Automatic Selection of the Number of Poles for Different Gender and Age Groups in Steady State Isolated Vowels

Thayabaran Kathiresan[1], Dieter Maurer[2], Volker Dellwo[1]
[1] Phonetics Laboratory, University of Zurich, Switzerland, [2] Institute for the Performing Arts and Film, Zurich University of the Arts, Switzerland
thayabaran.kathiresan@uzh.ch, dieter.maurer@zhdk.ch, volker.dellwo@uzh.ch www.phones-and-phonemes.org

Session: 2aSPb
JASA, 2016, 140 (4), p. 3058.

Background

According to the actual phonetic practice, formant frequency estimation of vowel sounds using a linear prediction (LP) algorithm is based on the assumption of age and gender specific parameters, such as number of poles (corresponding to the maximum number of formants) related to a specific frequency range of analysis.

Yet, when visually crosschecking the calculated formant frequencies along with a spectrogram or a spectrum, investigators often change the number of poles because of a lack of correspondence [1]. Thereby, the "misprediction" is often considered as due to a high variation within the formant tracks [2], or LPC filters of analysis not matching spectral peaks, possibly combined with an unexpected low or high number of occurring spectral peaks (e.g., "formant merging", "spurious formants"; see e.g., [3]).

However, this manual change of the analysis parameters in the course of analysis lacks methodological objectification (see [4] for the inherent circularity).

The present paper addresses this question and presents an attempt for an objective procedure to select appropriate number of poles for LP in formant pattern estimation of a given vowel sound in terms of comparing the Euclidean distance (ED) for formant patterns calculated with different pole numbers, and selecting the pattern showing the lowest ED measure.

Below, the method proposed is described and the results of a first evaluation on a sample of vowel sounds are given.

Approach

For a given vowel sound, using an LP algorithm, the first three formant frequencies F_1–F_2–F_3 for three different settings of the number of poles are calculated. Thereby, a frequency range of analysis equal for the three settings is also set, considering general phonetic and practice of analysis. (For an example, see below.)

Subsequently, for each of the three formant patterns, the standard deviation (σ) of the formant tracks is used to assess the Euclidean distance, and the formant pattern related to the lowest value of ED is selected as "best option".

This approach is based on the expectation that appropriate number of poles of LP analysis may generally relate to the constancy of the calculated formant tracks, and in consequence to lowest ED measures, if values of LP analyses for different pole numbers are compared.

[1] Hillenbrand, J., Getty, L., Michael, J., and Kimberlee, W., 1995. "Acoustic characteristics of American English vowels", J. Acoust. Soc. Am. 97, 3099–3111.

[2] Kabir, A., Barker, J., and Giurgiu, M., 2010. "Robust Formant Estimation: Increasing the Reliability by Comparison among Three Methods", Proceedings of the International Conference on Circuits, Systems, Signals, 341–344.

[3] Chanwoo, K., Kwang-deok, S., and Wonyong, S., 2005. "A Robust Formant Extraction Algorithm Combining Spectral Peak Picking and Root Polishing", EURASIP Journal on Applied Signal Processing Volume 2006, Article ID 67960, 1–16.

[4] Ladefoged, P., 1967. "Three Areas of Experimental Phonetics". Oxford U.O., London.

The General Procedure

The following is applied to a single vowel sound independent of the age and gender of the speaker:

Step 1: The frequency range of analysis is set to a fixed range, e.g., 0–5.5kHz.

Step 2: LP analysis is performed three times in parallel, with three different number of poles related to the frequency range set; e.g., for a frequency range of 0–5.5kHz, option A = 10 poles (four formants in maximum); option B = 12 poles (five formants in maximum); option C = 14 poles (six formants in maximum).

Step 3: SD (σ) is calculated for the first three formants in all the three resulting formant tracks.

Step 4: ED is calculated from the σ for all three resulting formant tracks.

Step 5: The formant pattern related to the lowest ED value is selected.

Note that, in the example given, the options correspond to settings usually considered as appropriate for children (A), women (B) and men (C).

Euclidean Distance Measure

ED is calculated using the formula given below

$$ED = \sqrt{[\sigma(F_1)]^2 + [\sigma(F_2)]^2 + [\sigma(F_3)]^2}$$

where $\sigma(F_1)$, $\sigma(F_2)$, $\sigma(F_3)$ are the standard deviations of the first three formant tracks.

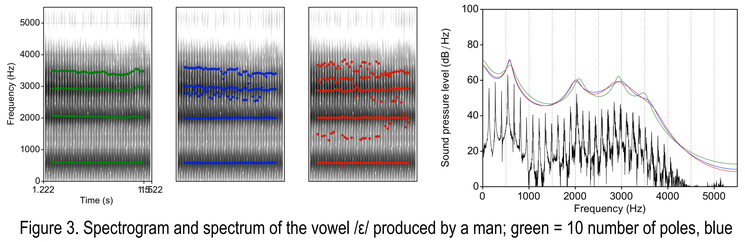

Figure 3. Spectrogram and spectrum of the vowel /ɛ/ produced by a man; green = 10 number of poles, blue = 12 number of poles 5, red = 14 number of poles.

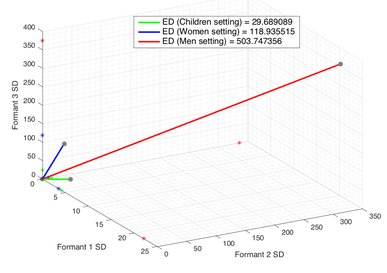

Figure 2. ED projection of the first three formants (in 3 dimensional space) of the vowel sound of /ɛ/ produced by a man.

Evaluation

The method presented was evaluated on a sample of the eight long Standard German vowel sounds /i, y, e, ø, ɛ, a, o, u/ produced by eight children (f_0 = 260Hz), 10 women (f_0 = 220Hz) and 10 men (f_0 = 130Hz), i.e., on 224 vowel sounds. – Steady state sounds were produced in isolation. Sound duration was 1–2 sec. Acoustic analysis was performed on the middle 0.3 sec of a sound (sound nucleus).

For a single sound nucleus and a fixed frequency range of 0–5.5 kHz, three patterns F_1–F_2–F_3 were calculated in PRAAT using LP (Burg) algorithm with three different number of poles 10, 12 and 14. ED was calculated for the three patterns F_1–F_2–F_3, and the pattern showing the lowest ED measure was selected as "best option".

For each single sound, the three formant tracks were visually crosschecked on the basis of the spectrogram and the spectrum of the vowel, and the following assessments were made for the selected pattern F_1–F_2–F_3 ("best option").

•Correspondence of the number of poles with the expected default value for the age and gender group of the speaker (children = 10, women = 12, men = 14).
•Best match between the formant tracks and the spectrogram/spectrum.
•Correspondence of F_1–F_2–F_3 and expectation according to phonetic knowledge.

Results

Poles	Automatic selection approved			Automatic selection not approved, manual selection needed			
	Need of additional editing		Total	Need of additional editing		Total	
	no	yes		no	yes		
10	20	10	30	0	1	1	Children
12	16	2	18	0	2	2	
14	1	2	3	5	3	8	
Total (of 64)	37	14	51	5	6	11	
10	6	4	10	0	9	9	Women
12	32	7	39	0	2	2	
14	13	4	17	1	2	3	
Total (of 80)	51	15	66	1	13	14	
10	2	1	3	0	3	3	Men
12	30	5	35	1	1	2	
14	37	0	37	0	0	0	
Total (of 80)	69	6	75	1	4	5	

Table 1. The results obtained from the algorithm and visual check by a phonetician.

The results of the automatic selection of a number of poles and of the evaluation of the corresponding F-pattern are given in Table 1. The results indicate that

•If further editing is included, automatic selection is approved for 80% (children) to 94% (men) of the sounds.

•In a substantial number of cases, the selected and approved number of poles does not correspond to an "expected" age and gender related default setting.

Discussion

As the results show, for sounds of a given age and gender related speaker-group, the number of poles for the estimation of formant frequencies that match best with the respective spectrogram/spectrum varies greatly. **Therefore, this selection should be objectivized as much as possible.** The present approach presents an example of such an objectification.

Further improvement may be achieved by:
•Integrating objectivized phonetic knowledge in terms of a table of frequency ranges related to speaker groups and vowel categories
•Formulating rules for overruling the automatic selection
•Formulating rules for the manual editing process