VOICE QUALITY AND VARIATION IN ENGLISH

Marion Coadou and Abderrazak Rougab

Laboratoire Parole et Langage CNRS UMR 6057
Université de Provence, Aixen Provence
marion.coadou; Abderrazak.rougab@lpl.univ-aix.fr

ABSTRACT

This study is, to our knowledge, the first to compare the voice quality of several accents of the British Isles. Our hypothesis is that voice quality can vary according to the regional accent of the speaker.

The Long Term Average Spectrum (LTAS) was measured for each of the 50 speakers. Then, in order to test our hypothesis, a Principal Component Analysis (PCA) was carried out to compare the spectra.

The results showed that at least two accent groups could be isolated from the others. The spectra of the Belfast accent were particularly concentrated around the negative part of the first component. This can be explained by the fact that the Belfast accent is still strongly influenced by the Celtic languages spoken in the region.

Keywords: Voice quality, British accents, LTAS, PCA.

1. INTRODUCTION

The study of voice quality has always interested many researchers from different fields such as voice pathology (e.g. [8]), speaker recognition (e.g. [9]) or the emotional dimension of speech (e.g. [3], [5]).

Our definition of voice quality is based on the model of Laver, meaning that it is composed of both laryngeal and supralaryngeal settings ([7]).

Variations of voice quality according to social origins have also been observed in a few studies, for example [4] or more recently [11].

Yet, to our knowledge, no study has ever dealt with the voice quality variations across several accents of the British Isles. Our hypothesis is that voice quality can vary according to the regional accent of the speakers. In order to measure acoustically the variation of the voice quality, the long term spectral analysis (LTAS) was used in this study.

2. CORPUS

The corpus used for this study was extracted from The Intonational Variation in English corpus (IviE) created by Grabe, E., Low, L. and Nolan, F. in 1997. The advantages of this corpus are that several variables such as age or social class background are controlled through the method used for the recordings, which were all made in comprehensive schools with 16-year-old pupils. In order to neutralize any variation due to the content and style of the recordings, the text reading was chosen. Another advantage of the use of the text is that all the recordings last approximately 50 seconds, which is essential for the LTAS analysis. The sampling frequency of 16 KHz is also common to all the recordings.

Concerning the selection of the accent, Cambridge was chosen to represent a form of Standard English, although we do not confuse this accent and Received Pronunciation. Liverpool and Newcastle accents were chosen to represent the North of England and the accents of Cardiff and Belfast were chosen because they are both strongly influenced by Celtic languages.

Finally, the corpus is composed of ten speakers (5 male and 5 female) per accent.

3. SCIENTIFIC BACKGROUND

3.1. General background

Many studies have used the LTAS in order to measure voice quality in different areas of research. This is probably due to the fact that the main characteristic of voice quality is its “quasi-permanence” ([1]). Consequently, in order to observe voice quality acoustically, an analysis on a long-term basis is necessary. Note that stictly speaking, the LTAS does not only analyze voiced sounds but also voiceless ones.

Among these studies, [2] used the LTAS to measure the variation of voice quality between Castilian and Catalan dialects. The results showed

that the LTAS varied more between dialects than within one dialect. While this study successfully showed that the LTAS was a good tool to measure voice quality variation, this has not always been the case with other studies ([9], for a review). Indeed, the LTAS can be influenced by many different variables. For instance, some intra-speaker variation was observed in [6]. Considering those problems, the experimental method of our study had to be controlled very carefully, beginning with the method of calculating the LTAS.

3.2. Methods of calculating the LTAS

Spectral analysis provides a representation of the signal in the frequency domain. One of the major tools is the measurement of the power spectral density (PSD), defined as the Fourier transform of the self-correlation function of the signal, shown in the following equation.

(1) $$S(f) = \lim_{N \to \infty} E\left[\frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j2\pi fn} \right|^2 \right]$$

The mathematical expectation (E) means that an average of several realizations is calculated.

The method of Schuster was the first to propose the periodogram, an estimation of PSD. Later, a modified version of Schuster's periodogram integrated a weighting window. The first version of an average of spectra was proposed by Bartlett in 1948. He began by dividing the signal into several segments and then used Schuster's method for each segment. Finally, the method used here is that of Welch ([12]), which is a modification of Bartlett's periodogram introducing an overlap of the segments and a weighting window (see section 4).

4. METHODOLOGY

As mentioned in section 3, the method of Welch with an overlapping of 50% was selected for this study. The algorithm used to calculate the LTAS is shown in the following equation (2).

(2) $$\hat{S}_{Welch}(f) = \frac{1}{KLU} \sum_{i=0}^{k-1} \left| \sum_{n=0}^{L-1} \omega(n) x(n + iD) e^{-j2\pi fn} \right|$$

$$U = \frac{1}{N} \sum_{n=0}^{N-1} |\omega(n)|^2$$

The length of the window is of 512 samples, so that its duration is 32 ms.

The algorithm was implemented using Matlab. Before the LTAS analysis, the programme allows the user to choose between different types of windows and also different durations and then to specify the number of files to be treated.

Once the LTAS has been calculated, the programme provides a list of variables that can be used in order to sum up the different parameters or for the representation of the results.

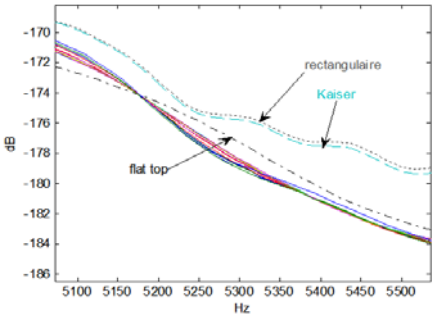
5. RESULTS AND DISCUSSION

5.1. Preliminary studies

Before analysing our corpus, two preliminary studies were made, one in order to test the type of window and the other to check whether the recordings of the corpus were long enough for the LTAS not to be influenced by the variation caused by the phonemic inventory of the different accents. Indeed, many studies have shown that the LTAS is influenced by the different phonemes of a language or a dialect if the recordings are not long enough (see [10] for a review). However, the required duration varies according to the different studies; which is why we wanted to test the effect of duration of the signal for our own corpus.

As mentioned in section 4, several types of weighting windows can be selected in our programme. The first preliminary study was, then, to observe the effect of the type of window on the LTAS. A comparison was made between sixteen different windows for one speaker.

Figure 1: Comparison between the sixteen types of window.



The results showed that three windows were particularly different from the others, namely the flat top, the Kaiser and the rectangular windows. We decided consequently not to use these for our study. From the others, we decided to choose the Hamming window, which was the preferred window for most previous studies.

As far as the duration of the signal is concerned, the LTAS was calculated on different durations for one male speaker of the Cambridge accent. The signal was divided into ten parts so that one tenth of the duration could be added each time. The LTAS was then calculated for all the durations that were obtained. Note that the final duration corresponds to the initial duration of the signal for the speaker.

We then compared the LTAS by means of the cross-correlation coefficient. The results showed that, for this speaker, the cross-correlation coefficient did not vary at all between 52 and 58 seconds, showing that the LTAS was no longer influenced by the duration of the signal. Note that, from 29 seconds the cross-correlation coefficient is above 0.999, a result of 1 indicating a total similarity.

5.2. Results of the LTAS analysis

The LTAS of the fifty speakers were measured for this study. However, only some examples of the results are shown in this presentation. The following figures (2 and 3) represent the LTAS of the speakers from Cambridge.

Figure 2: Representation of the LTAS of the female speakers of the Cambridge accent.

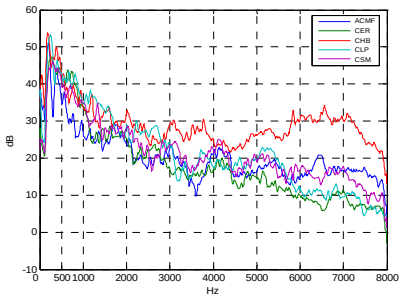
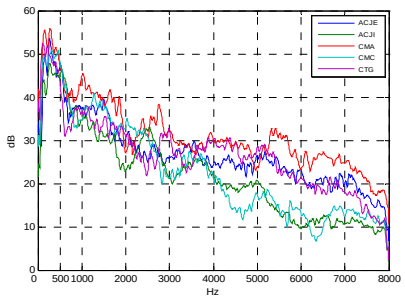


Figure 3: Representation of the LTAS of the male speakers of the Cambridge accent.



The LTAS of the female speakers (fig.2) are relatively coherent (apart from that of one female speaker) in particular from 4 kHz upwards. The LTAS of the male speakers (fig.3) are coherent in the lower frequencies and then separate from each other around 4 KHz. We next wanted to compare the results between the male and female speakers of the same accent. To do so, an average between the five LTAS of the male results was calculated. The same manipulation was done for the female results.

Figure 4: Comparison between the average LTAS of the male and female speakers of the Cambridge accent.

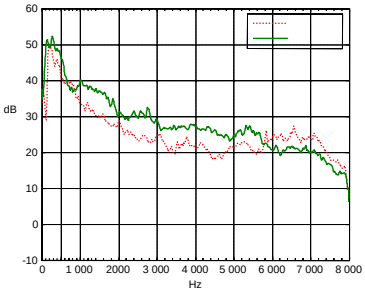


Figure 4 shows that only a few decibels (around 10 dB between 1000 Hz and 5500 Hz) separate the female speakers' LTAS (dotted line) from that of the male speakers. The small difference between the two LTAS can be explained by the fact that the speakers of the corpus are all 16 years old, which means that the recordings of the male speakers are from adolescents whose voices have only just broken.

The visual observation of the results is necessary but not sufficient to determine whether the LTAS form different groups according to the five accents. In order to do so, multidimensional scaling analysis seems an appropriate tool.

5.3. Principal component analysis

Principal Component Analysis (PCA) was used in this study in order to compare the results according to a different type of representation. The PCA is a mathematical technique that reduces the number of dimensions of a complex system of correlation. It represents the majority of the total variance of the data. The first principal component corresponds to the maximum variance that can be obtained out of the analysis, the second one being the maximum of the remaining variance etc. Figure 5 shows that the percentage of representation of the variance is greater than 70% for the first two components, meaning that a great majority of the

variance between the LTAS are represented with these two principal components.

Figure 5: Percentage of absorbed variance for the fifteen main axes.

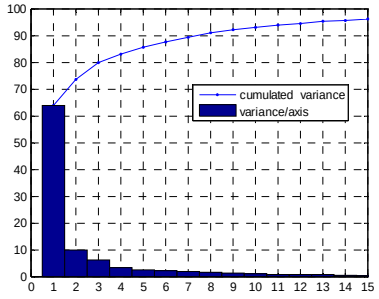
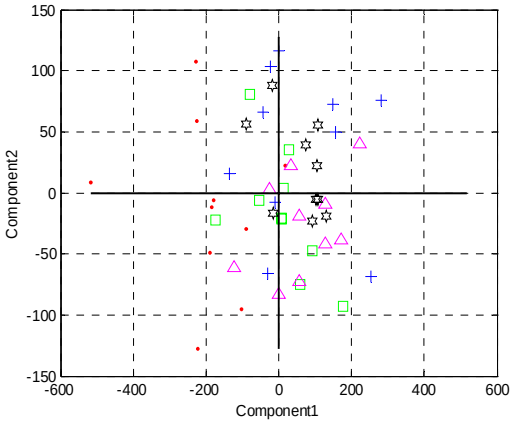


Figure 5 shows that the first component is the most efficient, given that around 65% of the variance is represented by the first axis.

The following figure is the representation of the results of the PCA analysis on the fifty speakers of the corpus.

Figure 6: Analysis of the five accents: representation of the first two components. (• Belfast, + Cambridge, □ Cardiff, △ Liverpool, ☆ Newcastle)



The distribution of the different symbols suggests that the LTAS form a more or less uniform group according to the accent. The group of the Belfast accent (red dots), tend to be particularly concentrated in the negative part of the first component. As far as the second component is concerned, this group is more spread out than the others. Another remark can be made about the speakers from Newcastle, which tend to be concentrated in the positive parts of the two axes. As far as the other groups are concerned, no tendency seems to appear, all speakers tend to concentrate around the two axes. Considering gender effect, no particular tendency was found.

6. CONCLUSION

To conclude, the results show that the PCA analysis was successful in discriminating two groups of accents, Belfast and Newcastle. Thus, thanks to the LTAS analysis, we demonstrated that our hypothesis is correct for those two accents.

Our short-term perspectives of research are to use other methods in order to compare the LTAS and see if the other accents can be isolated. On a long-term basis, it would be very interesting to use a much larger corpus, which would allow a comparison between age and gender groups as well as accent groups.

7. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh : Edinburgh University Press.

[2] Bruyninckx, M., Harmegnies, B., Llisterri,J. and Poch, D. 1994. Language-induced voice quality variability in bilinguals. Journal of Phonetics. 22: 19-31.

[3] Campbell, N. 2004. Accounting for voice-quality variation. In proceedings, Speech Prosody 2004, Nara, Japan.

[4] Esling,, J.H. 1978. Voice quality in Edinburgh: a sociolinguistic and phonetic study. Unpublished PhD dissertation, University of Edinburgh.

[5] Gobl, C. and Ni Chasaide, A. 1992. Acoustic characteristics of voice quality. *Speech Communication* 11, 481-490.

[6] Harmegnies, B. and Landercy. 1988. Intra-speaker variability of the long term speech spectrum. *Speech Communication 7,* 81-86.

[7] Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge University Press : Cambridge.

[8] Mackenzie Beck, J. 2005. Perceptual analysis of voice quality : the place of the Vocal Profile Analysis. . In Hardcastle W.J. and Mackenzie Beck J. (Eds.). *A Figure of Speech*. Lawrence Erlbaum Associates, Inc., Publishers: New Jersey. 285-322.

[9] Nolan F. 2005. Forensic speaker identification and the phonetic description of voice quality. In Hardcastle W.J. and Mackenzie Beck J. (Eds.). *A Figure of Speech*. Lawrence Erlbaum Associates, Inc., Publishers: New Jersey. 385-411.

[10] Pittam, J. 1987. The long term spectral measurement of voice quality as a social and personality marker : a review. In Tadjen, Kris. Jan-March, 1-12.

[11] Stuart-Smith, J. 1999. Glasgow: Accent and voice quality. in P. Foulkes and G. Doherty (eds), *Urban voices: accent studies in the British Isles*, (Edward Arnold: London, 1999), 201-220.

[12] Welch, P.D. 1967. The Use of Fast Fourier Transform for the Estimation of Power Spectra. *IEEE Trans. Audio Electroacoustics*, 15, 70-73.