Adaptation of wavelet transform analysis to the investigation of biological variations in speech signals

Julia M. Rees [a],*, Gavita Regunath [b], Sandra P. Whiteside [c],
Meghana B. Wadnerkar [c], Patricia E. Cowell [c]

[a] *Department of Applied Mathematics, Hicks Building, Hounsfield Road, University of Sheffield, Sheffield S3 7RH, UK*
[b] *Department of Chemical and Process Engineering, Newcastle Street, Sheffield S1 3JD, UK*
[c] *Department of Human Communication Sciences, University of Sheffield, Sheffield S10 2TA, UK*

Abstract

The purpose of this study was to adapt wavelet analysis as a tool for discriminating speech samples taken from healthy subjects across two biological states. Speech pressure waveforms were drawn from a study on effects of hormone fluctuations across the menstrual cycle on language functions. Speech samples from the vowel portion of the syllable 'pa', taken at the low- and high-hormone phases of the menstrual cycle, were extracted for analysis. Initial analysis applied Fourier transforms to examine the fundamental and formant frequencies. Wavelet analysis was used to investigate spectral differences at a more microbehavioural level. The key finding showed that wavelet coefficients for the fundamental frequency of speech samples taken from the high-hormone phase had larger amplitudes than those from the low-hormone phase. This study provided evidence for differences in speech across the menstrual cycle that affected the vowel portion of syllables. This evidence complements existing data on the temporal features of speech that characterise the consonant portion of syllables. Wavelet analysis provides a new tool for examination of behavioural differences in speech linked to hormonal variation.
© 2007 IPEM. Published by Elsevier Ltd. All rights reserved.

Keywords: Wavelet analysis; Ovarian hormones; Vowel production; Motor speech behaviour

1. Introduction

It has been found that temporal aspects of speech vary across the menstrual cycle [1]. Specifically, the micro-acoustic parameter, voice onset time (VOT), was different for specific plosive sounds (e.g., /p/ in 'pat' and /k/ in 'cat') in a comparison of women at phases of the menstrual cycle corresponding to low and high levels of circulating ovarian hormones. This work also documented differences between men and women that were consistent with the direction of the hormonal effects hypothesised to be operating within women. These findings have recently been replicated and expanded within our laboratory [2] and are in agreement with a larger body of work showing that ovarian hormones play a key role in the organisation and function of motor and motor speech behaviours. Specifically, high levels of estrogen have

been linked with improved performance on a range of motor and verbal tasks [2–4]. Results from our laboratory support the view that ovarian hormones affect the neurocognitive, neuromuscular and articulatory systems involved in speech production.

Traditionally, speech analysis of parameters such as VOT has been conducted within the domain of acoustic phonetics. In such studies, specific temporal features of the speech signal are extracted for analysis. Previous reports from our laboratory have explored hormone-mediated effects on VOT of plosive consonants [1,2]. In contrast, the aim of the current study was to adapt signal processing tools to investigate the source of hormone effects on another part of the speech signal, namely the vowel portion in the spoken CV syllable /pa/ ('pa'). Wavelet analysis is a specialist tool for deriving an in-depth understanding of underlying components of complex curvilinear data. It has been used to study waveforms in other domains of research such as meteorology [5], and has also been employed in the analysis of speech [6,7]. The objec-

* Corresponding author. Tel.: +44 114 2223782; fax: +44 114 2223739.
E-mail address: j.rees@shef.ac.uk (J.M. Rees).

tives of this study were to elaborate on the use of wavelet analysis on two-dimensional speech pressure waves, and to apply wavelet analysis to inform the study of differences in speech signals produced at hormonally distinct phases of the menstrual cycle.

2. The wavelet transform

One of the most well-known tools for signal processing is the Fourier transform. This breaks down a signal into its constituent sinusoids of different frequencies. One of the drawbacks of Fourier analysis is that in transforming to the frequency domain, time information is lost. For stationary signals, i.e., those that do not change much over time, this is not a serious drawback. However, for nonstationary data, spectral techniques that retain information from the time domain are more appropriate. Nonstationary waveforms are commonly encountered in experimental acoustics. In the 1940s, Nobel prize winner Dennis Gabor adapted the Fourier transform in order to analyse only a small section of the signal, known as the spectral window, at a time. This so-called short-time Fourier transform (STFT) maps a signal into a two-dimensional function of time and frequency, so one can determine approximately when an event of a particular frequency occurs. However, although STFT has proved useful in numerous applications, it is ill suited for the study of signals where the frequency content ranges over several orders of magnitude as the size of the window is the same for all frequencies. For signals such as these, which typically include sound waveforms, a more flexible approach is needed where one can vary the window size: longer windows are required to detect low-frequency content, though these will be insensitive to high-frequency sounds of short duration; shorter windows are needed to detect high-frequency components.

A further possible complication of using the STFT, or similar time–frequency responses, is that signals such as voiced speech and music can exhibit a strong periodic structure which can give rise to a characteristic 'bed-of-nails' structure in the spectrum. These evenly spaced spectral spikes are spurious, being an artefact of the transform kernel design. If the dominant frequency is known *a priori*, then it is possible to develop a pitch-dependent kernel to suppress such artefacts [8]. Use of the wavelet transform overcomes this potential problem.

Wavelet analysis is based on a windowing technique employing regions of varying size and can provide the time and frequency information simultaneously whilst overcoming the resolution limitations of the STFT. Long-time intervals are used to provide more precise information at low frequencies; shorter intervals are used to extract characteristics of high-frequency components of a signal. The wavelet transform is a powerful signal-processing tool that has been successfully used in the analysis of nonstationary data from a wide range of physical processes, for example, from engineering systems and in meteorology, that exhibit multiscale

features [5,9,10]. It has also been used extensively for speech compression and recognition [11,12] and for a number of studies on speech waveform analysis [6,7,13].

Whereas in Fourier analysis a signal is decomposed into sine waves of different frequencies, wavelet analysis involves the decomposition of a signal into shifted (i.e., translated) and scaled versions of the so-called "mother wavelet". Sinusoids are of infinite duration and are periodic. In contrast, the basis functions of the wavelet transform are of limited duration and tend to be irregular and asymmetric. There are many different types of mother wavelet. In general, the results of wavelet analysis are not crucially dependent on the exact type of mother wavelet chosen, but some wavelets are better suited to particular types of application. For example, the Mexican hat wavelet, which is popular in vision analysis, is not suitable for analysing speech as its flat frequency response results in low formant resolution. The Daubechies mother wavelet [14] is often chosen for the analysis of speech records [6]. For the present study the Daubechies wavelet of order 10 (db10) was used.

The continuous wavelet transform (CWT) of a signal is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the mother wavelet. Thus mathematically, the wavelet transform $C_{a,b}(t)$ of a signal $f(t)$, where t represents any independent variable (in this study t denotes time), is defined as:

$$C_{a,b}(t) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) \mathrm{d}t,$$

where the function $\psi(t)$ is a mother wavelet and the real numbers a ($a \neq 0$) and b denote the scaling and translation respectively. $C_{a,b}(t)$ are known as the wavelet coefficients. The constituent wavelets of the original signal can be found by taking the product of each coefficient with the appropriately scaled and shifted wavelet. This facilitates the detailed analysis of the signal over a range of frequencies. The scale a is related in a broad sense to frequency by:

$$F_a = \frac{\Delta t \times F_c}{a}, \tag{1}$$

where F_a is the pseudo-frequency in Hz corresponding to the scale a in Hz, F_c is the centre frequency in Hz of the mother wavelet and Δt is the sampling period. For the case of the db10 mother wavelet, the centre frequency is 0.68421 Hz.

3. Digitised speech data

3.1. Speech data collection

3.1.1. Participants

Data from three healthy women aged 20–25 years (mean age 20.39 ± 0.34 years) were selected for this study from a larger scale research project designed to examine hormones, speech and related behaviour. Participants were right-handed native English speakers who had regularly occurring men-

strual cycles (mean duration 29.67 ± 0.88 days) and were not using oral contraceptives, pregnant or lactating for at least 1 year prior to the study. All were free of health conditions that could adversely affect brain and behaviour. Data were collected with written informed consent of participants using a research protocol approved by the Department of Human Communication Sciences Research Ethics Committee.

3.1.2. Procedure

Testing took place across two test sessions for each participant using a within-subjects repeated measures design. Participants were tested across two phases of the menstrual cycle, once during the menstrual phase (Day 2–5 when ovarian hormone levels were low (menstrual low-EP phase); mean day 4.00 ± 0.58) and once during the midluteal phase (Day 18–25 when ovarian hormone levels were high (high-EP phase); mean day $19.33 \pm .88$). The midluteal phase was confirmed by counting backwards from the onset of the next menstrual cycle (mean 10.33 ± 1.67 days).

Test sessions included an interview to collect demographic information, family, hormonal, medical and educational history. Baseline testing for manual laterality and IQ (mean 119.33 ± 5.23) was conducted. The same speech tasks were administered at high EP and low EP. Administration of the first test session was counterbalanced, such that two of the subjects were first tested during the low-EP phase, and the remaining was first tested during the high-EP phase. The speech task consisted of speeded repetition of a single syllable 10 times as quickly and as accurately as possible. Four versions of the task were conducted so that 10 repetitions of the following consonant–vowel (CV) syllables were produced, /pa/, /ka/, /ba/ and /ga/. Speech data were recorded onto a digital recorder (Marantz Portable Professional Solid State Recorder, Model PMD670) using a sampling rate of 44100 Hz (file type PCM .wav file, stereo 16 bit). The digitised audio files were transferred into Adobe® Audition™ (version 1.5) for further editing and analysis.

3.2. Speech samples and signal pre-processing

In the current study, the aim was to characterise hormone effects on speech through detailed analysis of the vowel portion of CV syllables using wavelet analysis. The syllable /pa/ ('pa') was chosen for further analysis because it had already been shown to be particularly sensitive to VOT differences across the menstrual cycle [1,2].

Six digitised samples representing the speeded syllables repetition task for /pa/ ('pa') were selected for analysis. These samples represented data from three subjects at low EP and at high EP. The women whose digitised speech patterns were selected for this study are referred to as Subject 1, Subject 2 and Subject 3. Further to this, the third and eighth repetitions were selected from each of the speeded 10 repetitions for wavelet analysis. These samples were chosen in order to minimise 'warm up' and fatigue effects in the speeded syllables. This resulted in the wavelet analysis of 12 syllables

(third and eighth repetitions from three speakers at high-EP and at low-EP).

Each of the 12 digitised syllables was then visually inspected to edit the samples for wavelet analysis. The burst and aspiration phases of the plosive /p/ were removed at zero-crossings to leave the quasi-periodic phase of the CV syllable which represented the vowel 'a'.

4. Results of signal processing and wavelet analysis

Each of the 12 digitised data records had a total duration of 0.1–0.2 s. For consistency, the vowel cycles occurring between 0.03 and 0.06 s were selected for detailed wavelet analysis.

Fig. 1 shows the speech signal for Subject 1 for the third syllable repetition of /pa/ at high EP. The signal comprises an irregular wavelike pattern which has duration of approximately 0.09 s. The section selected for wavelet analysis has duration of 0.03 s and is indicated by the boxed region. This section of the signal is shown in more detail in Fig. 2. By choosing this section of the signal, the initial plosive /p/ was excluded to allow focused analysis on the more regularly repeating cycles of the vowel /a/. Latter portions of the syllable, where the amplitude of the wave packet decays, were also excluded.

The corresponding digitised speech signal for Subject 1 for low EP is shown in Fig. 3. Again, the section of the signal chosen for detailed analysis has duration of 0.03 s and is indicated by the boxed region. The same criteria were used in selecting this section of data as those used for the high-EP case. A close-up of this region of the data record is provided in Fig. 4. Both signals exhibit an underlying fundamental with a period of approximately 0.005 s, which corresponds to a frequency of 200 Hz.
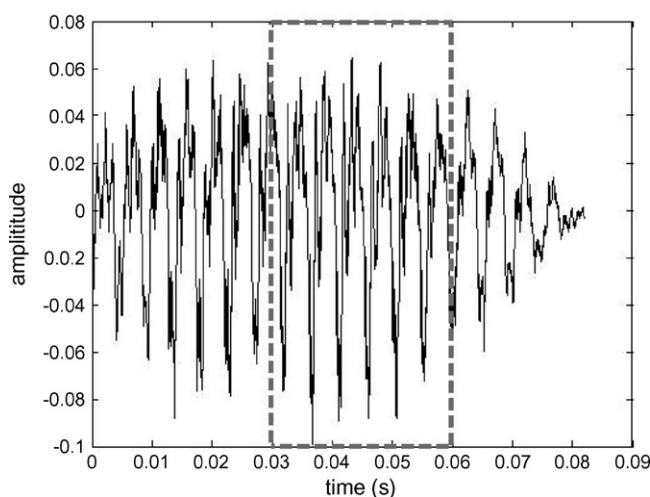

Fig. 1. Digitised speech signal for Subject 1, third syllable /pa/ repetition at high EP. The horizontal axis shows the time in seconds, and the vertical axis depicts a relative amplitude scale. The boxed region indicates the section of the data record used for detailed analysis.
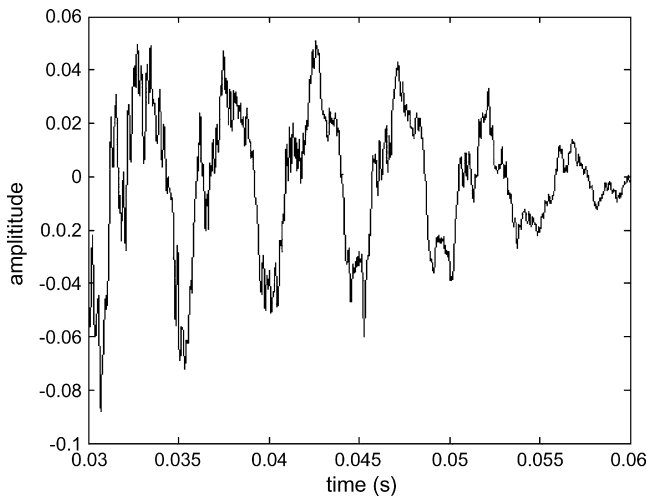
Fig. 2. Data section within boxed region of Fig. 1 expanded to in order to show fine details.

Each signal was then imported into Matlab® for the purpose of applying signal analysis routines. The aim was to extract information about signals' harmonic and wavelet structures, and related statistical properties. The first procedure applied in each case was the calculation of the Fourier spectrum. Although this technique has limitations for the study of nonstationary data, it was found to be useful in the current study for the determination of the fundamental frequency and the corresponding harmonic frequencies.

The Fourier spectrum for Subject 1 during the high-EP phase is shown in Fig. 5(a), and that for low-EP phase is depicted in Fig. 5(b).

For both cases, the fundamental frequency was close to 200 Hz, but the signal amplitude was approximately doubled
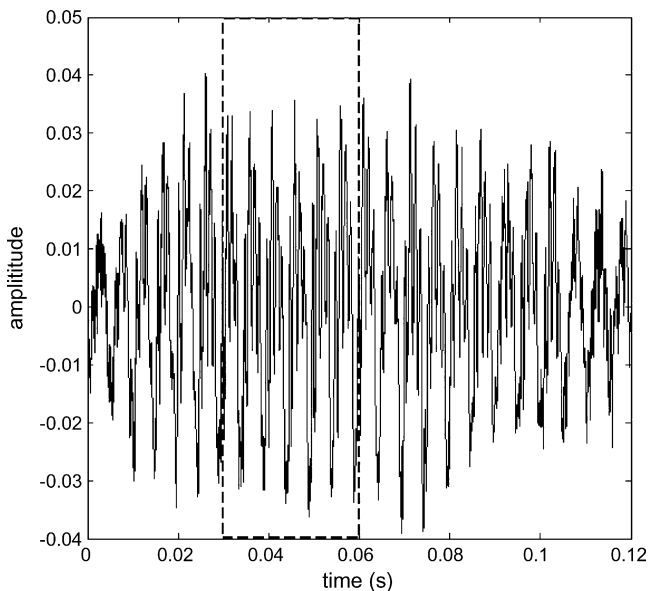

Fig. 3. Digitised speech signal for Subject 1, third syllable /pa/ repetition at low EP. The boxed region indicates the section of the data record used for detailed analysis.
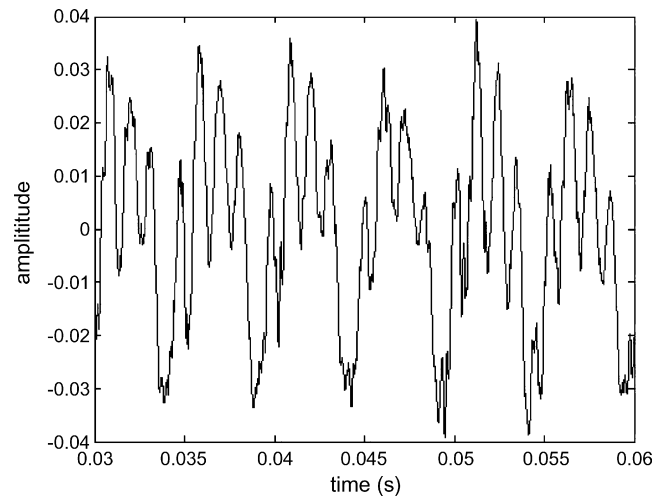

Fig. 4. Data section within boxed region of Fig. 3 expanded to in order to show fine details.

for the high-EP sample for Subject 1 (Fig. 5a). Since the fundamental frequency is determined by the rate of vibration of the vocal folds, it is therefore possible that the difference in hormone levels between low EP and high EP contributed to this effect. The harmonic frequencies occur at intervals of approximately 200 Hz (i.e., 400 Hz, 600 Hz, 800 Hz, etc.). However, these modes are poorly resolved by the Fourier decomposition. It should be mentioned that the Fourier spectra of some of the data records studied exhibited the spurious 'bed-of-nails' structure as described in Section 2. This is likely to be caused by the predominantly periodic structure of the vowel /a/.

In order to examine spectral differences on the microscale, we used the continuous wavelet transform (CWT) using the Matlab® wavelet toolbox. The Daubechies mother wavelet of order 10 (db10) was chosen for this study. The db10 mother wavelet is plotted in Fig. 6, together with the sinusoidal wave corresponding to the approximate central frequency of 0.68421 Hz.
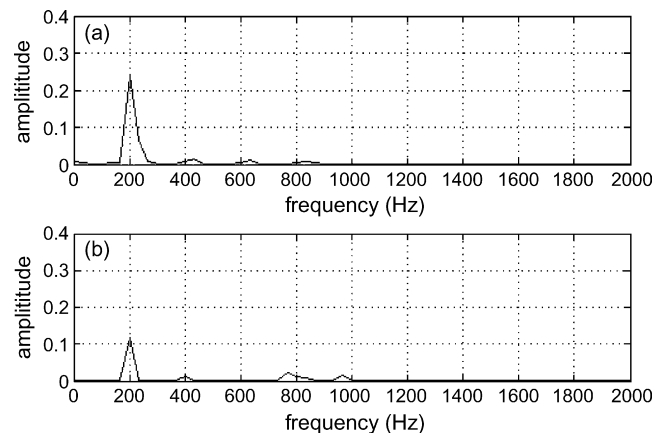

Fig. 5. (a) Fourier spectrum for the data shown in Fig. 2, i.e., from Subject 1 at high EP. (b) Fourier spectrum for the data shown in Fig. 4, i.e., from Subject 1 at low EP.
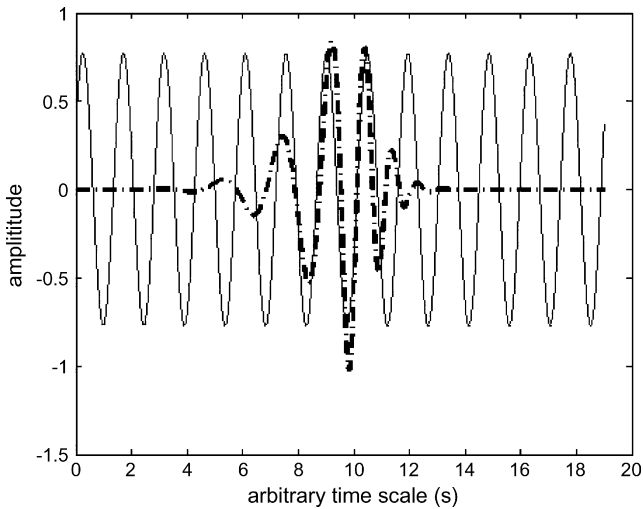
Fig. 6. The continuous line indicates the Daubechies mother wavelet of order 10 (db10). The dashed line shows the sinusoid corresponding to the centre frequency of 0.68421 Hz.

The relationship between wavelet scale and pseudo-frequency (see Eq. (1)), allowed the Fourier spectrum to be used as a guide for determining which wavelet frequencies should be examined further. Information about the frequency that revealed the maximum energy was available; thus, the scale used for the CWT analysis could be easily chosen. For example, from the relationship between scale and pseudo-frequency, a frequency of 200 Hz (which in this case corresponded to the fundamental frequency) will be associated with a wavelet scale of approximately 151 (Fig. 7).

The CWT for the high-EP case is shown in Fig. 8 and that for low EP is shown in Fig. 9. The lighter the image at a particular point in time–scale space, the more energy there is in that particular domain of the wavelet spectrum. The maximum wavelet amplitude scale for the high-EP case
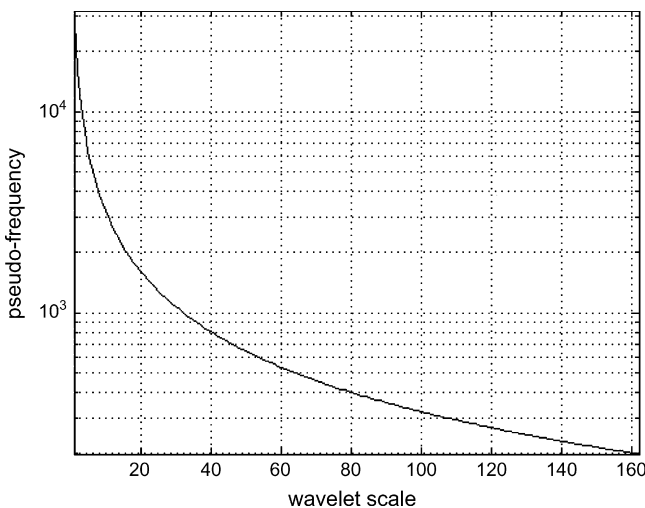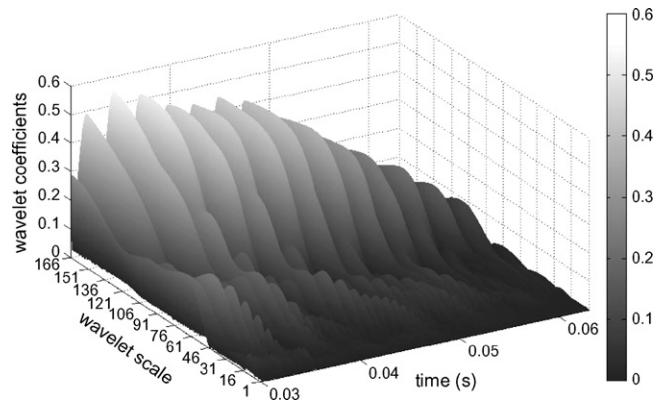

Fig. 8. The continuous wavelet transform for Subject 1 at high EP. The grey-scale key indicates the wavelet coefficients.

is approximately twice that for the low-EP case, which supports the evidence gained by the amplitude of the fundamental mode of the Fourier spectra shown in Fig. 5. In Fig. 9, across the lower wavelet scales, the spectrum is markedly lighter than that for the high-EP case shown in Fig. 8. This indicates that at low EP there was relatively more activity at smaller scales than at larger scales compared to the situation for high EP, where the fundamental frequency was more dominant.

We examined cross-sections of the wavelet spectrum for the fundamental frequency (Fig. 10) and the harmonics. For speech signals obtained during the high-EP phase, wavelet coefficients for the fundamental frequency exhibited consistently larger amplitudes than those for low EP. This effect is examined in more detail in Section 5 where quantitative differences are observed to evaluate whether this effect can be applied as a tool for investigating the influence of estrogen and progesterone on the spectral properties of speech waves. However, as can be seen in Fig. 5(b), typically the harmonic frequencies for low EP show no consistent trend. It is also worth mentioning that some harmonic frequencies have more energy during low EP than during high EP, although no consistent trend has been identified.

Fig. 7. Relationship between the wavelet scale and the pseudo-frequency.
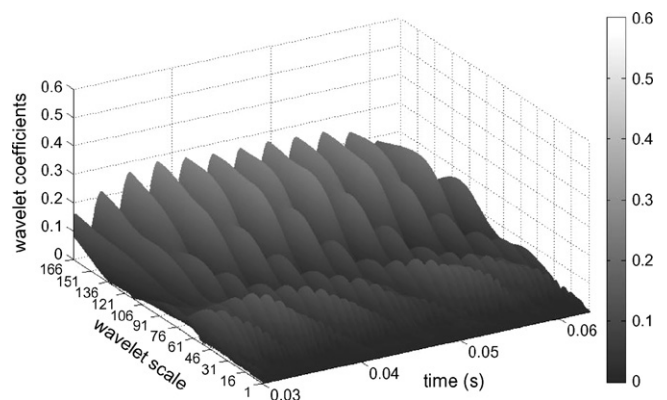

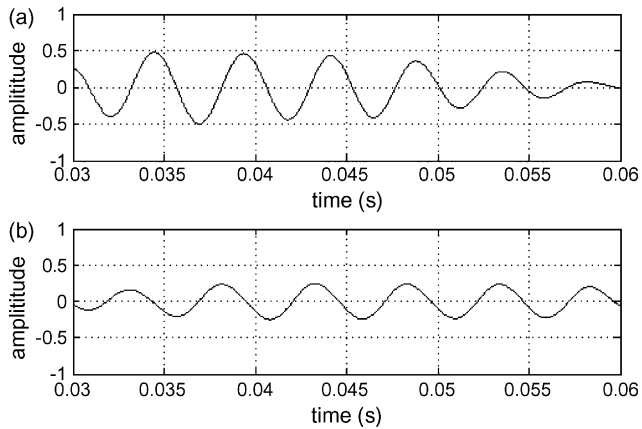Fig. 9. The continuous wavelet transform for Subject 1 at low EP.

Fig. 10. Cross-sections of the wavelet coefficients corresponding to the fundamental frequency of 200 Hz for Subject 1 for (a) high EP and (b) low EP.

Table 1
The variance of the wavelet coefficients for the fundamental frequency for each of the three subjects, for the third and eight vowel repetitions at both high EP and low EP

Subject	Vowel repetition	Variance at high EP	Variance at low EP	*F*-test statistic
1	3rd	0.0682	0.0241	1.5600
1	8th	0.0586	0.0143	42.5000
2	3rd	0.0165	0.0156	6.0625
2	8th	0.0189	0.0116	1.0714
3	3rd	0.0137	0.0052	3.2308
3	8th	0.0081	0.0029	5.8462

The *F*-test statistic calculated from the data in columns 3 and 4 is also shown.

5. Conclusions and discussion

For each of the 12 data records examined, the variances of the wavelet coefficients associated with the fundamental frequency, between times of 0.03 and 0.06 s, were calculated. Previous studies on digitised wave data from a variety of physical systems have successfully exploited the variance of the wavelet coefficients for distinguishing different regimes of the underlying processes [5,9].

For each of the three subjects, the variance at high EP was greater than the corresponding value for low EP. The one-tailed *F*-test can be used to make a statistical comparison between the variances of two data sets. The hypotheses for the *F*-test were:

- Ho: there is no difference between the two variances.
- Ha: larger variance s_1^2 is significantly different than the smaller variance s_2^2.

At the 5% confidence level, the critical *F*-test value is 1.0947 for our data sets. For five out the six cases presented in Table 1, the *F*-test statistic calculated from the data was clearly greater than the critical value. For the remaining case (i.e., Subject 2, eighth vowel repetition) the *F*-test statistic is close to the critical value. From these statistics, we conclude

that the wavelet variance of the fundamental mode at high EP is generally greater than that at low EP.

Typically it was found that the variance was lower for the eighth vowel repetition than it is for the third (the only exception being for the high-EP case for Subject 2). This feature can probably be accounted by a fatigue effect by the time the eighth repetition is reached.

The effects of hormones on vocalisation have been documented in a range of species, including humans [15]. However, the specific role of ovarian hormone fluctuations on the human voice across the menstrual cycle has received relatively little scientific attention. Using traditional methods of musical performance and voice analysis, it has been possible to study very specialised examples of voice from professional singers to detect changes related to ovarian hormones [16]. The development of wavelet analysis offers an alternative probe that is sensitive enough to measure changes in voicing from any healthy adult female in the context of speech production.

Conflict of interest

There are no conflicts of interest to report.

References

[1] Whiteside SP, Hanson A, Cowell PE. Hormones and temporal components of speech: sex differences and effects of menstrual cyclicity on speech. Neurosci Lett 2004;367:44–7.

[2] Wadnerkar MB, Cowell PE, Whiteside SP. Speech across the menstrual cycle: a replication and extension study. Neurosci Lett 2006;408: 21–4.

[3] Hampson E. Variations in sex-related cognitive abilities across the menstrual cycle. Brain Cogn 1990;14:26–43.

[4] Maki PM, Rich JB, Rosenbaum RS. Implicit memory varies across the menstrual cycle: estrogen effects in young women. Neuropsychologia 2002;40:518–29.

[5] Rees JM, Staszewski WJ, Winkler JR. Case study of a wave event in the stable atmospheric boundary layer overlying an Antarctic ice shelf using the orthogonal wavelet transform. Dyn Atmos Oceans 2001;34:245–61.

[6] Enders J, Geng W, Li P, Frazier MW, Scholl DJ. The shift-invariant discrete wavelet transform and application to speech waveform analysis. J Acoust Soc Am 2005;117(4):2122–33.

[7] Seok JW, Bae KS. Comparison of glottal closure instants obtained by using wavelet transform of speech signal and EGG signal. IEICE Trans Inform Syst 1999;E82D(11):1486–8.

[8] Kramer ML, Jones DL. A pitch dependent time-frequency response for periodic signals. In: IEEE international conference on acoustics, speech, and signal processing, conference proceedings, vols. 1–6. 1996. p. 2598–601.

[9] Staszewski WJ, Worden K. Wavelet analysis of time-series: coherent structures, chaos and noise. Int J Bifurcation Chaos 1999;9(3):455–71.

[10] Torrence C, Compo GP. A practical guide to wavelet analysis. Bull Am Meteor Soc 1998;79(1):61–78.

[11] Lung SY. Wavelet feature selection using fuzzy approach to text independent speaker recognition. IEICE Trans Fundam Electron Commun Comput Sci 2005;E88A(3):779–81.

[12] Srinivasan P, Jamieson LH. High-quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling. IEEE Trans Signal Process 1998;46(4):1085–93.

[13] Wang JF, Chen SH, Shyuu JS. Wavelet transforms for speech signal processing. J Chinese Inst Eng 1999;22(5):549–60.

[14] Daubechies I. Ten lectures on wavelets. Philadelphia, U.S.A: I.A.M; 1992.

[15] Rhodes HJ, Yu HJ, Yamaguchi A. Xenopus vocalizations are controlled by a sexually differentiated hindbrain pattern generator. J Neurosci 2007;27(6):1485–97.

[16] La FMB, Ledger WL, Davidson JW, Howard DM, Jones GL. The effects of a third generation oral contraceptive pill on the classical singing voice. J Voice 2007;21(6):754–61.