Chapter 15

Spectral Study with Automatic Formant Extraction to Improve Non-native Pronunciation of English Vowels

R. Munoz-Luna, A. Jurado-Navas and
L. Taillefer de Haya

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/57221

1. Introduction

The purpose of this paper is to develop the frequency domain of the study started in [1]. In particular, we present an algorithm which obtains the first two formants ($F1$ and $F2$) of a vowel segment. These two elements are most often enough to disambiguate an English vowel, being crucial for non-native speakers' oral training. $F1$ and $F2$, corresponding to mouth opening and tongue position respectively, provide the necessary information for a proficient pronunciation. The phonological information rendered by $F1$ and $F2$ frequency contents produces an algorithm which can help non-native students of English in positioning their tongue and lips.

2. State of the art

The most widely cited experiment on vowel perception and acoustics is a simple one conducted at Bell Telephone Laboratories [2]. In that paper, authors recorded repetitions of ten vowels in /h V d/ context uttered by 33 men, 28 women, and 15 children. From these recordings, the first three formant frequencies ($F1 - F3$) as well as the fundamental frequency ($F0$) were extracted. Nevertheless, there was considerable formant frequency variability among participants, and formant frequency patterns overlapped substantially.

Formant frequencies have been already well-studied in both American and British English vowels [2–7]. On another note, remarkable numerical investigations were performed by Jan Awrejcewicz involving vocal cord oscillations and primary resonances [8, 9] and other particular effects as stability and bifurcation phenomena [10].

As far as phonology teaching is concerned, Pavón implemented a software programme [1] as a learning tool for his university students of English. One of Pavón's software applications is the fact that users can record a specific phoneme and compare it with an already existing

phoneme in his software programme. This sound comparison results in a graphical degree of similarity expressed as percentages, showing the resemblance between user and programme sound waves.

Nevertheless, as Pavón himself states, this is an approximate value and it depends on recording conditions (e.g. room noise and external variables), which make an indicative result. Although the idea is conceptually good, a frequency domain analysis is required in order to draw out the degree of resemblance between users' wave forms and those included in the system. On the one hand, software programmes do not distinguish between male and female voice recordings even though fundamental frequencies and formants are different in both cases. Women present peak energy in higher frequencies when talking, and Pavón's software only includes female recordings. On a different matter, time domain comparisons are not significant: results are very often meaningless.

For this reason, this paper attempts to improve the afore-mentioned software including a frequency domain analysis by means of fundamental frequency and $F1$, $F2$ identification. This would allow a more significant comparison between users' recordings and programme audio database. At the same time, depending on formant position, learners will receive information on mouth opening and tongue positioning according to each vowel sound. Consequently, we are making use of authors' previous research on audio signal processing [11], knowledge on communication channels [12], numerical methods [13], analytical modelling [14] and English applied linguistics [15]. This theoretical framework backs up a useful tool for students of English who want to autonomously improve their pronunciation of English vowels.

Finally, we are only focusing on vocalic sounds since not all human sounds offer well-defined formants. Vowels, on their part, do have distinct formants and their study complements oral language teaching, in this case, of the English language.

3. Organs of speech

Vowels are the result of glottal source, supraglottal tract and their filtering effects. Same quality vowels have similar spectral shapes, without regard to the source fundamental frequency (this is a variable that changes considerably depending on the speaker's age, sex and emotions). The air coming from the lungs supplies the necessary energy to produce sounds. Thanks to vocal cords vibration, the rate of air flow through the glottis generates a complex periodic wave. Glottal source waves and spectrum vary depending on the type of phonation. The differences in the waveform are due to the different amount of time that the vocal folds are open during a glottal cycle. Figure 1 shows the organs of speech in a cross-section:

The fundamental frequency, $F0$, also called the glottal frequency of the vocal fold vibration,S is dependent on several factors such as mass, length and tension of the folds which are interrelated in a fairly complicated way. These are typical values for $F0$ (during normal speech production, voicing frequency varies over an octave):

- adult male voice: 125 Hz.
- adult female voice: 220 Hz.
- child voice: 300 Hz

http://dx.doi.org/10.5772/57221

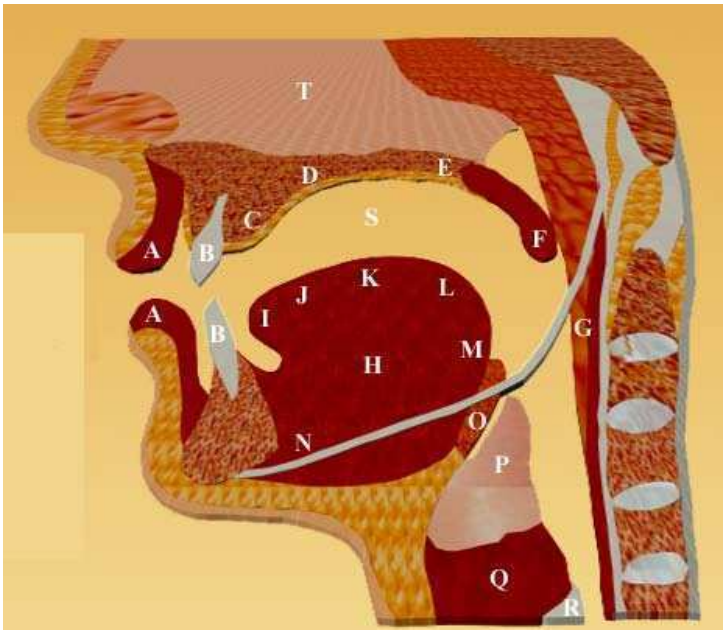

Figure 1. Organ of speech: A. Lips, B. Teeth, C. Teeth ridge, D. Hard palate, E. Soft palate, F. Uvula, G. Pharynx, H. Tongue body, I. Tongue tip, J. Blade, K. Tongue front, L. Back of the tongue, M. Tongue root, N. Jaw, O. Epiglottis, P. Thyroid cartilage, Q. Cricothyroid cartilage, R. Trachea, S. Oral cavity, T. Nasal cavity. Figure taken from [1].

Vocal tract filter selectively passes energy in the harmonics of the source. The size/shape of the vocal tract determines the amount of energy that is used in oral speech. For each vocalic sound, the so-called formants describe their characteristic resonance. In fact, the vocal tract transfer function for a particular vowel is defined by formant bandwidth and frequency. We can model the acoustic properties of the vocal tract as a tube open at one end, which is the mouth, and closed at the glottis. Assuming this tube uniformity, resonant frequencies can be calculated with the following formula:

$$F_n = \frac{(2n-1)c}{4L}, \tag{1}$$

where n is the number of the formant, c is the speed of sound, and L is the length of the tube. However, we also need to consider acoustic constrictions in the vocal tract. One way of modelling the acoustic properties of vowels is to represent the vocal tract as a concatenation of tubes [16]. An alternative approach is known as perturbation theory, which deals with vocalic acoustics in terms of relationship between air pressure and speed [17].

3.1. Formant frequencies of the vowels

First formant frequency ($F1$) is traditionally influenced by the shape of the vocal tract. $F1$ is inversely related to tongue height: low vowels have high $F1$ and high vowels have low

$F1$. On the other hand, second formant frequency ($F2$) corresponds to length and size of the speaker's oral cavity; in this case, front vowels have high $F2$ whereas back vowels have low $F2$; the formant frequencies decrease through the cardinal vowels, where the cardinal vowels can be consulted at [18]. Nevertheless, these relationships are not straightforward since there are other factors influencing sound production (e.g. lip rounding, tongue retroflexion, among others).

Articulatory properties of vowels are determined by these $F1$ and $F2$ formants in such a way that one is plotted against the other. Because of the inverse relationship between articulatory parameters and formant frequencies, zero frequency is at the top right corner. In Fig. 2 [1], we have displayed where English vowels are pronounced inside the oral cavity:

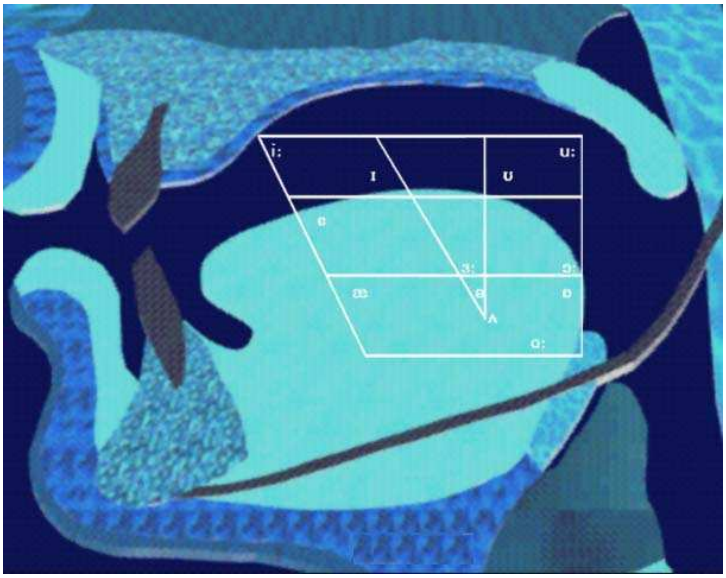

Figure 2. Vowel trapezium inserted in the oral cavity, indicating tongue movements for the pronunciation of the different vocalic phonemes [1].

4. Methodology

In this section, we present the algorithm implemented to find the frequency and amplitude of the first formants during any vowel-like segment. In order to analyse any speech fragment, a time-frequency analysis is needed. Short-time Fourier transforms (STFT), constant-Q [19] and wavelet transforms are some of the most commonly employed solutions in several systems. In this paper, the main idea is based on a previous work [20], tested on a large number of utterances produced by several different speakers; McCandless's discovery was found to be extremely successful. This algorithm is combined with some other ideas already developed by authors [11] in the context of polyphonic piano recordings.

We should remark that this manuscript comes to complement the work initiated in [1], so the recordings accepted by our system consists of only one vowel each, unlike the one presented in [20]. The latter developed a completely automatic algorithm which was meant to yield

http://dx.doi.org/10.5772/57221

the first three formants during all voiced sounds in continuous unrestricted speech. For this reason, the algorithm developed in this paper can be implemented more easily and productively.

4.1. Data acquisition and preprocessing

This stage consists in the recording of a vowel file. The audio data was kept in a WAV file at a sample rate of 44.1 kHz. The system accepts a monaural file as well as a stereophonic one. Then, the digitized signal is low-pass filtered in order to eliminate high frequency components.

4.2. Onset detection and temporal segmentation: windowing

As in [11], our system divides the vowel-segment into temporal slots and, afterwards, a frequency analysis of each slot is done. This temporal segmentation is based on the detection of onsets, so the system is prepared for detecting when a phoneme starts in the recording. This information makes it possible to discard frames whose total spectral energy is below a threshold for silence, and that must not be processed by the system.

After that, a Hamming window [13, Eq. 56] is applied to the segmented signal so that the extreme samples of the segments had less weight that the central samples. In this paper, we use a M-points Hamming window symmetric about the point $M/2$ of the form

$$w[n] = \begin{cases} 0.54 - 0.46\cos(2\pi n/M), & 0 \leq n \leq M, \\ 0, & \text{otherwise;} \end{cases} \tag{2}$$

owing to it is optimized to minimize the maximum (nearest) side lobe.

4.3. Sliding window procedure

A sliding window procedure [11] is employed to detect any increases in energy that exceed a certain threshold. This threshold has been selected to characterize the appearance of an onset. Rectangular windows that contain 4096 samples ($\approx$ 92.8 ms) of the signal to analyze are employed. The number of samples is chosen to be a power of two so that a fast Fourier algorithm can be employed to compute all values of the discrete Fourier transforms (DFTs) when performing a frequency analysis of the vowel-segment. Thus, the number of arithmetical operations required will be substantially reduced. Moreover, the character quasi-periodic and quasi-stationary of speech in that interval is seen as an additional justification for the size of these blocks, and will be of great utility in further upgrades of this system.

For any 4096 samples segmentation, a peak-picking method as the one employed in [20] was developed to extract formants. The justification of having the recording divided into frames of 92.8 ms is to detect such formants easily. Peaks can appear and dissapear from one frame to the next one due to resonances in the vocal tract and due to nasalizations, and the segmentation of the recording in frames of 4096 samples allows to successfully detect formants despite the mentioned fact of nasalizations.

In general, this latter effect presents a special problem because the nasalization is just a resonance of the nasal tract (it can be seen as a pole in the transfer function) whereas the oral tract is a closed side branch, which causes zeros (minimun energy in the spectrum). Frequently, the second formant, $F2$ is greatly reduced in amplitude, because of a nearby zero; and, in fact, often there is no peak corresponding to $F2$. In particular, the nasalization of a vowel is a problem of similar nature. In this case, the nasal cavity is an open side branch, causing extra zeros and extra poles. In a nasalized front vowel, typically, there is an extra small peak slightly above the first formant in frequency. In a nasalized back vowel, the apparent bandwidth of $F1$ becomes quite wide, because of a nearby zero, and sometimes there is no peak for $F1$. We will show this effect in the results included through this paper.

For each frame, a $N-$FFT is employed to compute all values of the DFTs. If the number of samples of the last frame is not a power of two, it is required to first zero-pad such a last frame previous to compute the FFT of the sequence [13]. As an interesting remark, for the computation of all N values of a DFT using the definition, the number of arithmetical operations required is approximately N^2, while the amount of computation is approximately proportional to $N \log_2 N$ for the same result to be computed by an FFT algorithm [21].

4.4. Processing of each frame: formant extraction

In this case, same steps as in [20] are developed. For each frame, fundamental frequency is first detected. Normally, it is always obtained as the peak with maximum energy. In our paper, all the tests were carried out by adult females, so fundamental frequencies were detected between 190-240 Hz in all cases, depending of the vowel produced by women. Tests have been restricted to women because the previous system implemented by Pavón in [1] was released with solely recordings of women. We must stress that, according to the signal processing problem, recordings obtained from women or recording produced by men are exactly the same problem, and the treatment and the way to solve both of them would be exactly the same.

Secondly, as in [11], we eliminate harmonics of fundamental frequency in each frame except, if we find a peak with higher energy placed in a potential harmonic. The constraint we impose in this step is that the amplitude of a peak placed in the frequency corresponding to the n-th harmonic must be lower than the amplitud of the peak positioned in the frequency corresponding to the $n-1$-th harmonic, with $n \leq 1$, where, in this notation, the 0-harmonic frequency is the fundamental frequency.

As a third step, our system fetches peaks finding the frequencies and amplitude of possible formants in the region from 150 to 3400 Hz. By executing this step in each 4096-sample frame, the system can detect peaks that appear, peak mergers as well as peak cancellations due to pernicious effect of the resonances and nasalizations commented above. Hence, we can take advantage of a very important feature in voice signals: the no continuity, i.e., how frequency formants can change from frame to frame, and new peaks can appear in a frame and disappear in the next one. A complete analysis of the voice signal without segmentation would entail, in many occasions, an error in the estimation of the formants, because some formants would not have enough energy to be detected.

After doing that, our system has selected some candidates to be the formants in each frame. One particular feature in the analysis of each frame is that the fundamental frequency is

http://dx.doi.org/10.5772/57221

moved to lower frequencies in subsequent frames. This fact let us obtain the bandwidth of this fundamental frequency for a most effective harmonic elimination. A final smoothing may be accomplished at each voiced frame in the same way as proposed by McCandless, to yield the formant tracks. The interpolated and smoothed values are valid if they are not too "different" from the original.

Finally, if any formant is not achieved, for instance, due to it has been merged with another one, an enhancement procedure is included using linear prediction analysis [20] employing, for this case, the linear prediction filter coefficients routine (*lpc.m*) included in MATLAB, based on an autocorrelation method of autoregressive (AR) modeling, as the one implemented in [12], to find the filter coefficients. Once the coefficients, a_k, are available, we can obtain in a straight manner the approximated spectrum by simply evaluating the magnitude of the transfer function, $H(f)$ of the filter represented by the coefficients a_k, at N equally spaced samples along the unit circle [20]:

$$H(f) = a_0 - \sum_{k=1}^{p} a_k \exp\left(-j2\pi nk/N\right) \tag{3}$$

For this purpose, the system can employ the function

```
filter([0, -a_k(2:end)], 1,xn);
```

as a previous step, where `a_k` are the coefficients, a_k, of the transfer function, `xn` is the original audio recording, and $n = 0, 1..., N-1$. As indicated in [20], two closely spaced formants frequently merge into one spectral peak, and cannot be resolved on the unit circle even with infinite resolution. However, they can often be separated by simply recomputing the spectrum on a circle of radius, r, less than 1. This amounts to reevaluating $H(f)$ at $x = r \exp j(2\pi n/N)$, r < 1. Because the contour comes in closer to the two poles, their peaks are enhanced, and a separation can be effected. Hence, by the estimated characters of linear prediction coding spectrum, in the region that the energy of signals is strong, i.e. the region closing to the peak value of the spectrum, the linear prediction coding spectrum is closing to the signal spectrum. However in the region that the energy of signals is weak, i.e. the region closing to the vale of the spectrum, both spectrums are significantly different. So to check the peak values of the linear prediction spectrum can confirm the formant.

5. Results and discussions

In this section, we are showing some results offered by the implemented system. As we have commented above, tests and recordings have been carried out in adult females, following the original system by [1], which was resealed in 2001. Nevertheless, we must remark that the signal processing problem would be identical in the case of males and children; the automatic formant extraction method would not change. After the process described in Section 3, , the system would have selected frequency peaks as candidate formants in each recording. The formant frequencies of vowels produced by males would be surely moved to lower frequencies in relation to the formant frequencies of vowels produced by women (see [4], for instance).

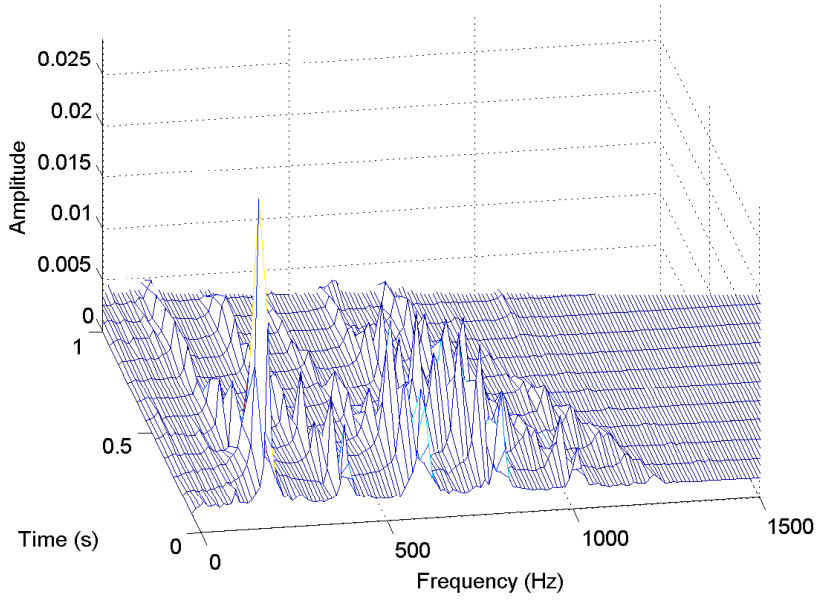

Figure 3. Time-varying spectral representation derived of a wrong-pronounced vowel number 5 = /a:/ by a woman of 29 years old.

In addition, the algorithm presented in this paper is based on the one by [20], which is effective in formant extraction during all vowel-like segments of continuous speech. In our particular case, voice recordings are even simpler, since they contain just a vowel sound, following the original system implemented by Pavón [1]. Our system compares users' recordings to those already included in its database, those latter which are the students' references in English learning. This algorithm will show users how to position their jaws and tongues for a correct vowel pronunciation by analysing formant frequency shift in vowels uttered by users in comparison to already-recorded model formants. This association comes with the relationship between $F1$, $F2$ and articulatory means. Consequently, there is a direct connection between first formant rising frequency and mouth opening: the higher $F1$ frequency is, the more open the vowel, and vice versa. Moreover, there is also a direct association between tongue backward movement and $F2$ frequency lowering: high $F2$ frequencies imply front vowels and vice versa. These conclusions can be verified in the results offered by [3, 4], especially in Table V in [4]). These authors confirm the correlation between first formant frequencies and vowel type (e.g. open, close, front and back).

As a significant result, we analyse a 29 year old female trying to pronounce vowel number 5 [18]. Initially, she does not position her mouth and tongue appropriately, being her mouth opening not wide enough. In addition, her tongue position is not so back as required. In Fig. 3 the temporal evolution of the spectrum derived when trying to pronounce the vowel number 5 [18] = /a:/ is displayed.

http://dx.doi.org/10.5772/57221

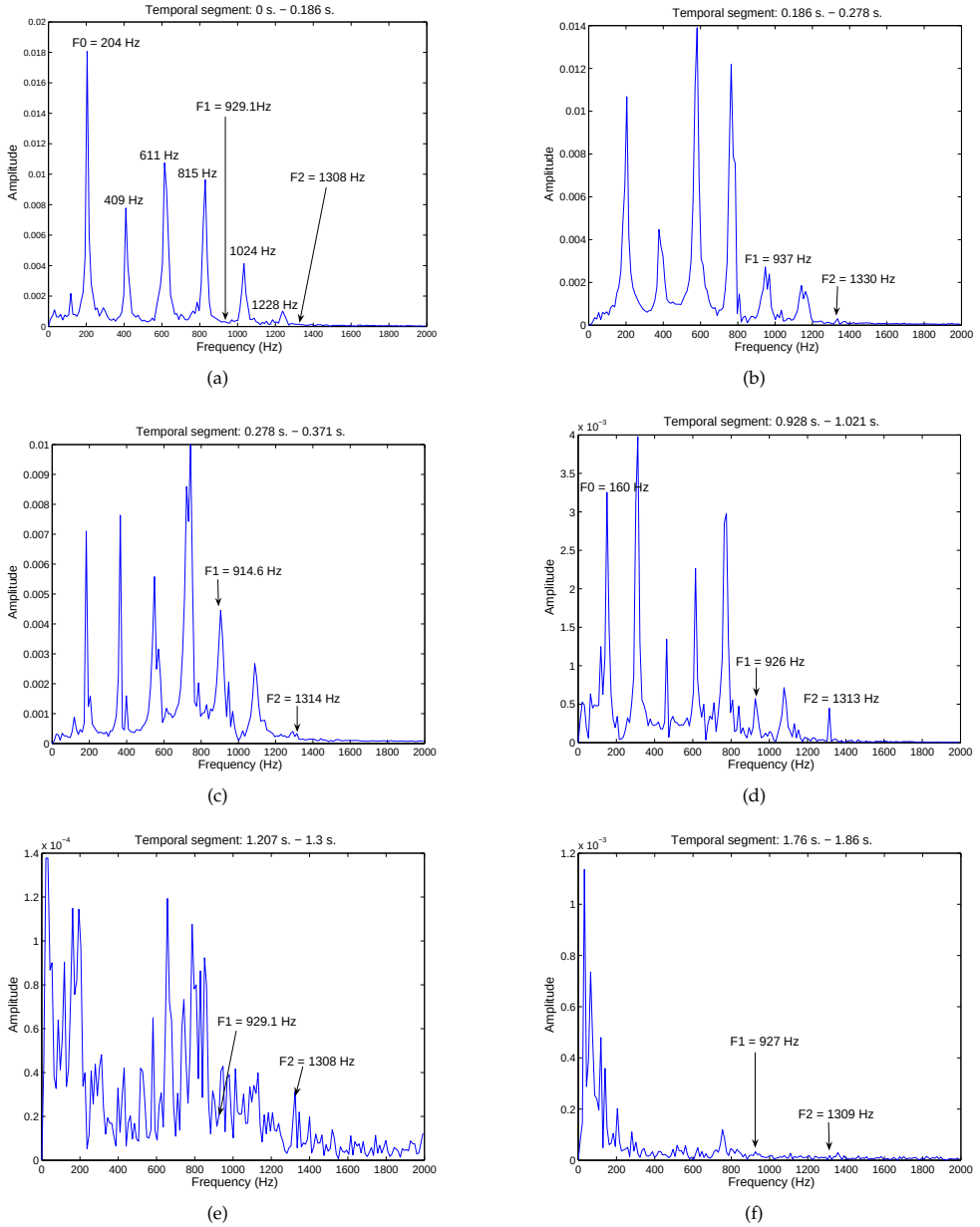

Figure 4. Spectra of different temporal segments after applying the sliding window procedure of a wrong-pronounced vocal number 5 = /a:/ by a woman of 29 years old. (a) 0 - 92.8 ms, (b) 0.186 - 0.278 s, (c) 0.278 - 0.371 s, (d) 0.928 - 1.021 s, (e) 1.207-1.3 s, (f) 1.76 - 1.86 s.

Now, in Fig. 4, we show some spectrums obtained from different temporal segments after applying the sliding window procedure detailed in previous section. As indicated above, any temporal segment is of approximately 92.8 ms. In particular, we are showing the following intervals: 0 - 92.8 ms (Fig. 4.a), 0.186 - 0.278 s (Fig. 4.b), 0.278 - 0.371 s (Fig. 4.c), 0.928 - 1.021 s (Fig. 4.d), 1.207-1.3 s (Fig. 4.e), 1.76 - 1.86 s (Fig. 4.f). We can clearly see the evolution of different peaks in the spectrum. Most of them are harmonics from the fundamental frequency ($F0 = 201.9Hz$). For instance, in Fig. 4.a, the fundamental frequency is placed in 204 Hz. Peaks at 409, 611, 815 1024 and 1228 Hz are considered the first five harmonics of $F0$. Through Fig. 4, we can see the evolution of the formants ($F1$ and $F2$) in each of the temporal segments. Even although these formants could have a low level of energy (above all in the second formant), the system operates successfully, as can be observed in Fig. 5.b. There, the system concludes that, for this recording of a 29 year old woman, the fundamental frequency is detected at 201.9 Hz, whereas the first two formants are positioned at 929.1 Hz and 1308 Hz, respectively. A peak at 872 Hz was also present, but it was discarded by the system after checking it is a harmonic of the 175.3 Hz-spurious peak. Finally, the spectrum of the whole recording (2.64 s in time, after detecting the onset of the vowel and rejecting the samples before the onset) is included in Fig. 5.a.

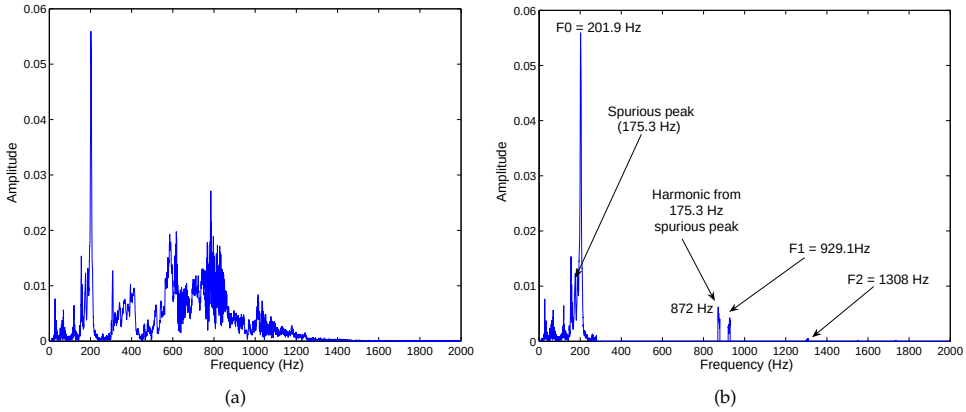

Figure 5. Spectrum of the whole recording (a), and amplitudes of fundamental frequency and frequencies of the two first formants (b).

At this stage, our system compares formant positions coming from this female recording to original recordings in [1]. According to Pavón, formants are placed at 940 and 1540 Hz, respectively. Therefore, this female subject has not achieved the correct articulatory mode or articulatory point. More specifically, her mouth is closer than required, and that is why $F1$ appears moved leftwards, from 940 Hz to 926 Hz. If $F1$ frequency had been higher, we would have had a too wide mouth opening. On the other hand, the articulatory point is not correct either: $F2$ appears at 1306 Hz, which is a much lower frequency than the 1540 Hz indicated in [1]. In this case, the subject has uttered vowel number 5 with a too backwards tongue position, while the system suggests a more central one. On the contrary, if her tongue had been more fronted, $F2$ could be detected in frequencies higher than 1540 Hz. The evolution of the first formant frequencies for each English vowel appears in Table V in [4], for American English vowels, and in [3], for British English.

http://dx.doi.org/10.5772/57221

Thanks to the corrections suggested by our system, the subject uttered vowel number 5 again, with the result shown in Fig. 6. As in the previous case, we are depicting in detail some time segments resulting from sliding window procedure described above. As indicated, any temporal segment is of approximately 92.8 ms. In this case, we are showing the following intervals: 0 - 92.8 ms (Fig. 7.a), 0.371 - 0.464 s (Fig. 7.b), 0.464 - 0.557 s (Fig. 7.c), 0.835 - 0.928 s (Fig. 7.d), 0.928 - 1.021 s (Fig. 7.e), 1.02 - 1.115 s (Fig. 7.f)

In this case, gesture corrections pointed out by our system allow the speaker to approach the target vowel sound. As we can see in Fig. 8.b, $F1$ and $F2$ are 934 and 1495 Hz, respectively, being $F1 = 940$ Hz and $F2 = 1540$ Hz the referential frequencies recorded in the system. Consequently, this new recording is closer to the adequate pronunciation range of vowel number 5. If we accept a +-5% error range, the speaker's new pronunciation can be considered correct since the system error calculation is the following:

$$\text{Error in F1 (\%)} = \frac{|934 - 940|}{940} = 0.6\%. \tag{4}$$

$$\text{Error in F2 (\%)} = \frac{|1495 - 1540|}{1540} = 2.92\%. \tag{5}$$

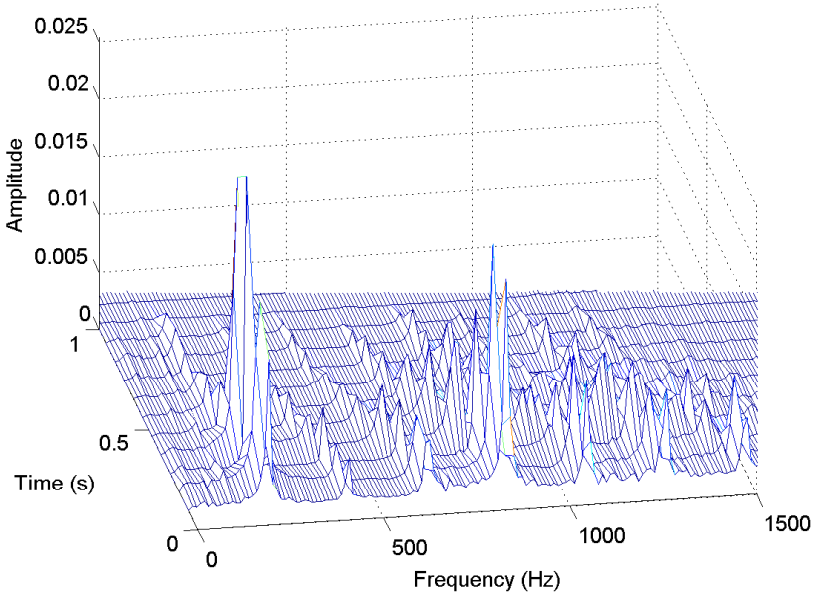

Figure 6. Time-varying spectral representation derived of vocal number 5 = /a:/.

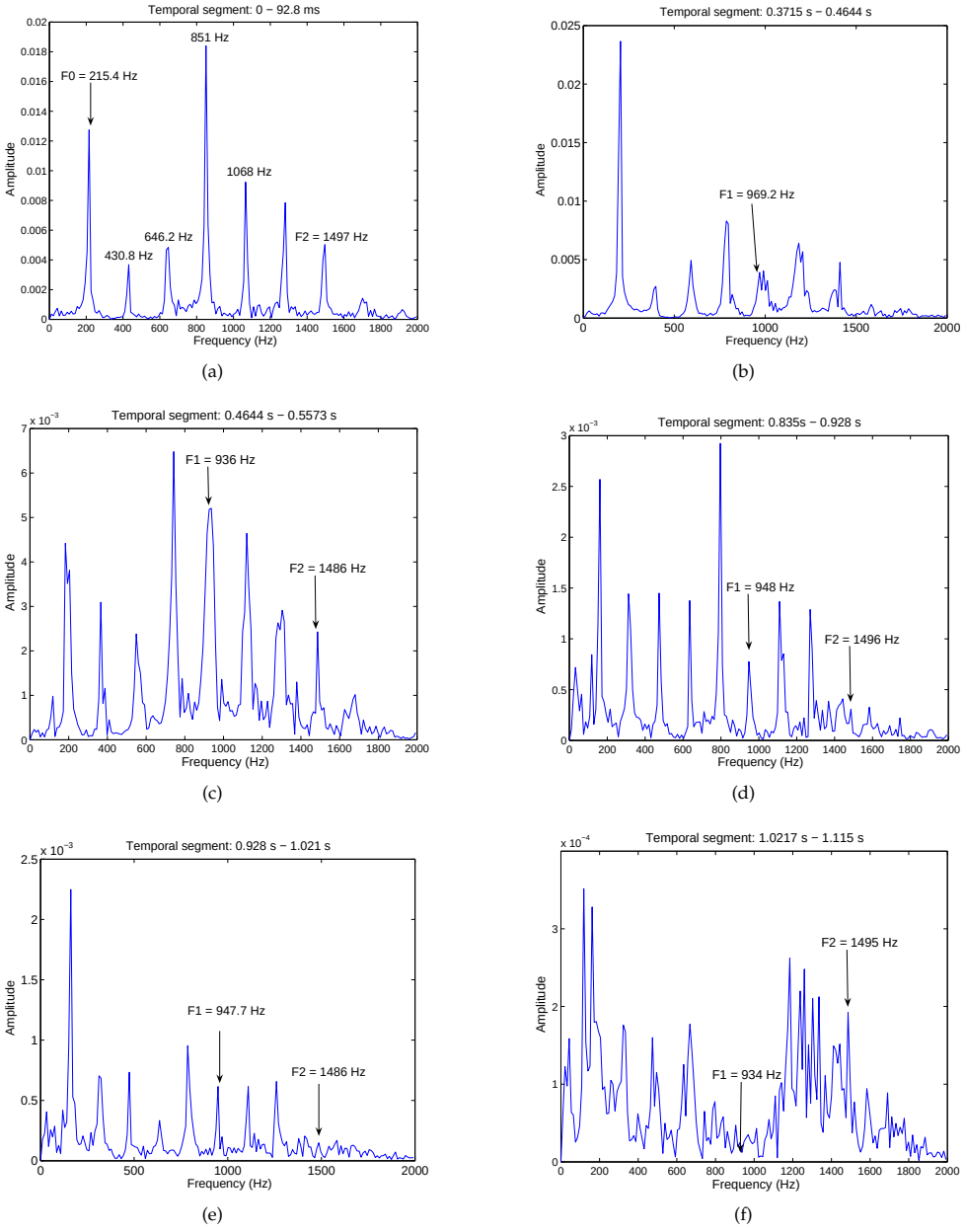

Figure 7. Spectra of different temporal segments after applying the sliding window procedure of a well-pronounced vocal number 5 = /aː/ by a woman of 29 years old.. (a) 0 - 92.8 ms, (b) 0.371 - 0.464 s, (c) 0.464 - 0.557 s, (d) 0.835 - 0.928 s, (e) 0.928 - 1.021 s, (f) 1.021 - 1.115 s

http://dx.doi.org/10.5772/57221

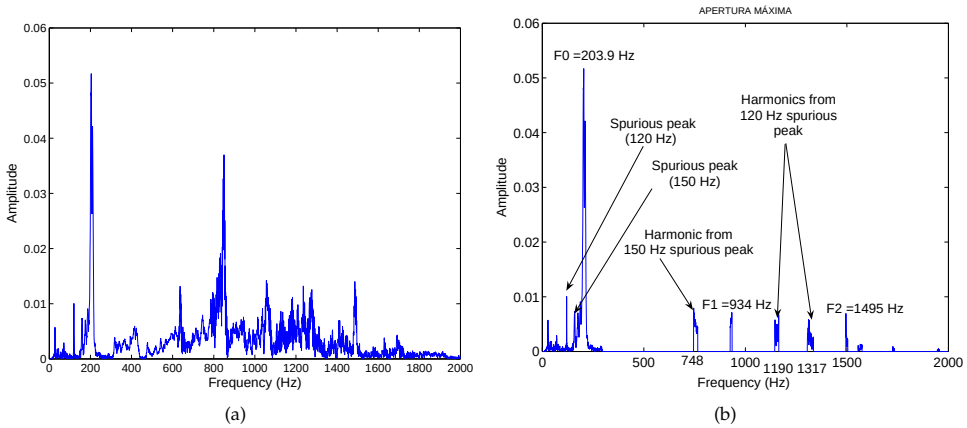

Figure 8. Spectrum of the whole recording (a), and amplitudes of fundamental frequency and frequencies of the two first formants (b).

With respect to vowel duration, our system does not pay attention to this feature because we understand that any user can distinguish a long duration with respect to a short duration of any vowel recording included in the system.

Finally, as in [20], the success of the automatic formant extraction algorithm is even higher than in McCandless's work because vowels are given to the system in an isolated manner and not in a sentence. Only when the formant was too strongly cancelled by a nearby zero (in nasals and nasalized vowels), or a peak merger was not resolve, the system does not achieve the correct result, but represent only a 10-15 percent of the total cases.

6. Concluding remarks

In this paper we have improved the tool implemented in [1], which consists in a software system for the teaching of English phonology. Pavón's contribution allows phoneme recordings, which are later on compared to similar sounds in the system. However, it offers a comparison based on the time domain, which is certainly not significant when providing help for learning a second language pronunciation. Moreover, it includes female voice recordings only, so male users (and children) would not obtain a significant result. Taking into account that Pavón's original idea is very good for those students who lack listening and pronunciation skills, this paper describes a new procedure to be added to the previous system and which is based on a frequency domain analysis. In this way, by means of a formant detection algorithm based on [20] and [11], the system can offer a more realistic contribution to the teaching of English pronunciation and phonology. F1 and F2 indicate oral cavity opening and tongue position respectively, and so the system specifies whether students have to open or close their mouths and which tongue part must be particularly employed in each vowel sound. As [1] makes use of female voice recordings only, our subjects are female adults. However, our formant detection algorithm would work with male and children voices equally. Male and children native' speakers are required for reference

in order to have their voices recorded and can be employed to appropriately compare with male and children non-native users of our system.

Acknowledgments

The authors are grateful for financial support from the Junta de Andalucía (research group "Communications Engineering (TIC-0102)").

Author details

R. Munoz-Luna[1], A. Jurado-Navas[2], and L. Taillefer de Haya[1]

1 Department of English, French and German Philologies. Faculty of Humanities. University of Málaga, Spain

2 Communications Engineering Department, University of Málaga, Spain

References

[1] Pavón, V. Sistema software para la contribución a la docencia de la fonética inglesa, v. 1.0, vocales y consonantes, 2001. [CD-ROM] Universidad de Córdoba, Servicio de Publicaciones, 2001.

[2] Peterson, G.E., Barney, H.L. Control methods used in a study of vowels. Journal of the Acoustical Society of America 1952; 24(2) 175–184.

[3] Deterding, D.. The formants of monophthong vowels in Standard Southern British English pronunciation. Journal of the International Phonetic Association 1997; 27(1-2) 47–55.

[4] Hilenbrand, J., Getty-M., L. A., Clark, J., Wheeler, K. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America 1995; 97(5) 3099–3111.

[5] Bauer, L. Tracing phonetic change in the received pronunciation of British English. Journal of Phonetics 1985; 13(3)61–81.

[6] Di Benedetto, M. G. Frequency and time variations of the first formant: properties relevant to the perception of vowel height. Journal of the Acoustical Society of America 1989; 86(1) 67–77.

[7] Hilenbrand, J., Gayvert, R. T. Vowel classification based on fundamental frequency and formant frequencies. Journal of Speech and Hearing Research 1993; 36(4) 694–700.

[8] Awrejcewicz, J. Bifurcation portrait of the human vocal cord oscillations. Journal of Sound and Vibration 1990; 136(1) 151–156.

[9] Awrejcewicz, J. Numerical analysis of the oscillations of human vocal cords. Nonlinear Dynamics 1991; 2(1) 35–52.

http://dx.doi.org/10.5772/57221

[10] Awrejcewicz, J. Numerical investigations of the constant and periodic motions of the human vocal cords including stability and bifurcation phenomena. Journal of Dynamics and Stability of Systems 1990; 5(1) 11–28.

[11] Barbancho-Pérez, I., Jurado-Navas, A., Barbancho-Pérez, A., Tardón, L. Transcription of piano recordings. Elsevier Applied Acoustics 2004; 65(12) 1261–1287.

[12] Jurado-Navas, A., Puerta-Notario, A. Generation of correlated scintillations on atmospheric optical communications. Journal of Optical Communications and Networking 2009; 1(5) 452–462.

[13] Jurado-Navas, A., Garrido-Balsells, J. M., Castillo-Vázquez, M., Puerta-Notario, A. A computationally efficient numerical simulation for generating atmospheric optical scintillations. In: Awrejcewicz J. (ed.) Numerical simulations of physical and engineering processes Rijeka: Intech; 2011. p. 157 - 180.

[14] Jurado-Navas, A., Garrido-Balsells, J. M., Paris, J. F., Castillo-Vázquez, M., Puerta-Notario, A. Impact of pointing errors on the performance of generalized atmospheric optical channels. Optics Express 2012; 20(11) 12550–12562.

[15] Taillefer de Haya, L., Silva Ros, M. T. New technologies in English Applied Linguistics. In: Proceedings of the 30th International AEDEAN Conference, Huelva, Ed. María Losada Friend et al. Huelva: U de Huelva, 2007.

[16] Fant, G. Acoustic Theory of Speech Production. The Hague: Mouton; 1960.

[17] Chiba, T., Kajiyama, M. The Vowel: Its Nature and Structure. Tokyo: Kaiseikan; 1941.

[18] IPA, The International Phonetic Association. http:www.langsci.ucl.ac.ukipa (accessed 20 May 2013).

[19] Brown, J. C. Calculation of a constant Q spectral transform. Journal of the Acoustical Society of America 1991; 89(1) 425–434.

[20] McCandless, S.S. An algorithm for automatic formant extraction using linear prediction spectra. IEEE Transactions on Acoustics, Speech, and Signal Processing 1974; 2(2) 135–141.

[21] Oppenheim, A.V. Discrete-Time Signal Processing. Upper Saddle River, New Jersey, USA: Prentice-Hall, 2nd edition; 1999.