Formant maps in Hungarian vowels – online data inventory for research, and education

Kálmán Abari[1], *Zsuzsanna Zsófia Rácz*[2], *Gábor Olaszy*[2]

[1] University of Debrecen, Hungary and [2] Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Hungary

abari.kalman@arts.unideb.hu, zsuzska.racz@gmail.com, olaszy@tmit.bme.hu

Abstract

This paper describes a project for creating an online system for studying the main formant movements of Hungarian vowels in spoken words, as a function of their sound environment. The speech material and the formant data corresponding to the vowels combined present research data for many other purposes as well. For efficient presentation of the data and to allow multilevel comparisons among formant features an online solution was developed. The inventory data can be regarded as a reference, because of the strict conformity between the defined formant data and the formants of the spoken words. A two-step manual verification phase after the completion of automatic formant tracking was performed. The on-line query ensures quick and wide spread studying of formant maps in vowels. The database is available at: "http://hungarianspeech.tmit.bme.hu/formant".

Index Terms: formant map, Hungarian vowels, live measurements, coarticulation, evaluation material.

1. Introduction

Formant measurements represent one of the oldest areas in speech research. A wide range of methods are known for formant tracking [1] [2] [3] and public software tools have also appeared in the last decade [4]. Research activity also aims at finding solutions for successful formant tracking in noisy environments [5]. Researchers know the advantages and limitations of these algorithms, namely that automatic formant measurements can guarantee results only with a certain error rate. The common situation when measuring formant data is to apply individual measurements. Researchers carry out their formant measurements to suite their own goals, and they publish only the results of the work, while the measured numerical formant data remain hidden. If a formant database is published then – in many cases – a number of questions can be answered by using the services of the database (e.g. formant distances between lower and higher formants, differences in mean values, the characteristics of formant movements). There is only a small number of public databases concerning the formant data of a given speech corpus. One of them was published by Deng Li et al. [6] for English to support speech research. In their paper the authors describe their goal: to create a publicly available database of the first three vocal tract resonances of 538 English sentences from the TIMIT speech corpus. This database is available for download. An Arabic formant database was used by Jemaa et al. [7] for the evaluation of a new automatic formant tracking algorithm based on Fourier ridges detection. For forensic purposes preliminary formant data have been determined for a Swedish dialect database [8].

Work on a Hungarian formant database begun recently [9], and the final version of the public online data inventory is described in this paper.

2. The goal of the project

The aim of the project was – on one hand – to construct a relevant data inventory about the formant maps of Hungarian vowels as a function of the adjacent sounds. On the other hand the goal was to create a query system, which allows multilevel comparison of the data. The results are given by numbers (mean, standard deviation, minima, maxima) and also in the form of different statistical distributions. An additional goal was to build up a fully stable and complex (audio and data) inventory for testing algorithms dedicated to define formant values in speech.

3. Material, method and development steps

The word list has 1832 isolated items, 3-8 syllables/word. It contains all CV, VC and VV combinations and also all combinations of CVC sequences as a function of 6 articulation positions for the consonants (6x6x9=324 different types). These positions are marked with BB for bilabial, LD for labiodentals, DA for dentialveolar, AL for alveolar, PA for palatal and VE for velar. Only the articulation position is taken into consideration, for example BB involves [b p m]. All fourteen Hungarian vowels are represented in the word list. The distribution of vowels in the inventory (Table 1.) represents their frequency in Hungarian.

Table 1. *The number of vowels in the word list (m = male, f = female)*

	[aː]	[ɔ]	[o]	[u]	[y]	[i]	[eː]	[ø]	[ɛ]
m	681	1402	645	230	101	738	447	239	1246
f	682	1395	645	224	100	723	444	234	1241
			[oː]	[uː]	[yː]	[iː]		[øː]	
m			268	65	67	85		148	
f			274	65	69	84		139	

The word list was read by a male (age 60) and a female (age 32) skilled, native Hungarian speaker. The corpus was recorded in an anechoic chamber at 16 bits and 22 kHz. The total speaking time was 28 minutes/speaker. The words were pronounced neutrally (without accent). The phonetic transcription was completed using an automatic method, sound symbols were adopted from a Hungarian TTS program [10]. The mapping of the Hungarian letters, the IPA symbols and the sound symbols used here are: á/[aː]/A:, a/[ɔ]/a, o/[o]/o, ó/[oː]/o:, u/[u]/u, ú/[uː]/u:, ü/[y]/U, ű/[yː]/U:, i/[i]/i, í/[iː]/i:, é/[eː]/E:, ö/[ø]/O, ő/[øː]/O:, e/[ɛ]/e. The sound boundaries were defined semi-automatically.

The formant definition process consisted of three steps, the first being the automatic formant tracking itself. The program Praat was selected [4] from the several computerised formant estimation methods available, because it is widely known and can be programmed using a simple scripting language which makes processing large data sets possible. Three measurement positions have been defined in the vowels to follow and describe the main formant movements (the 25%, 50% and 75% points of the vowel duration). In vowels at the beginning of the words only the 50% and 75% points, while in vowels at the very end of the words the 25% and 50% points represented a measurement position. The parameters of calculation (e.g. a window length of 25 ms) were the same as the defaults in Praat with the exception of the maximum formant frequency, where 5000 Hz and 5500 Hz were used for the measurement of the male and the female speech, respectively. Four formants have been defined by the program for every measurement point. In total 144 396 formant values make up the database. In the second step crude measurement errors were screened and corrected. The evaluation of the data was performed using a spreadsheet application and a statistical analyser. Criteria taken into account included formant ranges, formant distances and formant correlations (e.g. F1-F2, F2-F3, F3-F4 maps). The screening could only be partly automated, therefore a large part of it was completed manually [9]. In the third step visual examination was performed on plots (produced by another program [11]) that showed the measured formant data and the spectrum together (Figure 1.). 11014 formant values in total were corrected in the database. The higher the formant, the more inaccurate the estimation proved to be (Table 2).

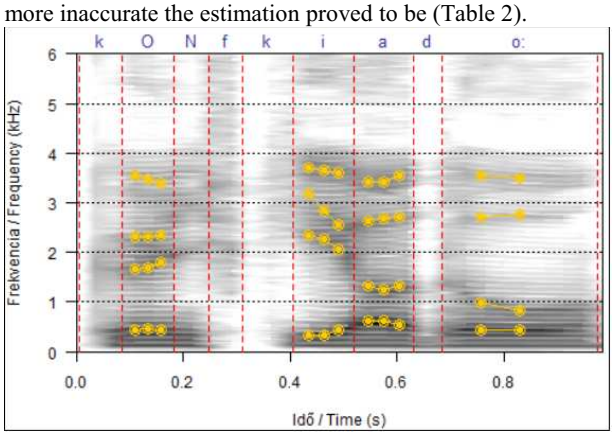

Figure 1: *The plot of a single word for visual inspection of the measured formant data in the vowels*

Table 2. *The number of manual corrections in the formants*

	F1	F2	F3	F4
male	198	245	272	302
female	203	1654	3267	4873

The efficiency of the automatic estimation used was found to be better for the male voice than for the female [9]. As a result of the evaluation and correction process a data inventory with a negligible number of errors has been created, that describes the formant values and movements in Hungarian vowels (represented by the spoken word list).

4. Design of the query system

An interactive query system makes the use of the database simple and efficient. It is built entirely with free software and relies on open standards. The web-based query system is written on a LAMP stack (Linux operating system, Apache web server, MySQL database management system and PHP for the server side programming). The user is able to access the formant data from any computer platform that has a standard compliant web browser and an internet connection. The web-based system uses Protovis (a JavaScript data visualisation library being developed at Stanford) [12] to draw the formant data.

Searches can be performed based on the following properties and their combinations: male/female voice, 14 vowels, the neighbourhood of the vowel defined by individual sounds or group identifiers as C=consonant, V=vowel, B=any sound within the word, #B/B#=any sound, #=word boundary position and the markers of the 7 articulation places given in Section 3. One can also search on absolute formant values, formant distances and different predefined trajectories. The data inventory has been designed first of all for the examination of coarticulation in CVC sequences regarding the place of articulation of the consonants. All possible combinations can be studied for 9 vowels (the short and long pairs can be regarded as identical) and 6 articulation positions. The results of each query are presented in two forms: statistical analysis and diagrams.

Before performing a search command, the user has to define the search space and also the form of presentation of the results.

The word list of the database can be displayed as well. Visual (spectrogram) and audio (sound) forms are available for each word of the list.

4.1. Search settings

Vowel definition. The selection is flexible: any subset of the 14 vowels can be defined.

Vowel neighbourhood. Three sounds (or sound groups), both before and after the vowel can be defined (Figure 2.).

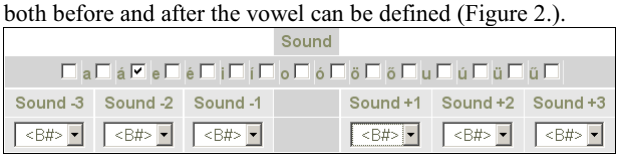

Figure 2: *Example for selecting a vowel and its environment. The sound [a:] is selected without any restriction ofits neighbourhood*

Search of formants. After defining the search details (measurement point, frequency boundaries) separately for the four formants, the results will be displayed.

Formant distances. The distance between two formants (F1-F2, F2-F3, F3-F4) may be defined using a threshold Hz value.

Formant trajectory. Nine basic patterns were predefined for selection. Each pattern consists of two linear parts (representing movement between measurement points 25%-50% and 50%-75%). The precise form of the pattern can be defined by threshold values in Hz (Figure 3.). Only the pattern defined by the thresholds is examined, independently of the absolute formant frequencies.

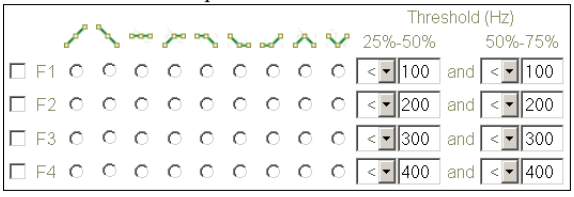

Figure 3: *The 9 possible predefined formant movement patterns and their threshold adjustment for selection*

Search by value. The user can define an arbitrary pattern by giving the formant frequency and the tolerance values in Hz. Using this adjustment the characteristic formant frequencies can be taken into account as well. The form of the formant movement and the frequency range of its occurrence can be selected at the same time.

Gender of the speaker. Data from both speakers or separate from the male or female can be displayed.

4.2. Display settings

It is possible to display the final result only or the detailed one (Figure 4.). By the latter all words that meet the filtering criteria are listed.

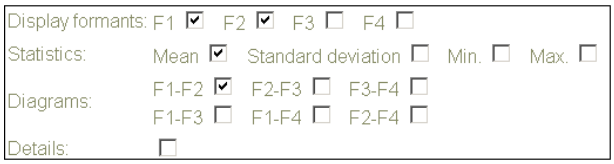

Figure 4: *The setting possibilities for results display*

The following data can be displayed: F1, F2, F3, F4 values, statistical results (average, standard deviation, minima and maxima). Diagrams are also available in different forms (for example Figure 5.).

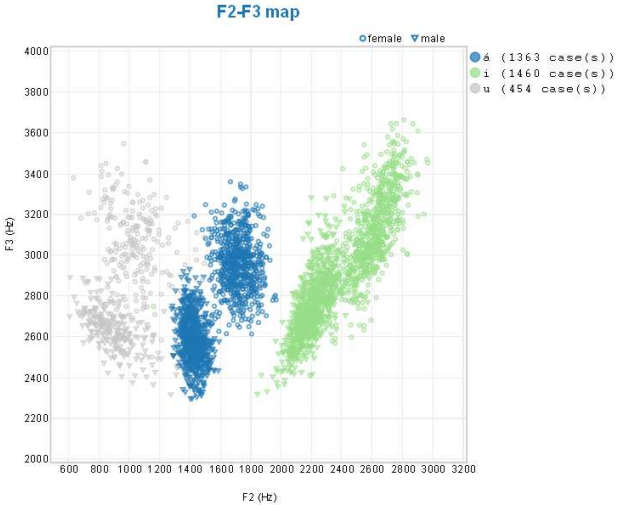

Figure 5: *An example of the diagrams the user can display in the online front-end*

5. Query examples

The query system gives maximal freedom for the user to get any data (average, distribution, sound environment of the vowel, direction of the formant movement, distance between formants etc.). Let us give some examples.

a) The mean values and the range of the formants of all vowels can be displayed depending on the measurement position. Figure 6. shows the plot from the data at the middle of the vowels.

b) Formant distances can also be measured on discrete vowel level (Table 3.). Results in Table 3 show the distances between the formant pairs in [aː]. There are no notable differences concerning the minimum distance values between the male and female data in the case of F1 and F2 or F2 and F3 but there are in the case of F3 and F4. Such measurements may be important to make speech analysis algorithms more sensitive.

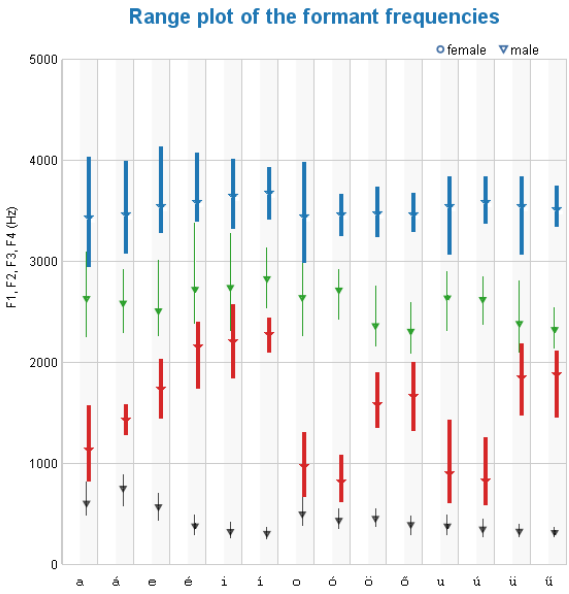

Figure 6: *The mean values and the range of F1, F2, F3, F4 formants (male) for the 14 vowels at 50% of their duration*

Table 3. *The minimum and maximum distances (male/female) between two formants of [aː] at the middle of the vowel (Hz)*

[aː]	F1-F2	F2-F3	F3-F4
min.	481/458	835/800	305/767
max.	710/1269	1564/1765	1402/1969

c) In the next example we give the answer to the question: which articulatory configurations result in constant formant values during the whole vowel? As the formant value is the same in the 3 measurement positions, the formant can be regarded to be level (not moving) within the vowel. The measurement was performed for F1 and F2 with the following adjustments: a predefined "level" pattern (see Figure 3.), an interval threshold of 5 Hz for F1 and 10 Hz for F2. The sound environment of the vowel was not limited (B#). Table 4. shows the results regardless of the gender.

Table 4. *The number of environmental cases where F1 or F2 have constant level along the whole vowel*

	[aː]	[ɔ]	[o] [oː]	[u] [uː]	[y] [yː]	[i] [iː]	[eː]	[ø] [øː]	[ɛ]
F1	0	6	16	27	16	134	32	9	0
F2	11	13	6	1	1	8	1	5	22

The results show that the lower the F1, the higher the number of cases where the formant has a constant value within the whole vowel. In case of F2 the tendency is the same. The precise sound environment can also be defined by further queries for all cases given in Table 4.

d) The following measurement shows the examination of the 'rise up-go down' movement in F1. The main questions are: which CVC combinations result in the most dynamic movement in F1? What is the maximum distance between the middle of the vowel and the 25% and 75% measurement positions? Let us adjust the settings as follows: from the *Formant trajectory* select the 'rise up-go down' pattern and set the thresholds to >210 Hz. The result of the search shows that only one female CVC item [raːng] fulfils the input definition,

in the word *ezüstpisztráng* (silver trout), and the F1 values in the three measurement positions are: 25%=676 Hz, 50%=912 Hz, 75%=678 Hz. (Figure 7.).

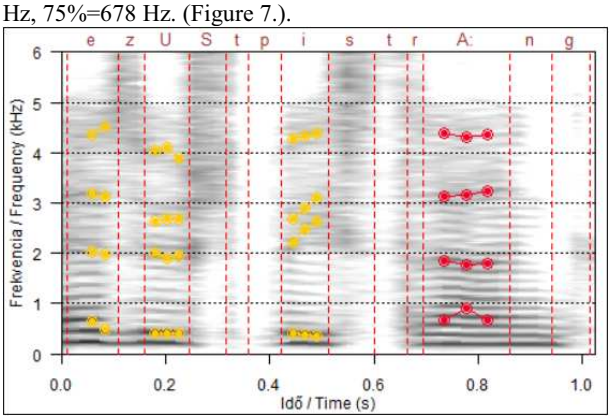

Figure 7: *The broadest 'rise up–go down' movement in F1 in the database (the vowel is highlighted)*

The second candidate (when reducing the threshold to >200 Hz) is the [saːr] sequence with the formants in [aː]: 786, 987, 774 Hz. The same measurement for male voice results in the following data. Using thresholds > 130 Hz the search shows that the CVC item [pɔk] fulfils the input definition in the word *kompaktlemez* (compact disc), and the F1 values in the three measurement positions are: 25%=543 Hz, 50%=680 Hz, 75%=522 Hz. The next hit is the [zɛt] sequence with the 275, 468, 341 Hz F1 values. The main conclusion is: the most active 'rise up-go down' formant movements in F1 occur in [ɔ aː ɛ] vowels where F1 is the highest in general.

e) In contrast, the examination of the 'go down-rise up' movement in F1 gives the following results. The most active movement occurs in the [voːt] sequence, i.e. 466, 358, 456 Hz at the measurement positions (threshold is >95 Hz, male voice). When reducing the threshold, the vowel [o] also appears among the hits. The conclusion is that the 'go down-rise up' form does not appear in those vowels having the lowest F1, like [u i y].

6. Formant tracker evaluation possibilities

The presented data inventory together with the labeled speech material can be used for evaluation of speech processing algorithms, especially formant trackers as well. The input for any formant tracker algorithm can be the spoken word list and the measured formant values can be compared to the corresponding values of the formant inventory. After a statistical analysis the effectiveness of the formant tracker may be assessed. As formant tracking in noisy environments represents an important research field for automatic speech recognition, a well controlled series of analysis can also be performed. Adding certain noise to the speech material of the original inventory, the robustness of a formant tracker (against noise) can be examined as well.

7. Conclusions

An online formant inventory for Hungarian was developed. It is based on a male and a female voice. The speech corpus belonging to the inventory contains separate spoken words. The labelling of the speech wave was performed semi-automatically the estimation of formants was realized by the Praat program. A two level correction procedure was used to find and correct the errors of the automatic formant estimation. The correction process resulted in a practically error free formant inventory of F1, F2, F3, F4 (144 396 items in total). The correction results showed that the automatic estimation works better for lower formants than for higher ones. Moreover formants in the male voice are estimated with a higher accuracy than those in the female one. A general purpose query system has been developed for the free and flexible use of the formant inventory. The presented system may help research and education, no preliminary recording, labelling and other work is needed to show the main features of formants (as a function of the sound environment). The speech database and the formant inventory can be used for the evaluation of new or existing automatic formant tracking algorithms as well.

8. Acknowledgements

This research was supported by the TÁMOP-4.2.2-08/1/KMR-2008-0007, TÁMOP-4.2.1/B-09/1/KMR-2010-0002) and the BelAmi: ALAP2-00004/2005 projects.

9. References

[1] S. McCandless S., "An algorithm for automatic formant extraction using linear prediction spectra", IEEE Trans. Acoust. Speech & Sig. Proc., Vol. 22, 1974, pp. 135-141.

[2] Bőhm T. and Németh G., "Algorithm for Formant Tracking, Modification and Synthesis", Infocommunications Journal LXII:(1), 2007, pp. 15-20.

[3] Zheng Y. and Hasegawa-Johnson M., "Formant tracking by mixture state particle filter", Proc. ICASSP, 2004, Vol.1, pp. 565-568.

[4] Boersma P. and Weenink, D., "Praat: doing phonetics by computer", (version 5.1.05. http://www.praat.org/, 2009. Computer program)

[5] Jonas L., "Formant tracking linear prediction model using HMMs and Kalman filters for noisy speech processing", Computer, Speech and Language 21. 2007. pp. 543-561.

[6] Deng L., Xiaodong C., Pruvenok R., Huang J., Momen S., Chen Y. and Alwan A., "A Database of Vocal Tract Resonance Trajectories for Research in Speech Processing", Proceedings of the ICASSP 2006, pp. 369-372.

[7] Jemaa I., Rekhis O., Ouni K. and Laprie Y., "An Evaluation of Formant Tracking methods on an Arabic Database", Proceedings of Interspeech 2009, pp. 1677-1670.

[8] Jonas L., "Preliminary Formant Data of the Swedia Dialect Database in a Forensic Phonetic Perspective", Proceedings of the 19th Annual Conference of the International Association for Forensic Phonetics and Acoustics, Trier, Germany, 2010.

[9] Olaszy G., Rácz Zs. and Abari K., "A formant trajectory database of Hungarian vowels", The Phonetician 97/98., 2008/2011, pp. 6-13.

[10] Olaszy G., Németh G., Olaszi P., Kiss G., Zainkó Cs. and Gordos G., "Profivox – a Hungarian TTS System for Telecommunications Applications", International Journal of Speech Technology. Vol 3. 2000, pp. 201-215.

[11] R Development Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[12] Bostock M. and Heer J., "Protovis: a graphical toolkit for visualization", IEEE Transactions on Visualization and Computer Graphics, 15(6), 2009, pp.1121-1128.