent their childhood in the mentioned regions. All of them lived more than sixteen years in the mentioned regions. People in the database are academic, educated and do not suffer from any problem in hearing or pronunciation. The database consists of 18 males and 18 females. Their ages are between twenty and thirty. The selection of people is based on geographic and gregarious criteria. It is notable that the people talk their dialects clearly.

III. LINGUISTIC MATERIALS

The ejected syllables are both opened syllables, which end with vowels, and closed syllables that end with constants. While the opened syllables are labeled as CV where the letter C means a constant and the letter V means a vowel [10], the closed syllables label as CVC. In the Syrian dialects, there are three long vowels namely /aa/-/ii/-/uu/ and three short vowels namely /a/-/i/-/u/. Every speaker says the following syllables twice: SUUS, SAAS, SIIS, SUS, SAS, SIS, SUU, SAA, SII, SU, SA, SI, DUUD, DAAD, DIID, DUD, DAD, DID, DUU, DAA, DII, DU, DA, DI. The syllables are written on a square card. The speakers record all syllables in an isolated studio. All records are saved in a WAV format. The speakers pronounce the twenty-four syllables twice in the studio, so the speech database consists of 1728 syllables. The enunciation is done in natural conditions of speed and loudness [11], [12].

IV. EQUIPMENT

Recording is done in an isolated studio whose walls, floor and ceiling are processed with soundproof materials. Studio without any windows are designed to prevent the noise infiltration into the studio. The studio's door is also soundproof. The studio is linked to a monitor room through several devices. The wall between the studio and the monitor room is Plexiglas. Each speaker records separately and all speakers finish recording within one session. Syllables are recorded by using head-mounted microphone positioned at a distance of 2 inches from the speaker's lips. The speech signal is sampled with the frequency of 44100 Hz with 16-bit accuracy and analyzed in a 20 ms Hamming window with 10 ms overlapping. During the recording operation, dynamic microphone, sound craft digital, Adobe Audition3 software and stereo channel are used.

V. EXPERIMENTAL RESULTS

By means of the developed software tool, the vowel polygons of all speakers are generated and Figure 1 illustrates the long vowel polygon for a male from Aleppo.

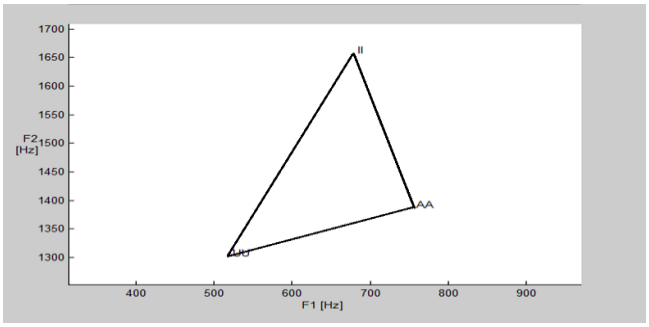

Fig. 1 Long vowel polygon for male no. 1 from Aleppo.

Obtained Syrian vowel polygons are generated in the F1-F2 plane and defined in four groups. These groups are long vowel

polygons for males, short vowel polygons for males, long vowel polygons for females and short vowel polygons for females. Each group is defined in F1-F2 formant plane. The main motivation of this study is finding the most suitable vowel polygons for dialects distinction by two different criteria. The first criteria is depended on the area of created Syrian vowel polygons. These areas are calculated using Matlab's polyarea function. For each group, the difference between the maximal value of the areas and the minimal one is calculated and set as ΔS. The max result have been achieved by long vowel polygons for females in F1-F2 formant planes. The distribution of the areas is illustrated in Fig. 2 for the group of long vowel polygons for females. The values of maximal areas, minimal area and ΔSmax of the four groups are shown in Tab. 2.

TABLE IIIII
VALUES OF MAX, MIN AND ΔSMAX AREAS FOR ALL GROUPS.

Group Name	Maximal Area [Hz^2]	Minimal Area [Hz^2]	ΔSmax [Hz^2]
long vowel polygons for males	31986	8437	23549
short vowel polygons for males	4081	189	3892
long vowel polygons for females	45458	17356	28102
short vowel polygons for females	13400	554	12846

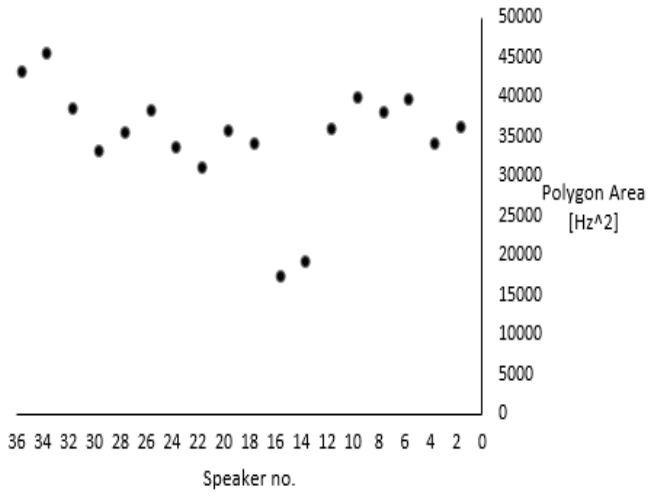

Fig. 2 The distribution of the areas of long vowel polygons for females.

The second criteria is depended on the length of created vector L generated by different polygons' centroids. The vector L directs from initial point created by Center Of Gravity (COG) of current speaker to COG of all other speakers. The length of each vector L is calculated and then the length differences between all possible vectors couples are observed in each group. Then the minimal value of the vector L in each group is set as ΔLmin and compared with ΔLmin in the other groups. The most suitable vowel polygons for dialects distinction are involved within the group that reaches the maximal ΔLmin. Long vowel polygon for females reach the best results. See Tab. 3. And the distribution of COG positions for long vowel polygons for females is illustrated in Fig. 3.

TABLE IVVVI
VALUES OF ΔLMIN VECTORS FOR ALL GROUPS.

Group Name	ΔLmin [Hz]
long vowel polygons for males	16
short vowel polygons for males	9
long vowel polygons for females	23
short vowel polygons for females	11

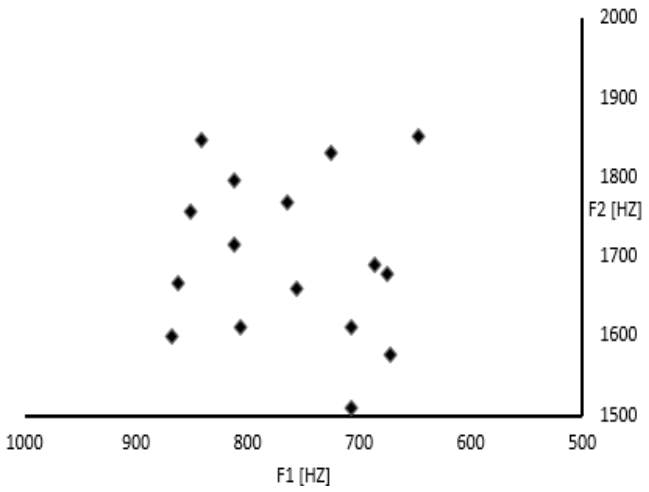

Fig. 3 The distribution of the COG positions of long vowel polygons for females.

VI. CONCLUSIONS

This work describes the steps for creating a speech database for various Syrian dialects. Information about speakers e.g. age, gender, sex and region are illustrated. The created speech database consists of 36 speakers from different regions and genders for studying the effect of the variety of regions and genders on the Syrian vowel polygons. The speakers are approximate ages so the effect of age is not studied. It will be useful to develop the created speaker database by adding speakers that differ in age - children, teens, youth and elderly. Two criteria are depended on to choose the most suitable Syrian vowel polygon for dialects distinction. The first criteria is the polygons' area and the second one is the length of vector between the polygons' centroids. Long vowel polygons for females reach the best results in both criteria. In future work,

it is important to create a telephone speech database for Syrian speakers that represent all Syrian groups whence dialect, gender, age and all available communications in the Syrian Arab Republic, then using the presented observations for vowel polygons in other languages.

ACKNOWLEDGMENT

The research that has led to this work has been supported in part by the Tishreen University Enterprise (RM5/2017), Research Project SDPP (Syrian Database Personality Perception). The authors wish to thank Nour Ghadban for her help on the psychological aspects of this work and for the technical support.

REFERENCES

[1] YAN, Q., VASENGHI, S. Analysis, Modelling and Synthesis of Formants of British, American and Australian Accents. In Proc. International Conference on Acoustics, Speech and Signal Processing. Hong Kong (China), 2003, pp. 712-715. DOI: 10.1109/ICASSP.2003.1198880

[2] STANEK, M., SIGMUND, M. Speaker Dependent Changes in Formants Based on Normalization of Vowel Triangle. In Proc. 23rd International Conference RADIOELEKTRONIKA. Pardubice. Czech Republic, 2013, pp. 337-341. DOI: 10.1109/RadioElek.2013.6530941

[3] STANEK, M., POLAK, L. Algorithms for Vowel Recognition in Fluent Speech Based on Formant Positions. In Proc. 36th International Conference on Telecommunication and Signal Processing. Rome (Italy), 2013, pp. 521-525. DOI: 10.1109/TSP.2013.6613987

[4] NABIL, A., HESHAM, M., Formant distortion after codecs for Arabic. In Proc. Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on, pp. 1-5. DOI: 10.1109/ISCCSP.2010.5463385

[5] STANEK, M. Software for generation and analysis of vowel polygons. In Proceedings of the 37th International Conference on Telecommunications and Signal Processing. Berlin (Germany), 2014, p. 721 - 724. DOI: 10.1109/TSP.2015.7296358

[6] SIGMUND, M., ZELINKA, P., Analysis of Voiced Speech Excitation Due to Alcohol Intoxication. Information Technology and Control, 2011, vol. 40, pp. 145-150. Print ISSN: 1392-124X, Online ISSN: 2335-884X

[7] ABDO, M. S., KANDIL, A. H., El-BIALY, A. M. Automatic detection for some common pronunciation mistakes applied to chosen Quran sounds. Biomedical Engineering Conference (CIBEC), 5th Cairo International, 2010, pp. 219-222. DOI: 10.1109/CIBEC.2010.5716073

[8] ALQAHTANY, M. O., ALOTAIBI, Y., SELOUANI, S. Analyzing the seventh vowel of classical Arabic. International Conference on Natural Language Processing and Knowledge Engineering, 2009. pp. 1-7. DOI: 10.1109/NLPKE.2009.5313729

[9] SEDDIQ, Y. M., ALOTAIBI, Y., Formant-based analysis of vowels in Modern Standard Arabic—Preliminary results, 11th international conference on information science, signal processing and their applications (ISSPA), 2012, pp. 689-694. DOI: 10.1109/ISSPA.2012.6310641

[10] GHULAM, M., KHALID, A., TAMER, M., MANSOUR, A. Automatic Arabic Digit Speech Recognition and Formant Analysis for Voicing Disordered People. Proceedings of IEEE Symposium on Computers & Informatics, 2011, pp. 699-702. DOI: 10.1109/ISCI.2011.5959001

[11] NABIL, A., HESHAM, M. Formant Distortion after Codecs for Arabic, Proceedings of the 4th International Symposium on Communications, Control and Signal Processing, (ISCCSP), 2010, pp TBC. DOI: 10.1109/ISCCSP.2010.5463385

[12] ALGHAMDI, M., EL HADJ, Y. O. M., ALKANHAL, M. A Manual System to Segment and Transcribe Arabic Speech, Proceedings of IEEE International Conference on Signal Processing and Communications (ICSPC), 2007, pp: 233–236. DOI: 10.1109/ICSPC.2007.4728298