Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Electroglottography based techniques in the analysis of age related changes in the adult male voice

Stephen Dirk Bier

Abstract

Ageing produces audible changes in voice quality. This thesis aims to describe the creation of new methodologies for measuring changes in voice quality and to develop them in the context of analysing how ageing affects voice quality.

Electroglottography (EGG) is a method of examining vocal fold behaviour that uses the changing electrical resistance across the larynx during voiced speech to determine the extent of contact between the vocal folds. Electroglottographic recordings were made of 30 male speakers divided evenly into two age categories; young (20-26 years) and old (56-71 years). Each speaker was recorded reading the Rainbow Passage, producing sustained vowels at varying target pitch and loudness levels, and reading hVd word lists.

In order to quantify differences in the dynamic behaviour of the vocal folds between the two age groups, algorithms were developed to find cycle by cycle values for both the fundamental frequency and contact quotient from the electroglottographic waveform. For the sustained vowels, this allowed calculation of the perturbation of both the fundamental frequency and contact quotient. For the Rainbow Passage and hVd words, the discrete cosine transform was used to model the transient changes in fundamental frequency and contact quotient.

Age-related changes found in the measures derived from the electroglottographic waveform depended on the type of phonation. For all three types of elicitation, contact quotient was significantly lower for older speakers. The F0 was found to be higher for older speakers in the two reading tasks, but not significantly so in the case of sustained vowels.

The perturbation of the contact quotient in the sustained vowels was significantly higher for the older speakers, and proved a stronger indicator of age-related voice quality changes than the perturbation of the F0. The discrete cosine transform analysis of the fundamental frequency and contact quotient contours in the hVd vowels showed age dependent and vowel dependent differences.

Electroglottography proved to be a useful method of analysing vocal fold behaviour, and the results of this thesis have shown the perturbation of the contact quotient and contours of the contact quotient to each be effective for different types of voice elicitation.
Acknowledgements

I would like to express my special thanks to all those who helped me complete this work. To my supervisors, Catherine and Clare, who persevered with me even when I was struggling to be productive. To the MAONZE research group who provided funding and support throughout this research – in particular to Margaret, who was a willing ear in a time of need. To the University of Auckland Faculty of Engineering, who also provided funding to allow this research to happen. To the department of Speech Science, who allowed an engineer to come and work with them and supported me despite that.

To my colleagues in the ECE department, who provided support and friendship, and the experience of a common goal. To my friends, who helped keep me grounded. To Lyndon, who provided support and counsel in the final months of writing.

To my parents, Hank and Judy, who provided an unbelievable amount of support the whole way through. I couldn’t have reached this point without you. To my siblings, Peter, Miriam, and Jim, who have all walked the postgraduate path before me, and provided wise counsel throughout my own journey.

But thanks most of all to my wife Roseanne. You have been an unbelievable pillar of support throughout this PhD, and you are the only reason I have made it this far. Knowing that you were there for me, regardless of what happened, was instrumental to the completion of this thesis. I dedicate this work to you.
This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. Please include one copy of this form for each co-authored work. Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4 is an adaptation of an article published in the Journal of Voice.

Bier, S.D., Watson, C.J., & McCann, C.M. (2014). Using the perturbation of the contact quotient of the EGG waveform to analyse age differences in adult speech. Journal of Voice 28(3); 267-273

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>First author and primary contributor. The research material represents work done by the PhD candidate under the supervision of the co-authors.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>70%</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catherine I. Watson</td>
<td>Supervision: critique of the methodology and analysis, editorial support for the writing</td>
</tr>
<tr>
<td>Clare M. McCann</td>
<td>Supervision: critique of the methodology and analysis, editorial support for the writing</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephen D. Bier</td>
<td>S. Bier</td>
<td>7/6/2017</td>
</tr>
<tr>
<td>Catherine I. Watson</td>
<td>C. Watson</td>
<td>7/6/2017</td>
</tr>
<tr>
<td>Clare M. McCann</td>
<td>McCann</td>
<td>7/6/2017</td>
</tr>
</tbody>
</table>

Last updated: 19 October 2015
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5 is an adaptation of an article published in the journal Logopedics Phoniatrics Vocology.


<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>First author and primary contributor. The research material represents work done by the PhD candidate under the supervision of the co-authors.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>70%</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catherine I. Watson</td>
<td>Supervision: critique of the methodology and analysis, editorial support for the writing</td>
</tr>
<tr>
<td>Clare M. McCann</td>
<td>Supervision: critique of the methodology and analysis, editorial support for the writing</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephen D. Bier</td>
<td></td>
<td>7/6/2017</td>
</tr>
<tr>
<td>Catherine I. Watson</td>
<td></td>
<td>7/6/2017</td>
</tr>
<tr>
<td>Clare M. McCann</td>
<td></td>
<td>7/6/2017</td>
</tr>
</tbody>
</table>
# Table of Contents

Chapter 1. Introduction .......................................................................................................................... 1  
1.1. Motivations .................................................................................................................................. 1  
1.2. Aim ............................................................................................................................................... 2  
1.3. Primary developments ................................................................................................................. 2  
  1.3.1 Development of perturbation measures ............................................................................... 2  
  1.3.2 Development of dynamic measures of voice quality ............................................................. 2  
1.4. Thesis outline ............................................................................................................................... 3  

Chapter 2. Analysis of the ageing voice .................................................................................................. 4  
2.1. Source filter model of speech production ................................................................................... 4  
  2.1.1 Anatomy and physiology of the vocal folds during phonation .............................................. 5  
2.2. Measurement of voice quality and stability ................................................................................ 7  
  2.2.1 Acoustic measures ................................................................................................................. 8  
  2.2.2 Relating perceptual measures to acoustic measures .......................................................... 11  
  2.2.3 Relating perceptual measures to quantitative measures .................................................... 13  
2.3. Voice quality and stability in the ageing voice ........................................................................... 15  
  2.3.1 Physical changes in the vocal folds caused by age .............................................................. 15  
  2.3.2 Fundamental frequency (F0) ................................................................................................ 16  
  2.3.3 Intensity ............................................................................................................................... 17  
  2.3.4 EGG measures ...................................................................................................................... 18  
  2.3.5 Perceptual features .............................................................................................................. 18  
  2.3.6 Short term stability .............................................................................................................. 19  
2.4. Conclusions ................................................................................................................................ 20  

Chapter 3. Signal processing and measurement techniques in voice analysis ..................................... 22  
3.1. Common speech measurement techniques .............................................................................. 22  
  3.1.1 Acoustic recordings .............................................................................................................. 22  
  3.1.2 Laryngoscopy ....................................................................................................................... 23  
  3.1.3 Electroglottography ............................................................................................................. 24  
3.2. Signal processing techniques in the assessment of voice quality and stability ......................... 29  
  3.2.1 Fundamental frequency calculation ................................................................................. 30  
  3.2.2 Calculating the contact quotient ....................................................................................... 32  

List of Figures

Figure 2.1: Frequency domain representation of the source filter model .........................5
Figure 2.2: Coronal cross sectional view through the larynx, showing the mucosa, connective tissue and muscle layers that form the vocal folds Adapted from Colton, et al. (2006) ...............6
Figure 2.3: Cross Sectional View through a vocal fold, showing the three layers of the lamina propria in addition to the epithelium and thyroarytenoid muscles. Adapted from Colton, et al. (2006) .........................................................................................................................................7
Figure 2.4: Analysis of the sustained vowel /a/ in Praat ..................................................9
Figure 2.5: An example output from the Multidimensional Voice Program (sourced from http://www.clinicafoniatricaariasbotella.com/wp-content/uploads/2012/03/mdvp.gif on 20/10/2017). .............................................................................................................................10
Figure 3.1: EGG experimental setup. The electrodes and microphone are both attached to the microprocessor, which passes the recorded signals to the computer interface ...............26
Figure 3.2: EGG waveforms for (a) a young speaker and (b) an old speaker and acoustic waveforms for the same young speaker (c) and old speaker (d). All waveforms show a 30 ms segment of the vowel /i:/ at normal pitch and volume ............................................................ 27
Figure 3.3: Calculation of contact quotient from the EGG waveform .................................29
Figure 3.4: Use of interpolation to calculate threshold crossing times. Linear interpolation between the two samples on either side of the threshold allows a more accurate estimate of the time of threshold crossing to be estimated using Equation 3-1 ......................... 31
Figure 3.5: EGG waveforms showing the contact phase for (a) a young speaker and (b) an old speaker, both producing the vowel /i:/ at normal pitch and volume. The dashed line represents a threshold of 30% of the peak-to-peak amplitude, used in this instance to determine the start and end of the contact period .................................................................................................. 32
Figure 3.6: A pitch contour of a speech segment, showing time domain (top), spectrogram and pitch contour (bottom). Note that pitch contour is superimposed over the spectrogram, but they have separate y axis scales ............................................................................................................ 35
Figure 3.7: Normalised cosines corresponding to the first five DCT coefficients c1 through to c5 .............................................................................................................................................. 36
Figure 3.8: A reconstructed pitch contour using the first five DCT coefficients ...................37
Figure 3.9: Methodology for calculating F0 and Qx contours .........................................39
Figure 3.10: Rough period determination using zero crossings ......................................40
Figure 3.11: Calculation of thresholds for a given vocal fold cycle .....................................40
Figure 3.12: Using local threshold values to determine contact start and end points .............41
Figure 3.13: Calculating analytical measures from the F0 and Qx contours ......................42
Figure 4.1: Calculation of the periods and contact phases from the EGG waveform. The y axis shows the EGG magnitude. The dashed line represents the threshold of 30% of the peak to peak amplitude, which is used to determine the start and end of each contact phase. 

Figure 4.2: Target Pitch differences for sustained vowels, for (a) F0, (b) SD(F0), and (c) Qx. There were no visible differences for SD(Qx).

Figure 4.3: Target loudness differences for sustained vowels, for (a) F0, (b) SD(F0), and (c) Qx. There were no significant effects on SD(F0).

Figure 5.1: (a) F0 and (b) Qx for each vowel from the hVd words, for old (o) and young (y) speakers.

Figure 5.2: Phrase by phrase comparison of young and old speakers for (a) F0, (b) SD(F0), (c) Qx, and (d) SD(Qx). Each box represents the range of values across the 15 speakers in each age group, for a given phrase. The corresponding phrase for each code can be seen in Table 5.2.

Figure 5.3: Normalised F0 and Qx contours for hVd words, examined by age.

Figure 5.4: Normalised F0 and Qx contours for upward and downward sloped F0 samples of the syllable “rain.”

Figure 6.1: One vocal fold cycle from the young speaker producing the vowel /i:/ at normal pitch and low intensity. The images proceed in a clockwise manner.

Figure 6.2: One vocal fold cycle from the young speaker producing the vowel /i:/ at normal pitch and normal intensity. The images proceed in a clockwise manner.

Figure 6.3: One vocal fold cycle from the young speaker producing the vowel /i:/ at normal pitch and high intensity. The images proceed in a clockwise manner.

Figure 6.4: One vocal fold cycle from the young speaker producing the vowel /i:/ at high pitch and low intensity. The images proceed in a clockwise manner.

Figure 6.5: One vocal fold cycle from the young speaker producing the vowel /i:/ at high pitch and medium intensity. The images proceed in a clockwise manner.

Figure 6.6: One vocal fold cycle from the young speaker producing the vowel /i:/ at high pitch and high intensity. The images proceed in a clockwise manner.

Figure 6.7: One vocal fold cycle from the old speaker producing the vowel /i:/ at normal pitch and normal intensity. The images proceed in a clockwise manner.

Figure 6.8: Electroglottographic waveforms for target pitch and loudness levels examined with laryngoscopy.
List of Tables

Table 4.1: Significant results of sustained vowel parameters with age group.........................49
Table 4.2: Mean values of the Speech Studio parameters for each age group, with post-hoc analysis results. ..........................................................................................................................51
Table 4.3: Mean values of the Speech Studio parameters for each target loudness, with post-hoc analysis results. .................................................................................................................52
Table 4.4: Mean values of the measures calculated in R for each age group, with post-hoc analysis results. ..........................................................................................................................53
Table 4.5: Mean values of the measures calculated in R for each target loudness, with post-hoc analysis results. ..........................................................................................................................54
Table 5.1: Phonetic symbols for hVd words analysed..............................................................59
Table 5.2: ID codes for the seven phrases analysed in the continuous speech recordings .....59
Table 5.3: Mean values of the Speech Studio parameters for hVd words, with post-hoc analysis results. ..........................................................................................................................60
Table 5.4: Significant vowel differences in F0. Vowels are ordered from highest F0 to lowest. Significant differences are shaded. ..........................................................................................62
Table 5.5: Significant vowel differences in Qx. Vowels are ordered from highest Qx to lowest. Significant differences are shaded. ..........................................................................................62
Table 5.6: Comparing /a:/ and /i:/ between the hVd word analysis and the sustained vowel analysis. ....................................................................................................................................63
Table 5.7: Mean values of the Speech Studio parameters for analysed phrases, with post-hoc analysis results. ..........................................................................................................................64
Table 5.8: F0 DCT coefficient comparison between age groups.............................................69
Table 5.9: Qx DCT coefficient comparison between age groups.............................................69
Table 6.1: Estimated Qx from stroboscopic video analysis...................................................81
Table 6.2: Mean EGG derived Qx values for speakers in stroboscopic analysis. .................81
Chapter 1. Introduction

Voice quality is something that can affect everyday life. The sound of our voice is part of our identity, and it changes with age, as do many other aspects of our body. Just as understanding the health implications of physiological ageing on the rest of our body is important, understanding the effect of age on voice quality is important.

Alongside the changes present with normal ageing, there are a large number of voice disorders that can have significant effects on voice quality (Colton et al., 2006). When examining the voice of those who experience a voice disorder, normative data for the changes that occur with age can help us differentiate which voice quality aspects are the result of ageing and which aspects are due to the voice disorder itself.

Understanding the changes in voice quality that occur with age could also have an impact on speech recognition systems and forensic speech analysis, where it is important to be able to accurately estimate within subject variation.

1.1. Motivations

The motivation for this research comes from a few areas. For one, there is the wish to better understand how age affects our bodies, particularly as New Zealand has what is currently considered to be an ageing population.

A second aspect is a desire to contribute to wider understanding of measurement of voice quality. It is hoped that the techniques developed in this thesis, being applied to examination of age-related voice quality differences, will be able to be applied to measurements of voice quality in other contexts as well. With some extension, this research could also be applied to a clinical setting, for analysing and diagnosing speech pathologies.

There is the additional factor of local context for this research – the voices analysed in this thesis are all from speakers of New Zealand English. The data collected over the course of this PhD are able to provide some normative information for the New Zealand context, which sets it apart from a lot of other research into age-related changes in the voice. Voice quality differences that are dependent on vowel have the potential to differ between accents that have different vowel spaces.
1.2. Aim

This research aims to use signal processing techniques to create software algorithms to help with the analysis of voice quality, and to illustrate their potential use by examining age-related changes in voice quality. In particular, the algorithms developed will be applicable to electroglottographic recordings, and will show more information about voice quality than standard acoustic measures alone.

The algorithms developed also aim to address the dynamic nature of speech and illustrate the differences between static and dynamic measurement of voice quality.

1.3. Primary developments

This thesis develops two key methods of examining and quantifying voice quality, and its changes that occur with age. These methods primarily relate to the dynamic nature of speech.

1.3.1 Development of perturbation measures

Perturbation measures are a commonly used method for quantifying variance in voiced speech. The two most commonly used perturbation measures are jitter, which corresponds to fluctuations in pitch, and shimmer, which corresponds to fluctuations in loudness (Horii, 1982).

This thesis develops a third perturbation measure which measures fluctuation in the contact quotient, which is a measure of the period of contact within vocal fold vibration (Orlikoff, 1991). The contact quotient perturbation is calculated by examining electroglottographic waveforms, along with jitter. The contact quotient perturbation proved to demonstrate stronger differences between young and old speakers than jitter and shimmer, and as such, may prove to be a better tool for analysis of age-related changes in voice quality. This possibility is explored in this thesis.

1.3.2 Development of dynamic measures of voice quality

While the commonality of perturbation measures in the literature makes it a worthwhile topic to explore, it suffers from some inherent limitations. They are generally designed to deal with speech in a somewhat stationary state, in the form of sustained vowels. Phonemic and prosodic variations in continuous speech make it more difficult to apply perturbation measures. In addition to this, there is a question as to whether sustained vowels can really be considered speech.
Introduction

Therefore, this thesis investigates analysis of F0 and contact quotient contours, using the discrete cosine transform as a method of quantifying the shape of these contours. Comparing F0 and contact quotient contours across the 11 monophthongs of New Zealand English allows examination of some of the phonemic impacts on these measures, and how these impacts may vary with age. By looking at contours for a collection of instances of the diphthong /ei/ taken from a read passage we are able to see how these contours vary with different prosodic variations as well.

The discrete cosine transform provides a relatable way of quantifying the shape of these contours, and provides some convenient advantages for this analysis with inherent low pass filtering and time normalisation.

1.4. Thesis outline

Chapter two of this thesis describes the anatomy and physiology of the vocal folds during phonation; examines the current literature regarding measurement of age-related changes in voice quality and examines the methods of quantifying voice quality. Chapter three describes the different speech measurement techniques that have been developed and used in this research and provides an overview of the signal processing techniques used for this analysis.

Chapters four and five constitute the major contributions of the thesis. Chapter four develops the measurement of contact quotient perturbation based on the electroglottographic signal, and applies it to sustained vowels in young and old speakers. Chapter five then examines the differences between sustained vowels, word lists, and connected speech. It expands on the usage of the discrete cosine transform as a method of quantifying F0 and contact quotient contours calculated from the word lists and connected speech. Chapter six then discusses the considerations of using electroglottography or laryngoscopy in possible future work in this field. Chapter seven outlines some of the limitations of the current study and concludes the findings.
Chapter 2. Analysis of the ageing voice

Speech is produced by excitation of the vocal tract, which changes shape in order to produce the different sounds that make spoken language. The excitation can be in the form of turbulent noise for the unvoiced portions of speech, or, for the voiced portions of speech, by a glottal pulse generated by air flow through vibrating vocal folds, producing voiced speech. This is usually described in terms of the source filter model of voice production, with the excitation being the source and the vocal tract being the filter.

2.1. Source filter model of speech production

Speech can be modelled as a source filter system (Fant, 1960, Harrington and Cassidy, 1999), with the source being movement of air through the vocal folds (i.e. the glottal pulse) and the filter being the shape of the vocal tract. In order to create intelligible speech, the shape of the vocal tract is changed to alter the frequency spectrum of the output sound, creating the vowel and many of the consonant sounds that make up spoken language. For speech synthesis the modelled source then needs to be convolved with an appropriate vocal tract filter configuration.

It should be noted that the frequency response of the vocal tract/filter for vowel sounds has a series of peaks referred to as formants, which correspond to resonant frequencies of the vocal tract shape for a given segmental unit (such as a vowel). A stylised representation of the frequency domain for the source filter model in its simplest form can be seen in Figure 2.1. The glottal pulse section shows frequency content for a periodic signal produced by the vocal folds, which corresponds to the source. The vocal tract provides a filter with resonant peaks, which multiply the source in the frequency domain to give the resulting frequency content for the acoustic waveform. This can be calculated by convolving the source and the filter. In practice, lip radiation also needs to be taken into account (Fant, 1960, Wong et al., 1979).

The application of the source filter model of speech production is reliant on some assumptions (for example, see Rothenberg, 2008). The key assumption is that the vocal tract is a linear system. This has generally been found to be a safe assumption, which allows the source filter model to provide useful, meaningful results (Rothenberg, 2008).
Applying the source filter model has allowed use of linear predictive coding (LPC) to make estimates of glottal flow. This was initially developed by Rothenberg (1973), but has become one of the main approaches to estimating glottal flow. LPC estimates glottal flow by deconvolving the airflow volume velocity (as in Rothenberg 1973) or the acoustic signal with an estimate of the vocal tract filter. The most prominent method for this is using the LF model, named for its developers Liljencrants and Fant (Fant, Liljencrants, & Lin, 1985), which is particularly useful for estimating the magnitude of glottal flow. The LF model operates by estimating the derivative of glottal flow. However, it does include abrupt termination of the glottal flow as part of the waveform generated, which does not allow for situations in which there is not complete closure. Breathy phonation, for instance, can produce discontinuities in the middle of descending parts of the waveform.

This thesis primarily relates to vocal fold behaviour and its effects on speech, which is essentially an examination of the source within the context of the source filter model.

2.1.1 Anatomy and physiology of the vocal folds during phonation

A cross-sectional view of the larynx can be seen in Figure 2.2. The thyroid cartilage provides a support to hold the muscles and connective tissue that are used in voice production. The vocal folds themselves are made up of a mixture of muscle and elastic fibres which vibrate back and forth when driven by pressure from the lung.
Analysis of the ageing voice

![Diagram of the larynx](image)

**Figure 2.2:** Coronal cross sectional view through the larynx, showing the mucosa, connective tissue and muscle layers that form the vocal folds Adapted from Colton, et al. (2006).

Figure 2.3 (below) shows a cross-sectional view through a vocal fold. Beneath the epithelium is a layer of connective tissue called the lamina propria, which is divided into three layers. Underneath the lamina propria lie the muscles of the vocal folds. The vocal fold is sometimes described in terms of body (thyroarytenoid muscle and the deep and intermediate layers of the lamina propria) and cover (epithelium and superficial layer of the lamina propria) (Hirano et al., 1989).

The action of the vocal folds during phonation has been described many times but is detailed well by Colton et al. (2006). When adducted by the actions of the surrounding muscles, the vocal folds obstruct air flow. The subsequent air pressure then forces the vocal folds open. Whilst open, a combination of decreased air pressure in the glottis and elastic tension in the vocal folds causes them to shut. This cycle repeats itself to generate the glottal flow that provides the source for voiced speech. As the vocal folds vibrate, their surface exhibits a wavelike movement, referred to as the mucosal wave.
2.2. Measurement of voice quality and stability

There are audible differences in the speech of young and aged adults, and the differences between speakers in the voiced segments of their verbal output can be called differences in voice quality. Quantification of voice quality can be performed in a number of ways and several measures have been used in order to measure this (Buder, 2000). Measurements can be based on analytical measurement of voice signals or can be perceptual measurements based on listener responses. Analytical measures relate to many facets of the voice signal including F0, loudness, spectral content, and various stability measures. As Buder (2000) showed, the manner in which these things are measured varies as well. For instance, the short term perturbation measure jitter has several possible equations for calculation, and there are longer term measures that also measure the variability of F0.

Voice quality measures can utilise different signals for analysis. The most common and straightforward of these is the acoustic signal, which can yield information about loudness and pitch, typically represented by measures of sound pressure level and F0. Moving beyond acoustic analysis, more direct measures of physiological activity are also possible, such as electroglottography (EGG) and laryngoscopy. These provide more information about the movement of the vocal folds, along with measures of F0, and vocal fold vibratory behaviour. Vibratory behaviour is commonly described in terms of the open quotient (for example,
Winkler and Sendlmeier, 2006) or contact quotient henceforth referred to in this thesis as Qx (for example, Bier et al., 2014, and Orlikoff, 1991).

In this thesis, the aspects of voice quality that will be focused on will be those that relate to the vibration of the vocal folds. This includes F0, which is determined by the rate at which the vocal folds vibrate, and Qx, which provides information about the vibration itself. For both of these, the stability will be investigated. The methods used to analyse the stability of these features will be detailed in Chapter 3.

### 2.2.1 Acoustic measures

When examining voice quality in particular, it is common to use measures relating to the short term (cycle to cycle) variation of the voice (e.g. Baken, 2005, Torre, 2009, Ferrand, 2002, Awan, 2006 and Harnsberger et al., 2008). In this thesis the short term variability of the voice is taken as an indication of stability of the voice. Typical measurements of stability take the form of standard deviations or perturbation measures. Separating short term variation in voice features from longer term variation due to prosodic and phonetic variation can be difficult. A common way to avoid this problem is to examine sustained vowels where the speaker is producing a constant target pitch and target loudness such as in Baken, (2005), and Murry et al. (1995). When examining such a sustained vowel, both the standard deviation of F0 and the perturbation of F0 (jitter) provide an indication of a speaker’s ability to hold a stable pitch. Likewise, the standard deviation of sound pressure level (SPL) and the perturbation of SPL (shimmer) give an indication of the stability of loudness. Similar measurements can be made regarding the vibratory characteristics of the vocal fold behaviour, with measures of the standard deviation of the Qx or its perturbation (Bier et al. 2014).

These stability measures work well in the context of sustained vowels where the features of interest are held relatively constant for the duration of phonation. However, they become more difficult to implement or interpret in connected speech, where prosodic and phonemic effects need to be accounted for. Standard deviation measures cease to indicate stability around a target mean, and begin to encompass the prosodic and phonemic variance within the speech. Perturbation measures are often calculated via equations that do not account for transient behaviour in F0 or SPL as discussed by Yiu et al. (2000) and Bier et al. (2014). However, any measurements based on sustained vowels still need to be related to connected speech, as ultimately, this is the context in which listeners are able to perceive differences in voice quality.
Analysis of the ageing voice

Standard software packages that measure voice quality are available. Many studies use Praat (www.praat.org) to measure the acoustic features of F0, SPL, jitter and shimmer, for example Amir et al (2009), Maryn et al (2009), and Haldun et al (2011). An example of examination of a sustained vowel in Praat can be seen in Figure 2.4. The top half of the display shows the time domain waveform, with blue lines signifying pitch pulses. The bottom half of the display shows the spectrogram, with red dots indicating formants and the blue line showing the F0 contour. Acoustic features can be extracted for statistical analysis. In addition to allowing analysis of the speech waveform, Praat can be used to label wave files.

Praat has the advantage of being free software, allowing anyone to use it and making measurements comparable between various studies. There is, however, a danger in that measurements are easily applied without taking into consideration the methods and algorithms used in the back end, which may lead to some ill-advised interpretations of the validity of the results.

Another commonly applied software package is the Kay Pentax Multidimensional Voice Program (MDVP), which allows up to 33 different parameters to be extracted from its analysis (Kay Elemetrics, 1993). MDVP displays results in the form of a radial graph, an example of

Figure 2.4: Analysis of the sustained vowel /a/ in Praat.
which can be seen in Figure 2.5. Bhuta et al (2004) point out that MDVP is regarded as an objective and reliable measure of voice quality, though it also has its limitations. MDVP is primarily used for analysis of sustained vowels, as perturbation measures are usually reliant on this mode of voice (Yiu et al., 2000). Yiu et al. examined the ability of MDVP to distinguish dysphonic voices from normal voices. They found that only relative amplitude perturbation was sensitive to dysphonia. Therefore, while the MDVP manual states that connected speech is a viable speech source, Yiu et al. dispute this.

Nicastri et al. (2004) examined normative results for the MDVP with euphonic adult subjects. They focused on measures relating to amplitude variation, and found that their results were comparable to other centres also using MDVP (Ursino et al., 1999, De Colle & Schindler, 2001). Despite the comparability in results, they concluded that standardisation of analysis procedures is a matter of importance, and should be implemented sooner rather than later. In this particular case, all three studies used the same software package. There is greater potential for variation when different analysis packages are used.

There are a few studies that have compared Praat and MDVP, such as Amir et al. (2009), Maryn et al. (2009), and Haldun et al. (2011). Using analysis of sustained /a/ and /i/ vowels, Amir et al. found that both Praat and MDVP produced similar mean F0 values, but Praat produced significantly lower jitter, shimmer, noise to harmonic ratio, and the percentage of unvoiced
Analysis of the ageing voice

segments than MDVP. As a result, a group contrast between speakers with and without nodules and cysts produced a significant difference with the jitter values from MDVP, but this group contrast was lost with the jitter values from Praat. Similarly, Maryn et al (2009) found that perturbation measures calculated with Praat were significantly lower than those taken from MDVP. Their study compared the two programs using two recording systems and they concluded that frequency perturbation outcomes could hardly be compared across the two programs and systems, and that amplitude perturbation outcomes could hardly be compared across the two recording systems. Given the range of ways in which perturbation measures can be calculated (Buder 2000), this is not entirely surprising. Haldun et al. (2011) found that the two programs had comparable F0 and shimmer results, but the jitter and noise to harmonics ratio were significantly different between the two programs. Based on Maryn et al. and Haldun et al., it would seem that amplitude perturbation, or shimmer, is treated similarly by the algorithms in the two programs but requires matched recording systems for comparable analysis. Meanwhile, all three of these studies indicate that frequency perturbation or jitter comparisons across the two programs are of very limited use.

Bielamowicz et al. (1996) compared another three speech analysis systems; CSpeech, Computerised Speech Laboratory, and Soundscope. For these three software packages, F0 was comparable, but once again there was less agreement for jitter, shimmer and harmonics to noise ratio.

Overall, it is apparent that there are a large number of ways that acoustic recordings are being quantitatively measured in order to derive voice quality features, and there is some consensus about which parameters are worth investigating. However, beyond the most basic measurements of F0, there is often disparity between the methods and results for the various acoustic measures. As such, it would be prudent for any given study to make sure it is self-consistent in how each of these parameters is calculated, and care must be taken when making comparisons with results from other studies. It is also sensible to be explicit in describing how any measure that is analysed is calculated. To comply with this, the analysis in this thesis will be transparent in terms of explaining how each voice quality measure is calculated. These methods will be detailed in Chapter 3.

2.2.2 Relating perceptual measures to acoustic measures

While direct analysis of acoustic and physiological waveforms provides reliable measurement of voice quality differences, the reality is that they are being used to measure perceptual
Analysis of the ageing voice

differences in voice quality. These include voice characteristics such as hoarseness, coarseness, breathiness and roughness. By virtue of being perceptually measured features, there is likely to be some listener discrepancy in any attempt to quantify them. Blaustein and Bar (1983) demonstrated low interjudge reliability between three listeners assessing a group of 161 children for speech pathologies, despite the fact that each of the listeners in this instance were certified by the American Speech and Hearing Association, and were currently practising speech pathologists with at least two years’ experience.

One approach to quantising perceptual features of the voice is to use a standardised perceptual measurement, the most prominent of which in voice quality measurement is the GRBAS (Grade, Roughness, Breathiness, Asthenia, Strain) measurement protocol, introduced by Hirano (1981). GRBAS is primarily a clinically based measurement protocol, in which each of the five measures is graded on a four point scale, from 0-3, where ‘0’ corresponds to normal, ‘1’ to slight, ‘2’ to moderate, and ‘3’ to severe. Carding et al (2000) compared GRBAS to two of the other prominent perceptual protocols (The Vocal Profile Analysis and The Buffalo III Voice Profile) and concluded that GRBAS was the best of the three, and should be considered an absolute minimum standard for UK-based voice clinicians.

As De Bodt et al (1997) point out, there is potential for the background of listeners to have a profound impact on the recorded results of perceptual scales. GRBAS manages to keep intrajudge and interjudge variance low while maintaining high intervoice variance (Dejonckere et al, 1993), but as De Bodt et al. points out, this is at the cost of accuracy and precision of the scale. GRBAS also has an advantage in that there are standardised tapes demonstrating typical voice samples represented by the GRBAS scale (Hirano, 1981). This allows for standardised training of listeners, which can go a long way toward improving interjudge reliability. The diminished precision of the GRBAS scale renders it of limited use for the assessment of differences in normal voices. It is primarily targeted at the diagnosis of dysphonic voices. In the context of analysis of voice quality changes that occur with normal ageing, most analysed speech samples would be expected to fall at the lower end of the scale for all five of the GRBAS parameters.

Another method that can be used to accommodate listener discrepancy while maintaining more precision is to allow listeners to create their own scales for each measure, as performed by Gorham-Rowan and Laures-Gore, (2006). Their study allowed listeners to create their own scales for measures of hoarseness and breathiness, which were then compared to acoustic
measurements of SD(F0), shimmer, and noise to harmonic ratio. There were moderate correlations found between the perceptual measures and the acoustic measures in this instance.

While perceptual measures reflect the audible age-related changes in voice quality, for a full understanding of how these occur we need to be able to relate them to the quantifiable behavioural differences that can be measured from the voice production system. The changes in vocal fold behaviour that are discussed in this thesis will be contributing factor in the perceptual differences observed, even if the mechanism of how the vocal fold behaviour contributes is yet to be fully understood.

2.2.3 Relating perceptual measures to quantitative measures

Ultimately, the reliability and repeatability of quantitative acoustic measures makes it a preferable form of analysis, provided it can be related to the perceptual features we identify in voice quality. Gorham-Rowan and Laures-Gore (2006), discussed above, is one example of several relating a perceptual scale to direct acoustic measurements.

Bhuta, Patrick and Garnett (2004) examined potential correlation between the GRBAS scale and measurements taken with the MDVP. Nineteen of the measures provided by MDVP were analysed in this study. Looking at each of the GRBAS measures, Grade (which is a measure of the severity of voice abnormality) was found to correlate with three of the 19 measures from MDVP (voice turbulence index, noise to harmonic ratio, and soft phonation index). Roughness only correlated with one measure (noise to harmonic ratio), while Breathiness and Asthenia both only correlated with a single measure (soft phonation index). Bhuta et al. (2004) noted that the acoustic features that produced correlations were primarily associated with the noise, or non-periodic component of the speech signal, so perhaps noise is the marker of the dysphonic voice. They did not find any significant results for the oft reported jitter and shimmer in this particular study.

Bhuta et al. (2004) also pointed out that their results differed slightly from those of Dejonckere et al (1993), who used a different statistical analysis technique which produced a correlation between Grade and both shimmer and noise to harmonic ratio, a correlation between Roughness and jitter, and a correlation between Breathiness and shimmer.

Alongside studies comparing acoustic measures to GRBAS, there are other studies using more generic measures of dysphonia. Eskinazi et al. (1990) used a scale of 1-7 to represent five different aspects of dysphonia of the voice (Overall severity, Hoarseness, Breathiness,
Analysis of the ageing voice

Roughness, and Vocal Fry), and compared it to various acoustic measures derived from LPC coding of the speech signal. The two features that best predicted dysphonia in their study were the “pitch amplitude” (which is found by performing a normalised autocorrelation function of the LPC residual signal and taking the largest peak that is not the origin) and the harmonics to noise ratio. The pitch amplitude would relate to steadiness of the pitch, but another measure relating to steadiness of the pitch, the jitter, was less useful in predicting dysphonia, although it did contribute to prediction of hoarseness and breathiness.

Eskinazi et al. (1990) analysed both normal and dysphonic voices in their study, but found that there was very low interjudge reliability on assessments of the normal voices. As such, their analysis was focused on the results for dysphonic speakers. This illustrates one of the issues regarding assessment of normal voices using perceptual measurements; without examples at the extremes for each measurement scale, listeners’ judgments of perceived quality are unlikely to match well. As discussed above, a standardised protocol such as GRBAS limits these interjudge reliability issues, but at the cost of precision, which would be most detrimental when the speech samples being analysed all fall within the relatively narrow band of “normal” voice quality. In these instances, the use of quantitative measurements derived from acoustic waveforms show their strength, as they have greater ability to discriminate between various “normal” voices.

Not all studies have found connections between perceptual scales and acoustic measurements, however. Wolfe et al (2000) used a 7 point scale to rate dysphonia much like Eskinazi et al. (1990), and found no strong correlations between ratings of dysphonia and the acoustic variables they analysed; cepstral peak prominence, Pearson R at autocorrelation peak, breathiness index, power spectrum ratio and period deviation, along with a measure of F0. In this study, untrained listeners were used for the perceptual measures, but they found interlistener reliability measures were comparable to other studies.

Each of these studies tends to note that there is relatively little agreement on which acoustic features correlate to perceptual terms such as breathy, hoarse, or harsh. The inconsistency of the results between the studies further illustrates that this is the case. Perhaps the biases and experience of the different listeners in each perceptual study have an impact on what acoustic features are relevant to the perceptual measures.

While there is certainly value in measuring the perceptual features, the inconsistency between results of different studies makes cross-study comparisons difficult. As Bhuta et al. (2004)
point out, the current thinking is that if they can be correlated, a perceptual measure (such as GRBAS) and an analytical measure (such as MDVP) can be combined to limit the disadvantages of each. While there is certainly some evidence of correlations, the incompleteness of this has not allowed the development of a standardised combined approach thus far.

Work on relating the perceptual features of voice quality to the underlying mechanisms will no doubt continue. However, there is still much that can be done in further examining acoustic measures, and developing new quantifiable measures of voice quality. It may be that the measures that will best correlate to perceptual features are yet to be developed. This thesis will focus on quantifiable measures derived from a range of speech measurement techniques (that will be described in Chapter 3), but it is still useful to reflect on the fact that we are ultimately trying to understand how perceptual differences in voice quality occur.

2.3. Voice quality and stability in the ageing voice

One of the clear contributors to voice quality is chronological age. Older age results in perceivable differences in voice quality, and a considerable body of research has investigated the causes of these differences, using a combination of the recording paradigms. The results are not always consistent, but there are a number of general trends that have emerged.

2.3.1 Physical changes in the vocal folds caused by age

The biological process of ageing is called senescence (Comfort, 1956). One of the vital underlying mechanics for senescence is a limitation on the number of times a human cell can divide, referred to as the Hayflick limit (Burnet, 1974). The mechanism behind this limit appears to be shortening of the telomeres in each cell's DNA with each cell division. Once they reach a critical length, the cells will no longer divide.

Having lost their ability to divide, they are no longer able to replace damaged cells or repair tissue damage. This leads to an increased degradation in tissues with age. Other proposed mechanisms for degradation in tissues include chemical damage to DNA. With regard to the vocal folds, the different tissue layers each experience changes due to senescence, causing changes in the properties and function of the vocal folds as described in Hirano et al (1989), Sato et al. (2002), and Colton et al. (2006). Elastic fibres in the lamina propria become rough and variable in size with increased age, and operate in a branched network rather than being
aligned parallel to the edge of the vocal folds. There is also degeneration and atrophy of the elastic fibres in the superficial layer of the vocal folds. Muscle fibres within the thyroarytenoid muscle also show atrophy and stiffening. These changes affect the movement of the vocal folds during phonation.

Hirano et al. (1989) used histological analysis of vocal fold tissue from human larynges ranging in age from 70 to 104 years, and compared the results with previous data from younger age groups. They found that vocal folds were shorter in older speakers, and this shortening was more marked with males than females. The thickness of the cover (the epithelium and superficial layer of the lamina propria) increased with age for females. For males, the thickness of the cover increased until the age of 70, after which it decreased with age. There was more atrophy of the elastic fibres within the lamina propria for males than for females.

2.3.2 Fundamental frequency (F0)

F0 is usually found to gradually increase with age in men, and gradually decrease in age with women, with the rate of change of F0 increasing from about 50 years onwards (Schötz, 2007, Harnsberger et al., 2008, Torre et al., 2009, Ma and Love, 2010). For instance, Torre et al. (2009) used CVC (consonant vowel consonant, e.g. had) words in carrier phrases to show higher F0 in older men (mean age 75.2 years) than younger men (mean age 25.5 years). Harnsberger et al. (2008) used the second sentence of the Rainbow Passage to show higher F0 for older men (62-92 years) than younger men (18-30 years). Ma and Love (2010) found a similar age difference in F0 in a combination of sustained vowels, the phrase “a baby boy,” and a reading of “The North Wind and the Sun”. Awan (2006), Ferrand (2002), and Goy et al. (2013) all found F0 decreases with age in women, using a combination of sustained vowels and readings of the Rainbow Passage. In contrast, Reubold et al. (2010) found F0 decreases with age in both men and women, using longitudinal data from archival BBC excerpts. It is generally accepted that after initially dropping, F0 does eventually rise again with age in women, which Baken (2005) suggests occurs around 70 to 80 years of age. Schötz (2007) found that standard deviations of F0 are usually found to increase with age.

As Torre (2009) points out the results vary from study to study, so only the general trends in how F0 changes with age should be given. The more consistent trends in the studies show a gradual raise in F0 from young adulthood to old age for men, whereas women generally exhibit a gradual decrease in F0 with the potential for a re-raising much later in life (for example Baken, 2005). Also commonly reported is an increase of the cycle to cycle variability of F0, or
Analysis of the ageing voice

jitter of the voice (Schötz 2007). This is perhaps indicative of a loss of control of F0 that occurs with ageing. Ferrand (2002), however, found this is not the case and suggest harmonics to noise ratio as a better measure of the change in vocal stability with age.

Perhaps of more interest is the demonstration that the standard deviation for F0 is higher in older speakers, as seen in Gorham-Rowan and Laures-Gore (2006). This would support the hypothesis that ageing brings a gradual loss of F0 control. Given the variance of F0 in normal speech, this could be a factor that allows listeners to help identify the age of speakers. That said, the standard deviation calculations in Gorham-Rowan and Laures-Gore (2006) are performed in Hz rather than log(Hz), and as such may be attributed to raised average F0 with age. Given the nonlinear nature of pitch perception, with a raised mean pitch, the same amount of variance in perceived pitch would result in a larger measured variance in Hz. A longitudinal study by Decoster and Debruyne (2000) found decreasing standard deviation in a longitudinal study of radio broadcasters.

There are of course a number of other factors that influence the F0 of speech, such as language. Bier et al. (2012) showed that elderly male speakers of both Māori and English have been noted to speak with a higher average F0 when speaking in Māori. It is unclear whether the F0 used for Māori or the F0 used for English is their preferred register, but either way there appears to be a distinction made regarding F0 between the languages when speaking. This distinction was not made by the younger male speakers.

The type of elicitation used in the analysis also has an effect on measures of F0 (Fitch, 1990). Fitch found that sustained vowel tasks produced higher measured F0 values than spontaneous and read speech tasks for a cohort of young adult speakers. Both sustained vowels and continuous speech tasks are common in measurement of F0, but the effect of the task on the outcome is not always addressed.

2.3.3 Intensity

Loudness is not as commonly reported, perhaps due to the stricter recording conditions necessary to allow comparability between speakers, as microphone distance has a clear impact.

Along with decreased control of F0, ageing brings a decrease in control of the intensity of the voice, as described Baker et al. (2001). This is likely due to a combination of degradation in the muscles and elastic fibres of the larynx and a decrease in the capability to maintain constant air pressure from the lungs. However, the effect of larynx degradation appears to be greater
Analysis of the ageing voice

than the effect of respiratory control, with older men actually raising their subglottal pressure to combat the increased stiffness of the vocal folds (Baker et al. 2001). In addition to this, Baker et al. also found lower firing rates in laryngeal electromyography, suggesting that nerve control of the larynx is weakened.

Schötz (2007) reviewed several studies and showed that where loudness is reported, it usually remains stable or decreases with age. Changes in vocal intensity with age are not consistently found, however. Morris and Brown (1994) found no significant difference in mean vocal intensity for young and old women in both connected speech and sustained vowels. When examining standard deviations of SPL, Schötz (2007) found that they usually increase with age.

2.3.4 EGG measures

Fewer studies have examined age-related voice quality with EGG, but Ma and Love (2010), and Higgins and Saxman (1991) both found that Qx decreased with age in men, and increased with age in women. However, Winkler and Sendlmeier (2006) found that Qx decreased in age for both men and women, but the difference was larger for men.

2.3.5 Perceptual features

In addition to acoustic features that are able to be measured such as F0 and intensity, there are perceptual features in speech that can be used to describe voice quality. These include hoarseness, coarseness and breathiness, and require the use of perceptual studies such as Gorham-Rowan and Laures-Gore (2006) to make meaningful comments about how they change with age. Breathiness is certainly linked with both ageing and incomplete closure of the glottis during voiced speech, but the relationship between the prevalence and extent of incomplete glottal closure is not straightforward. Indeed, Soderstern et al. (1995) found that as many as 80% of young female speakers exhibited a posterior glottal chink under laryngoscopy without having an abnormal voice. Older speakers do usually show more instances of incomplete glottal closure, and this can occur in different places along the vocal folds rather than being limited to a posterior chink (Soderstern et al. 1995).

Gorham-Rowan and Laures-Gore (2006) showed that both hoarseness and coarseness are observed to increase with age, and are also shown to be correlated with each other. What is not certain, is how each listener separates and defines these characteristics in their heads. This is a
weakness of the perceptual studies, but in the absence of trained listeners that allow a degree of inter-listener reliability, it is a difficult issue to amend. Like with breathiness, the implied cause for increases in coarseness and hoarseness is usually incomplete glottal closure.

As reported in Section 2.2.2, perceptually-based measures of voice quality can have low inter-judge reliability, or low precision, as is the case with GRBAS (Bodt et al., 1997). When comparing normal speech from young adults with normal speech of old adults, the contrasts are not as great as those apparent in the examination of dysphonic voices.

2.3.6 Short term stability

Short term variability of the voice can be taken as an indication of stability of the voice as discussed in Section 2.2. Typical measurements of stability take the form of standard deviations or perturbation measures such as jitter and shimmer. Separating short term variation in voice features from longer term variation due to prosodic and phonetic variation can be difficult. A common way to avoid this problem is to examine sustained vowels where the speaker is producing a constant target pitch and target loudness. When examining such a sustained vowel, both the standard deviation of F0 and the perturbation of F0 (jitter) provide an indication of a speaker’s ability to hold a stable pitch. Likewise, the standard deviation of sound pressure level (SPL) and the perturbation of SPL (shimmer) give an indication of the stability of loudness.

In studies where the effect of age on jitter and shimmer is examined, there is usually an increase in jitter and shimmer seen with age (Baken 2005, Schotz, 2007, Dehquan et al. 2013), if there is any statistically significant difference observed. Baken found that in adult speech, jitter and shimmer have both been shown to increase with age from approximately 0.5% in subjects in their 20’s up to 1% for subjects over 70 years of age. Orlikoff (1990) also found an increase in jitter with age. However, several studies such as Ringel and Chodzko-Zajko (1987), Ferrand (2002) and Schotz and Muller (2007) found no significant increase in jitter with age. Ringel and Chodzko-Zajko (1987) did find an increase in shimmer with age.

Measurements of jitter and shimmer can vary considerably between software implementations and algorithms. Amir et al. (2009) and Haldun et al. (2011) both found discrepancies between measurements made with Praat and MDVP, two of the more commonly used speech analysis software packages. The lack of consistency in finding age-related differences in perturbation measures may in part be due to inconsistency in the application of these measures. Perturbation measures are also primarily able to be related to sustained vowels. Yiu et al (2000) pointed out
the unreliability of jitter and shimmer measures taken from connected speech, as it requires examination of a more complex signal encompassing coarticulation, intonation, and multiple onset and offset of phonation.

2.4. Conclusions

There are currently a large number of approaches used to analyse voice quality. Many of them use underlying assumptions based on the source filter model of speech production in order to relate measured acoustic features to the physiological components of voice production. In this thesis, the focus is on the vocal folds, or the source for voice production.

There are a number of physical changes to the vocal folds (and the wider voice production system) that occur with age which contribute to the differences in voice quality. Studies analysing acoustic measures of young and old speakers are not entirely consistent, but there are some broad trends that occur. F0 is generally found to decrease with age in women, with an eventual increase starting at about 65 to 70 years of age. For men, F0 is generally found to increase with age. Intensity or SPL is generally found to decrease with age. Measures of short term stability of F0 and SPL (perturbation measures) are usually found to increase with age.

Voice quality can primarily be assessed with either perceptual measurements or with quantitative measurements derived from the acoustic waveform and/or other physiological measurements of voice production such as electroglottography and laryngoscopy. Both of these approaches have their advantages and flaws. Perceptual tests are the closest to what we are trying to identify. Investigation of age-related changes in voice quality is a result of there being perceivable differences in voice quality between young and old speakers. However, perceptual tests can have a lack of inter-judge and intra-judge reliability, and protocols that aim to limit these issues (such as GRBAS) sacrifice the level of precision of the measure in order to do so. Qualitative measurements are more reliable, with better capability for repeatable results. However, they are still marred by inconsistency between different applications of the various acoustic measures. Different software programs often produce different results, limiting comparability between studies.
This thesis will focus on quantitative measurements rather than qualitative measurements. The efforts made to limit inconsistencies in the implementation of the quantitative measures will be described in Chapter 3, where the algorithms used in this thesis are detailed.
Chapter 3. Signal processing and measurement techniques in voice analysis

This chapter will examine three methods of measuring the speech signal, along with signal processing methods used to analyse these measurements. Acoustic recordings form the baseline for voice research, and as such any such research needs to be considered in the context of the acoustic signal. In terms of examining the mechanisms for voice production, two more analysis techniques are discussed; laryngoscopy which provides a video view of the vocal folds during phonation, and electroglottography (EGG), which uses the changing electrical resistance across the larynx to give an estimation of vocal fold contact. Given that all three methods are examining the same overall physiological behaviour, they are all interrelated, but for the purposes of this thesis the primary measurement used will be EGG. These methods can be applied to measuring voice and voice quality in general, they will all be used in the analysis of age-related changes in voice quality within this thesis.

3.1. Common speech measurement techniques

3.1.1 Acoustic recordings

The acoustic waveform is what constitutes the speech signal that is passed from a speaker to a listener. It contains all of the information that a listener can infer through hearing alone. In research where we wish to understand perceivable features of the speech signal, the particular features that can be perceived from the acoustic are important. This thesis primarily deals with changes in voice quality that occur with age, and these changes are often able to be perceived by listeners (Baken, 2005).

The acoustic speech signal is very rich in information, and ultimately is the most relevant in terms of understanding how we perceive speech. Researchers have been employing acoustic speech recordings in their research for decades and there exists a plethora of different microphones and mechanisms for recording speech signals.

Several measures based on the acoustic speech waveform have been developed over the years. Two of the simpler, common measures derived from acoustic recordings are the fundamental frequency (F0) and sound pressure level (SPL) (Schötz 2007, De Cheveigne and Kawahara,
2001, Noll, 1967). Accurate calculation of the SPL does require the microphone used for the recordings to be calibrated (Winholtz and Titze, 1997).

In terms of measuring voice quality and stability, tracking the movement of F0 and sound pressure level is a common approach. Human hearing is sensitive to both pitch and loudness, so measurement of F0 and sound pressure level allows quantitative measurement of features which relate to perceivable differences in pitch and loudness. It is possible to look at standard deviations of these measures to provide indications of stability, but a common approach is to use perturbation measures of F0 and SPL. Perturbation measures random fluctuation of a given feature from cycle to cycle. In the case of acoustic analysis, the cycles used for analysis are defined by the period of F0. The perturbation of the F0 is referred to as jitter, while the perturbation of SPL is referred to as shimmer.

### 3.1.2 Laryngoscopy

Laryngoscopy is one of the most direct ways of observing vocal fold vibration during phonation. A view of the vocal folds is achieved through either the use of a rigid scope through the mouth, or a flexible scope usually inserted through the nose. Unlike EGG, laryngoscopy provides information about the changing glottal area, and allows examination of the extent of closure. In terms of video laryngoscopy, the two main approaches are high speed videoendoscopy and videoostroboscopy (Olthoff et al., 2007, Deliyski and Hillman, 2010). While the use of high speed videoendoscopy provides the most detail for analysis of cycle to cycle variations in vocal fold vibration, videoostroboscopy is still useful for real time visualisation of vocal fold vibration with simultaneous audio playback. It is also the most widely used laryngeal imaging technique in clinical settings (Mehta and Hillman, 2012).

Typical frame rates for high speed laryngoscopy reach 2 kHz (Kendall, 2009), which allows examination of the cycle to cycle variations in vocal fold vibration. However, this requires expensive equipment, and examination of the vibratory behaviour requires later analysis, because the vibrations are too fast for the human eye to follow in real time.

Stroboscopy is more commonly used in a clinical setting when making a diagnosis (Olthoff et al, 2007, Mehta and Hillman, 2012). By strobing the light source for the camera at a rate close to the fundamental frequency of the voiced speech being produced, an aliasing effect is achieved, slowing the apparent rate of vibration to a speed that is easier for a clinician to interpret. For instance, a strobing speed that is 1Hz higher or lower than the F0 of the voice
Signal Processing and measurement techniques in voice analysis

will produce an apparent vocal fold vibration rate of 1 Hz. As the individual frames recorded are each from a different vocal fold cycle, the observed vibratory waveforms are in a sense averaging across many cycles. This reduces the capability to analyse the cycle to cycle stability, but it is possible to make qualitative assessments of the vibratory behaviour based on stroboscopic video.

Aside from the lower level of detail, the primary disadvantage of stroboscopy is the need for real time tracking of F0 to allow appropriate strobing speeds. As such, it is best applied to sustained phonation tasks, preferably with stable phonation characteristics (Deliyski and Hillman, 2010). Concurrent audio and EGG recordings can provide the F0 information necessary to select appropriate strobing speeds, in addition to allowing assessment of the cycle to cycle stability within the segments analysed.

Unfortunately, by virtue of how the camera is able to view the vocal folds, both high speed laryngoscopy and stroboscopy are more invasive than the other procedures described in this chapter. When using a rigid laryngoscope inserted via the mouth, the best view of the vocal folds is obtained when the tongue is maximally depressed, as it is in the production of the vowel /a:/ . Flexible laryngoscopy via the nose benefits from maximal tongue fronting from vowels like /i:/ because this helps prevent the tongue from obscuring the view of the vocal folds. In addition, the presence of a laryngoscope in either the mouth or nose can be uncomfortable.

Subsequently, there is potential for reduced naturalness of the voice during laryngoscopic recordings. Vowel choice for analysis is often limited to the vowels /i:/ or /a:/ in order to satisfy the need to keep the tongue from obstructing the view of the vocal folds.

Ideally, we would like to have a method for getting vibratory data from the vocal folds that is more direct than models created from the acoustic waveform, but is less invasive and more conducive to normal voice production than laryngoscopy. In this thesis an appropriate middle ground is found in EGG.

3.1.3 Electroglottography

Electroglottography (EGG) is a method of monitoring movement of the vocal folds during speech more directly than the acoustic waveform. This was first reported by Fabre (1957) as a means to examine laryngeal function without the obvious problems associated with obscuring
the oral cavity. Given its ease of use, the method quickly spread across Europe and into the United Kingdom for use with clinical populations of motor speech and voice disorders (Baken, 1992). Two electrodes are placed on either side of the larynx and these are used to measure the changing resistance during speech, as seen in Figure 3.1. A microphone (in this case, a lapel microphone) is used to concurrently record the acoustic signal. When the vocal folds are in full contact, there is a greater path for electrical current to flow, causing a relatively low electrical resistance across the larynx, and when the vocal folds are apart, there is a decrease in the conductive path for electricity, causing a relative increase in the resistance across the larynx. As such, tracking the changing electrical resistance gives a waveform which relates to the vocal fold contact area (Childers, 1990), with lower resistance corresponding to increased vocal fold contact, and higher resistance corresponding to decreased vocal fold contact.

The use of EGG to monitor the extent of abduction and contact in the vocal folds was pioneered by Rothenberg and Mahshie (Rothenberg and Mahshie, 1988). They explored the use of duty cycle based measurements as markers of vocal fold contact, and compared how different thresholds for duty cycle affected measurement of different types of phonation. As the EGG signal is related to the vocal fold contact area, the period of time during which there is no vocal fold contact provides very little change in the EGG waveform. Rothenberg and Mahshie noted that this results in the typical appearance of an EGG signal being akin to an asymmetrically clipped sine wave, they suggested that extrapolating out an unclipped waveform might give an indication of the actual vocal fold movement. As such, the 50% threshold for the duty cycle of vocal fold movement would likely correlate to a lower percentage threshold of the EGG waveform. Comparing EGG duty cycle calculations with thresholds of 50% and 30% with concurrent glottal flow measurements, Rothenberg and Mahshie found a threshold of 30% to be more reliable in distinguishing between normal and tight voice than a threshold of 50%. Other studies have continued with the use of threshold based measurement of the extent of vocal fold contact. It has remained common to use a threshold of 30% (e.g. Hertegard and Gauffin, 1995) although there has been some variation in the thresholds used (e.g. Orlikoff, 1991, Sapienza et. al., 1995, Herbst and Ternstrom, 2009). The Laryngograph electroglottograph (www.laryngograph.com) which is used in this thesis is provided with analysis software (Speech Studio 3.5.6) which uses a threshold of 30%. This software package is the main point of comparison for the analysis algorithms developed in the thesis, so 30% was chosen to be the threshold used in this research.
An example of typical EGG waveforms can be seen in Figure 3.2. The initial EGG signal is a changing voltage, although it is stored treated by the Laryngograph as a wave file, providing a normalised magnitude. When comparing the EGG signal to the corresponding speech signals in Figure 3.2, it can be seen that the EGG signal is dominated by the F0, while the speech signal contains more high frequency information. An increase in the EGG signal corresponds to greater vocal contact; the examples in Figure 3.2 show a sharp rise in the EGG waveform corresponding to rapid closure of the vocal folds, followed by a more gradual opening. Figure 3.2 also demonstrates some of the differences between the EGG and speech waveforms of a young speaker (a and c) and an old speaker (b and d). Comparing (a) to (b), you can see that the more positive part of the EGG cycle takes up a greater portion of each cycle for the younger speaker (a) than for the older speaker (b), which indicates that the younger speaker has contact between their vocal folds for a larger portion of each vocal fold cycle.

Figure 3.1: EGG experimental setup. The electrodes and microphone are both attached to the microprocessor, which passes the recorded signals to the computer interface.
Figure 3.2: EGG waveforms for (a) a young speaker and (b) an old speaker and acoustic waveforms for the same young speaker (c) and old speaker (d). All waveforms show a 30 ms segment of the vowel /i:/ at normal pitch and volume.

Changes in the vocal fold contact area give useful information about the vibratory behaviour of the vocal folds, but there is a key issue regarding how much we can infer from this waveform. While the EGG waveform shows the phases of the vibratory cycle that are “more open” and “more closed,” it does not indicate whether the vocal folds are reaching full closure, nor does it indicate how far apart the vocal folds are during the open phase. The peak of the EGG waveform corresponds to the greatest amount of contact in a given cycle, but this does not specifically imply full closure of the folds. During the open phase, the EGG waveform relates to the extent of vocal fold contact rather than the distance between the vocal folds. As such, when discussing measures provided by EGG, it is important to use names and descriptors that accurately reflect the information the EGG signal provides.

One of the common measures derived from the EGG waveform is the contact quotient, abbreviated in this thesis as Qx. The contact quotient is the ratio of the contact period divided by the total period for each vocal fold cycle. Its calculation is described in more detail in Section 3.2. The contact quotient has been shown to decrease with age in men (Ma and Love, 2010,
Signal Processing and measurement techniques in voice analysis

Bier et. al. 2012), and has also been shown to increase with increased loudness (Bier et. al., 2012, Orlikoff, 1991). Note that the contact phase is the portion of time during which no (or very little) air can flow through the glottis. Increasing the contact quotient would therefore decrease the mean airflow. However, it is the amplitude of pressure oscillation, rather than airflow, which determines the loudness of an acoustic signal. Quieter vocalisation has been shown to have higher mean airflow (Holmberg et al., 1988). Decreased contact quotient also leads to higher airflow, which is linked to a perception of increased breathiness in speech (Hartl et al., 2003).

The contact quotient is sometimes erroneously referred to as the closed quotient, due to an assumption that the vocal folds are fully closed during the high contact phase of the EGG waveform. To prevent confusion produced by this assumption, some authors such as Orlikoff (1991) opted for the term contact quotient whereas others like Hacki (1996) have used the term quasi-closed quotient. Others such as Sapienza et al (1998) have avoided confusion between these terms by instead dealing with the open quotient, which is approximately the converse of the contact quotient. In this thesis, the contact quotient was chosen as the preferred EGG derived measure, given that the most direct physiological interpretation of the EGG waveform is the extent of vocal fold contact.

While the EGG waveform gives us an indication of when there is increased contact, calculation of the contact quotient requires a method determining when the contact phase begins and ends. Methods for this vary, but the two common approaches are to use a threshold to determine the start and end point of the closing phase (Sapienza et al., 1998, Herbst & Ternstrom, 2006), or to use markers from the differentiated EGG waveform to determine the start and end points of the contact phase (Herbst & Ternstrom, 2006). For the latter approach, markers are often taken from the differentiated EGG waveform. For instance, Henrich et al. (2004) took the point of fastest closure from the differentiated EGG to be the start of the contact phase, and the point of fastest opening to be the end of the contact phase.
For the purposes of this thesis, a thresholding approach was used to determine the start and end of the contact phase, with a threshold of 30% of the peak to peak amplitude of each cycle, as shown in Figure 3.3: Calculation of contact quotient from the EGG waveform. This was largely due to thresholding being the default approach for the Speech Studio software used in this research (Laryngograph, 2010). The algorithms developed for this thesis also used thresholding to be consistent and to allow comparability with the values calculated within Speech Studio. In both cases, the threshold used was a value of 30% of the peak to peak amplitude for a given cycle.

In addition to being able to calculate the contact quotient, which is one of the primary measures of vocal fold activity used in this thesis, the EGG waveform is seen to be a more reliable source than the acoustic waveform for calculating the fundamental frequency of voice production (Orlikoff 1991, De Cheveigne and Kawahara, 2001). Given that EGG provides more accurate measures of F0 and jitter than the acoustic waveform (Kilic et al., 2004, Orlikoff, 1995), it follows that it may also provide more information on the mechanisms controlling F0 and jitter.

### 3.2. Signal processing techniques in the assessment of voice quality and stability

In this thesis, we are interested in examining speech stability, and how it varies with age. While qualitative assessment can be informative, ultimately we require signal processing of the speech measurements made with the techniques described in Section 3.1 to quantify the differences we observe.
The primary focus in this thesis is voiced speech, so voice stability measures relating to the vocal fold behaviour are the greatest concern. As such, there will be a focus on the fundamental frequency and the contact quotient.

### 3.2.1 Fundamental frequency calculation

There are many algorithms available to calculate F0 including both time domain and frequency domain approaches (Hess & Indefrey, 1984, De Cheveigne & Kawahara, 2001). It is most common to use acoustic waveforms to calculate F0, but as De Cheveigne and Kawahara point out, it is usually easier to use the EGG waveform for F0 determination.

With the EGG waveform (see Figure 3.3), the overall shape of the waveform is simple enough that a time domain based feature picking method can be used to determine periodicity, and hence, F0. As with calculation methods for finding the contact quotient in 3.1.2, feature picking approaches can include peak picking or threshold crossing, of the waveform itself, or a differentiated version of the waveform (Sapienza et al., 1998, Henrich et al., 2004, Herbst & Ternstrom, 2006). Taking the maximum peak of the EGG waveform would correspond to using the point of maximum vocal fold contact to define the start and end of the vocal fold cycle. Taking the maximum peak of the differentiated EGG waveform would correspond to the point of fastest closure of the vocal folds. In this thesis, the feature used was the point of vocal fold contact, as calculated using thresholding (see Figure 3.3), primarily because this feature is also being used for calculation of the contact quotient.

The acoustic waveform has greater high frequency content than the EGG waveform, and this makes reliably identifying quasi-periodic time domain features more difficult. Other, more advanced approaches can be used, such as the autocorrelation method, which for many years has been one of the more reliable and robust pitch detection methods (Rabiner, 1977). Pitch detection algorithms that involve converting the signal to the frequency domain, such as the cepstral method are also useful with the acoustic waveform.

Threshold crossing is a reliable method of extracting F0 from the EGG waveform, that is robust regardless of whether the EGG signal is low-pass filtered first (Titze et al., 1987). Methods based on peak picking are less reliable in the presence of low pass filtering. Titze et al. demonstrated that using interpolation between samples to find the exact time point for threshold crossing allows increased accuracy in the measurement of cycle to cycle variation in F0, even
when the sampling frequency is somewhat limited. This form of establishing the time point of threshold crossing is seen in Figure 3.4.

![Figure 3.4: Use of interpolation to calculate threshold crossing times. Linear interpolation between the two samples on either side of the threshold allows a more accurate estimate of the time of threshold crossing to be estimated using Equation 3-1](image)

Using a simple linear interpolation, the horizontal position of the crossover point between $x_1$ and $x_2$ will be proportional to the position of the threshold ($y_{\text{Threshold}}$) between $y_1$ and $y_2$. As such the threshold time $x_{\text{Threshold}}$ can be calculated with Equation 3-1.

$$x_{\text{Threshold}} = \frac{(y_{\text{Threshold}} - y_1)}{(y_2 - y_1)}(x_2 - x_1) + x_1$$

Equation 3-1

Two EGG based F0 calculations are utilised within this thesis. The first of these is the in-built F0 calculation in Speech Studio, the software used to make electroglottographic recordings, which is used for the first set of analysis in each of Chapters 4 and 5. Due to a lack of transparency in the method of calculation within Speech Studio, a second method of calculating F0 was developed within the analysis software, R. This method utilised the threshold crossing of Titze et al., with interpolation to provide accurate determination of the time of threshold crossing. This second method was used for the analysis relating to the perturbation of the contact quotient in Chapter 4, and the contour analysis in Chapter 5. The calculated F0 values from the two calculation methods are not identical, but are strongly correlated.
3.2.2 Calculating the contact quotient

The Qx for a single period can be calculated according to Equation 3-2, where $T_{\text{contact}}$ is the duration of the vocal fold cycle during which the vocal folds are considered to be in contact, and T is the total period of the vocal fold cycle. The equation relies on determination of the contact period and the total period for each vocal fold cycle. Figure 3.5 shows how the contact period can vary for utterances with different vowels.

$$Qx = \frac{T_{\text{contact}}}{T}$$

Equation 3-2

As with F0, there are two methods of contact quotient calculation used in this thesis. The first method, implemented within Speech Studio, uses a threshold of 30% of the peak to peak magnitude of the EGG waveform. However, Speech Studio only provides four measures of Qx from each segment; the mean, the minimum, the maximum, and the standard deviation. As with F0, the calculation process is opaque, and does not allow easy access to the cycle to cycle contact quotient data. For any analysis of dynamic changes in contact quotient to be possible, it was necessary to develop an algorithm to find the contact quotient for each individual cycle.
In order to remain comparable with results provided by Speech Studio, the calculation of Qx within the developed algorithm also relied on a threshold of 30% of the peak to peak amplitude of the EGG waveform for each vibratory cycle. Due to the opaqueness of the calculation performed with Speech Studio, a potential point of difference is the use of interpolation to find precise time points for the start and end of the closed phase. This process can be seen in Figure 3.4. The y axis shows the EGG magnitude, which is speaker dependent. The dashed line represents the threshold of 30% of the peak to peak amplitude, which is used to determine the start and end of each closed phase.

A simple linear interpolation between the two samples on either side of the threshold crossing is used to calculate a more accurate time point. With this method, the F0 value for each consecutive cycle is also recorded.

### 3.2.3 Perturbation measures

Perturbation is the average difference between values for consecutive vocal fold cycles divided by the mean for each cycle. Jitter and shimmer are the two most commonly used perturbation measures in speech analysis and for each of these measures there are a number of ways they may be calculated (Schoentgen and Guchteneere, 1995). In this study, the Laryngograph is the device used to measure EGG. The jitter provided by the Laryngograph is calculated using Equation 3-3, where N is the total number of vocal fold periods in the segment analysed, and F0n is the nth vocal fold period (i.e. a time measure). Shimmer is calculated using Equation 3-4, where the peak amplitude of each period is substituted for F0n (i.e. an amplitude measure).

\[
jitter = \frac{100 \frac{1}{N} \sum_{n=1}^{N-1} |F0_{n+1} - F0_n|}{\frac{1}{N} \sum_{n=1}^{N} F0_n}
\]

\[
shimmer = \frac{100 \frac{1}{N} \sum_{n=1}^{N-1} |SPL_{n+1} - SPL_n|}{\frac{1}{N} \sum_{n=1}^{N} SPL_n}
\]

F0 and jitter are fundamentally related to the behaviour of the vocal folds. They can both be taken as either an acoustic or a physiological measure, because the link between vocal fold vibration and F0 is well understood. While it is common to determine F0 and jitter from an
acoustic waveform (Baken 2005, Kilic et al, 2004) more accurate results are obtained from EGG (Kilic et al, 2004, Orlikoff, 1995), which is more directly related to vocal fold vibration, without the filtering effects of the vocal tract.

Jitter relies on the calculation of the period for each vocal fold cycle. One of the factors that can have a strong effect on the calculation of the vocal fold periods in these algorithms is the sampling frequency of the speech being analysed. If, for instance, a sampling rate of 16kHz is used to measure a voice segment with an F0 of approximately 160 Hz, a vocal fold period difference of one sample each cycle would produce a jitter value of 1%, which is in the region of values we would expect to see in normal voice production. Higher sampling rates can alleviate this to a degree, but even with a sampling rate of 48kHz (the highest sampling rate available with the Laryngograph electroglottograph used in this research) sampling error could still account for approximately one third of a 1% jitter value. This potential error level is still unacceptable. The problem of the sampled signal giving rise to an incorrect jitter value was addressed by Titze et al. (1987), who found that interpolation at the start and end of each vocal fold period can improve the accuracy of jitter calculations (Titze et al., 1987).

Historically, the two perturbation measures of jitter (cycle to cycle variation in F0) and shimmer (cycle to cycle variation in loudness) have been commonly used to measure stability of F0 and loudness. However, there are some issues that plague their use, due to potential inconsistency of how these measures are applied.

What is often not made clear is the precise algorithm used to determine perturbation measures. The basic calculations to find jitter and shimmer in this thesis can be seen in Equation 3-3 and Equation 3-4, but the overall calculation method will be expanded upon in Section 3.2.3. Both of these equations take the mean difference in consecutive cycles of the measured feature, then divides this by the mean value of that feature over the analysed segment. For jitter, the equation requires knowledge of the start and end of each vocal fold cycle, but there are several ways to do this. As mentioned in Section 3.1.3, the method used in this thesis to find the start and end of the contact period is thresholding.

In the calculation of jitter, quantisation of the period presents a potential form of noise that could cloud the results. Typical jitter values for speech are in the order of 1% to 2% (Baken, 2005). If the speech sample being analysed has a sampling frequency of 16 kHz, then cycle length variations between periods of 1 sample Hz will produce a jitter of 1% for a speaker with
F0 of 160. Ultimately, this is where the benefit of the interpolation method described in Section 3.1.3 is found, as we can greatly decrease this form of quantisation noise.

### 3.2.4 Contour modelling

Given that the F0 and Qx calculation algorithms developed for this research provide data of how these features change over time, we have the opportunity to analyse the dynamic nature of these measures. Perturbation measures can provide some indication of variation in F0 and loudness over time, but this is limited to the case of sustained vowels with relatively constant pitch and loudness. Figure 3.6 shows a pitch contour taken from a speech segment recorded for the studies in this thesis.

![Figure 3.6: A pitch contour of a speech segment, showing time domain (top), spectrogram and pitch contour (bottom). Note that pitch contour is superimposed over the spectrogram, but they have separate y axis scales.](image)

By calculating the cycle to cycle F0 and Qx using the algorithms described in 3.2.2 and 3.2.3, we obtain contours for each of these measurements over the analysed segment. These contours can be quantified in simple terms, such as calculations of standard deviation and perturbation,
but for dynamic speech segments it is more useful to be able to describe the changing behaviour that occurs over time. One such way of describing the changes behaviour over time is to use the Discrete Cosine Transform.

The Discrete Cosine Transform (DCT) is a method that has been used to describe contours of various speech parameters, for example, formant trajectories (Watson & Harrington, 1999) and the F0 contour (Teutenberg et al, 2008). The DCT operates by describing a waveform as the sum of weighted cosine waveforms of increasing frequency. The frequencies are integer multiples of the frequency corresponding to a period of double the length of the waveform being analysed. The nature of this process provides two key benefits; time normalisation of the waveform in question, and an inherent low pass filtering if you limit the description to the lower order coefficients. Additionally, because the cosines are phase locked within this transform, each coefficient can be related to a specific descriptor of the shape of the waveform. The characteristic shapes of the first five DCT coefficients are illustrated in Figure 3.7, where it can be seen that $c_1$ corresponds to the mean of the waveform, $c_2$ gives an indication of overall slope, and higher coefficients continue to add higher frequency components to the shape of the waveform. Reconstruction of a contour using a limited number of DCT coefficients effectively low pass filters the contour, as seen in Figure 3.8.

![Figure 3.7: Normalised cosines corresponding to the first five DCT coefficients $c_1$ through to $c_5$](image-url)
Figure 3.8: A reconstructed pitch contour using the first five DCT coefficients

The DCT coefficients or cosine weights are calculated by Equation 3-5, where $c_k$ is the $k$th DCT coefficient, $M$ is the number of samples in the signal, and $f(t)$ is the original signal, which in the case of Figure 3.8 is the pitch contour of a recorded phrase.

$$ c_k = \sum_{t=1}^{M} f(t) \cos \left( \frac{\pi}{M} k \left( t + \frac{1}{2} \right) \right) $$

Equation 3-5

Teutenberg et al. (2008) examined the effectiveness of describing the F0 contour of phrases with the DCT. By increasing the number of coefficients used, the DCT model can more accurately recreate the original contour. However, for Teutenberg et al. (2008) five coefficients were sufficient to bring the contour approximation to a root mean square (RMS) error of less than 10 Hz.

In this thesis, the DCT is used to describe contours of both F0 and Qx. In both cases, low order DCT coefficients provide a useful descriptor of the behaviour. It is possible to apply the DCT to contours of varying length, ranging from single vowels (which will be seen in Chapter five), to entire phrases.
3.2.5 **Mechanisms for assessing voice stability and quality in this research**

While the perturbation measures of the contact quotient and F0 (Jitter) are mostly only applicable to the sustained vowel context, they are calculated from cycle by cycle contours of Qx and F0, allowing the same calculation procedure to be used for the dynamic measures of the F0 and Qx contours and the perturbation measures. There is an extra step required for the perturbation measures. The section that follows will outline the mechanisms used for the calculation of these parameters and will then be applied in Chapters four and five.

The measures developed here are taken from the EGG waveform, which was recorded concurrently with acoustic waveforms using a laryngograph electroglottograph. Initial analysis was performed using the propriety software for the laryngograph (Speech Studio), but the methods described here were performed on the extracted EGG waveform, while still being informed by the initial analysis within Speech Studio.

The boundaries of voice segments to be analysed were determined in the initial analysis in Speech Studio. Striations in the spectrogram of the waveform were used to help determine the start and end of a vowel when necessary. The start and end times for each vowel were then exported and stored in a database. All subsequent analysis was performed on these segments, so that the results within Speech Studio would be able to be compared with the results from the algorithms developed.

The wave files recorded within Speech Studio were stored in a stereo format, with the acoustic and EGG waveforms each being in a separate channel. A MATLAB script was used to split each of these wave files into separate mono wave files. This script created two different versions of the EGG wave; one exactly as recorded within Speech Studio, and the other with a simple FIR high pass filter applied to limit low frequency noise within the EGG waveform. The high pass filtered EGG waveforms were used in the subsequent analysis.

For each segment analysed, the filtered EGG wave file segment corresponding to the vowel start and end times was used. It was then subjected to a low pass filter, as the threshold crossing method used to determine features of the waveform is robust to low pass filtering (Titze, et al.). Two key features were then calculated from the EGG waveform; the F0 and the contact quotient. Each of these were calculated on a cycle by cycle basis. The mechanisms used to calculate these two features are outlined in Figure 3.9.
After the selection of the voiced segment to be analysed, the high pass filtered EGG segment was put through a low pass filter to remove any high frequency noise. For the subsequent algorithms applied to the EGG signal, a rough estimate of the start and end points of each vocal fold cycle was needed. This was obtained using a simple zero crossing method, using linear interpolation between sample points to gain a more accurate time for the threshold crossing itself. This linear interpolation method, illustrated in Figure 3.4, was applied whenever threshold crossing was used to identify a feature from the EGG waveform. It is not essential for the rough cycle determination at this stage of the process, but it is needed to ensure the accuracy of perturbation measures calculated further down the line.

Following the initial cycle estimation calculated from the upward zero crossings seen in Figure 3.10, a local maximum and minimum for each vocal fold cycle was extracted. These provide a local peak to peak amplitude that was then used to determine the threshold needed to indicate the start and end of the contact phase. The threshold used in this research was 30% of the peak to peak amplitude, which is consistent with the thresholding used in Speech Studio for its calculation of the contact quotient. Threshold determination for a rough vocal cycle estimate can be seen in Figure 3.11.
Once the local thresholds had been determined, threshold crossing was then reapplied to find both the upwards threshold crossings and the downwards threshold crossings, as seen in Figure 3.12. The upwards threshold crossings indicate the start of a given contact phase, as well as the start of a given vocal fold cycle. The downward threshold crossings indicate the end of the contact phase for their respective vocal fold cycles.
Once these start and end points were determined, contours for the F0 and Qx could be calculated using Equation 3-6 and Equation 3-7, respectively. The contours given by these equations can then be used for subsequent analysis.

\[
F_{0n} = \frac{1}{(\text{contactstart}_n + 1 - \text{contactstart}_n)}
\]

Equation 3-6

\[
Q_{x_n} = \frac{(\text{contactend}_n - \text{contactstart}_n)}{(\text{contactstart}_n + 1 - \text{contactstart}_n)}
\]

Equation 3-7

There are three sets of analysis tools that we took from these contours, and the processes for getting them can be seen in Figure 3.13 (below). The most straightforward is the standard statistical measures (mean and standard deviation) for each of the F0 and Qx. In addition to this, there is calculation of the perturbation measures. Jitter is already a widely used perturbation measure, as discussed earlier in this chapter, but an additional perturbation measure of Qx is a new measure developed in this research.
The calculation of the jitter can be seen in Equation 3-8 and the calculation of the Qx perturbation or CQP, is seen in Equation 3-9. Both of these perturbation measures are expressed as a percentage.

\[
jitter = 100 \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} |T_{n+1} - T_n|}{\frac{1}{N} \sum_{n=1}^{N} T_n}
\]

\[
CQP = 100 \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} \left| \frac{CP_{n+1}}{T_{n+1}} - \frac{CP_n}{T_n} \right|}{\frac{1}{N} \sum_{n=1}^{N} \frac{CP_n}{T_n}}
\]

The third analytical measure derived from these contours is a descriptor of the shape of the contours, and in this case the descriptors were calculated using the DCT. The DCT takes the F0 or Qx contour and describes it as the sum of phase locked cosines of increasing frequency. The predetermined phase of each cosine is very useful in that it allows each coefficient to relate to a specific shape within the contour. The first coefficient relates to the mean of the contour. The second coefficient gives an indication of the overall slope of the contour. The third
coefficient gives an indication of an overall peakyness or dippyness. As the coefficients go on, they add more high frequency information.

The DCT gives a number of coefficients equal to the number of F0 or Qx points used in the calculation. The transform essentially time normalises where each coefficient relates to a cosine shape over the entire duration of the contour. However, we are limited in the number of coefficients that we can analyse by the number of F0 or Qx points. For very short vowels, where there might only be three or four complete vocal fold cycles, we only have three or four coefficients to work with.

This exposes a limitation of using the DCT to model the contours, but this limitation is inherent in the data it is being used to analyse; the short length of some vowels in connected speech gives relatively little data to work with. Perturbation measures, which rely on a relatively steady state mean value, become less meaningful when there are few data points which may also have an underlying transition or contour. In contrast, sustained vowels give a large number of data points to allow calculations with greater statistical strength, and provide a very useful medium for the calculation of perturbation measures. However, as the speaker is trying to maintain a steady state of vocalisation, there will be little of note in the contour analysis of sustained vowels.

In this research, the standard statistical measures will be applied to all of the different elicitation types in the recordings. Perturbation measures will be used only on the sustained vowel elicitationss. Contour shape analysis using the DCT will be applied to the connected speech and word lists, where the vowels have transient contours.

**3.3. Conclusions**

Current technology provides information rich methods of measuring speech. Acoustic and EGG recordings both allow calculation of several quantitative measures of voice quality, but in particular they allow calculation of the F0 and the Qx during speech, both of which provide some description of the behaviour of the vocal folds during phonation.

In order to measure voice quality, we can measure how the F0 and contact quotient vary with time. Period to period variations are able to be assessed with perturbation measures, while longer term variations are able to be expressed as contours, which can then be modelled and
Signal Processing and measurement techniques in voice analysis

described. This chapter presented the methods used to calculate these measures for this research.

Chapter 4 focuses on the perturbation measures in particular, and demonstrates the use of the Qx perturbation in the measurement of age-related changes in voice quality using sustained vowels. Chapter 5 shows the need for contour modelling to describe how F0 and the Qx change over the course of words and phrases.
Chapter 4. Analysis of sustained vowels using the contact quotient perturbation

Perturbation measures jitter and shimmer are oft reported measures of voice stability in the analysis of sustained vowels. When the effect of age on these perturbation measures is examined, an increase in both jitter and shimmer with age is reported, suggesting that vocal stability decreases with age (Schötz, 2007, Bier & Watson, 2012, Dehquan et al, 2013). However, increases in jitter and shimmer with age are not consistently reported as being significant (Schötz, 2007). It is unclear why this is the case, but it could be a consequence of differences in the nature of the recordings or the implementation of the jitter and shimmer measurements. Having a specific and transparent methodology for calculating perturbation measures will help readers interpret and validate the results. This study uses the perturbation measures developed in Chapter 3 to analyse sustained vowels in young and old male speakers of New Zealand English.

4.1. Study methodology

4.1.1 Participants

For this study participants were taken from two age groups, referred to as “young” (20-26 years, mean = 23.3, SD = 1.8) and “old” (56-71 years, mean = 62.5, SD = 5.6). There were 15 young speakers and 15 old speakers in this study, giving a total of 30 speakers. Ageing is a graduated process, so an age gap of 30 years was selected to allow the differences to be large enough to be observed. A larger difference would likely be observed if the old speakers were even older (for example above 70 years), but we hypothesised that the 30 year gap used in this study should be sufficient to show measurable differences that demonstrate the usefulness of the software algorithms developed.

All speakers were male and spoke with a New Zealand English accent. Prior to being recorded all participants were required to complete a brief questionnaire related to vocal health. The questionnaire found that none of the speakers were current smokers, and only one young speaker had any history of smoking within the previous five years. Four of the old speakers reported varying degrees of hearing loss, but for all except one this was only in one ear. None of the young speakers reported any hearing loss. Hearing loss did not appear to be a
confounding factor in the recordings. None of the speakers were suffering from any respiratory illness at the time of recording. All of the participants except one old speaker reported at least moderate current health and fitness. A summary of the questionnaire responses can be found in Appendix A. Ethics approval for the experiment can be found in Appendix B.

4.1.2 Procedure

The recordings were obtained in a Whisper Room sound isolation booth (www.whisperroom.com) using an electroglottograph developed and produced by Laryngograph (http://www.laryngograph.com/) using the default sampling frequency of 16 kHz. The recording was composed of three parts. Firstly, participants were asked to read the Rainbow Passage (Fairbanks, 1960) in their natural speaking voice. Secondly, participants were recorded while producing the sustained vowels /a:/ and /i:/ for three seconds each. For each vowel the participant was recorded at three different target pitches within their speaking range (low, mid/normal, and high) and three different loudness levels (quiet, normal, and loud), in a manner similar to Painter (1990). Pitch and loudness levels were self-determined by the participants in a manner similar to that implemented by Holmberg et al., (1988). In total this meant that each participant produced 18 sustained vowels. A one second segment from the middle of each vowel was used for analysis, similar to the approach used by Awan (2006). A random selection of the vowel segments was checked by supervising researchers in order to ensure that the segmentation was performed with sufficient accuracy. Finally, participants were required to read word lists containing a range of vowels in an hVd frame. The duration of the entire recording process for each participant was no more than 30 minutes.

Given that the recordings required the speaker to wear an electrode around their neck, there were some practical considerations for the recording procedure. Neck measurements were taken to ensure that an elastic neck band of the appropriate length was used for each participant to hold the electrodes on either side of the larynx. The natural variability in the larynges of participants meant that the Laryngograph was not always able to get clear readings immediately. Therefore, some adjustment of the electrode positioning was needed to ensure tracking of the EGG signal. To limit the number of cords attached to the participant themselves, the lapel microphone used was attached to where the electrode cable split to the two separate electrodes, approximately 15cm from the participant’s mouth. The recording environment was quite unfamiliar and potentially daunting for the participants, which is why the most familiar task (reading a passage of text) was chosen as the first task.
Analysis of sustained vowels using the contact quotient perturbation

For this study, only the sustained vowel section of the recordings was used for analysis because these have been found to provide the best extraction of perturbation measures (Haldun and Kilic, 2011), as well as limiting the effects of prosodic and phonemic variations.

4.2. Analysis tools

4.2.1 Speech Studio

The Speech Studio (version 3.5.6) software for the Laryngograph was used for the sustained vowels. This provides measurements taken from the acoustic and EGG waveforms. The F0 and Qx are calculated from the EGG waveform, while the sound pressure level (SPL) is measured from the acoustic waveform.

The software also provides measures of perturbation, including jitter and shimmer. Both the jitter and shimmer measures used are calculated by the recording software. When calculating the jitter, the software only uses positive going EGG closures to signal the start of each vocal cycle. The shimmer measurement requires both the EGG and acoustic waveforms. The shimmer calculation uses the peak pressure in the acoustic waveform following vocal fold closure for each cycle.

4.2.2 R algorithms

The Speech Studio Software provided with the Laryngograph electroglottograph calculates most of the typical parameters including F0, Qx, SPL, and multiple different jitter and shimmer algorithms. However, it does not provide access to the workings of these calculations, so in order to measure the contact quotient perturbation, it was necessary to create algorithms in a different software package.

The new algorithms described in Chapter 3 were implemented in R and were developed to work on one second segments of sustained vowels. Each algorithm takes a list of start and end points for the different vowel segments, and reads those segments directly from the appropriate WAV file.

The EGG waveform was used to calculate time points for all of the new measures developed. The reference point in each cycle was the start of the contact phase, during the rapid closing of the vocal folds. This can be seen in Figure 4.3. Threshold crossing was used to determine which two samples lay on either side of each vocal fold period boundary. The algorithms then used
linear interpolation, as suggested by Titze et al., (1987), to find time points for the vocal fold period boundaries. In order to make the threshold crossing more reliable against noise, the EGG waveform was low-pass filtered prior to determining the vocal fold period boundaries. Low-pass filtering has minimal negative impact on threshold based boundary measurements (Titze et al., 1987).

Once the time points were found for all vocal fold period boundaries in a given speech segment, the vocal fold periods were calculated from the differences, then converted to frequencies to allow calculation of mean F0 for the segment. The Qx required further thresholding to determine the end point of the contact phase. The Qx for each period was then calculated, and these were averaged to find the mean Qx for the segment. The boundaries of the contact phase and vocal fold periods can be seen in Figure 4.1.

![Figure 4.1: Calculation of the periods and contact phases from the EGG waveform. The y axis shows the EGG magnitude. The dashed line represents the threshold of 30% of the peak to peak amplitude, which is used to determine the start and end of each contact phase.](image)

Once the period and contact phase was found for each vocal fold period of a given speech segment, calculation of the perturbation measures followed easily. Jitter was calculated from the EGG waveform using Equation 4.1, with no additional changes. To calculate the contact quotient perturbation (CQP), the contact quotient of each period was substituted in for $T_n$ giving Equation 4.2, where $CP_n$ is the contact phase time of the $n$th vocal fold period and $T_n$ is the length of the $n$th vocal fold period.
Analysis of sustained vowels using the contact quotient perturbation

\[
Jitter = 100 \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} |T_{n+1} - T_n|}{\frac{1}{N} \sum_{n=1}^{N} T_n}
\]

Equation 4-1

\[
CQP = 100 \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} \frac{CP_{n+1}}{T_{n+1}} \frac{CP_n}{T_n}}{\frac{1}{N} \sum_{n=1}^{N} \frac{CP_n}{T_n}}
\]

Equation 4-2

4.3. Results of sustained vowel analysis

Of the 30 speakers, Speech Studio was able to track Qx for 29, providing 522 tokens for analysis. A further seven tokens from the other speakers failed to track, leaving 515 tokens in the analysis.

A repeated measures ANOVA was performed for each of the parameters provided by the MDVP analysis, with the parameters being the dependent variables, target pitch, and target loudness were ‘within subject’ variables and age was a ‘between subject’ variable. Where a significant difference was found, post-hoc t-tests were performed with Bonferroni corrections to determine where the differences lay. A summary of the age dependent effects can be seen in Table 4.1. There is no significant difference in F0 between age groups, but this may be due to the nature of the sustained vowel task, where each participant was attempting to hold a target pitch for each utterance. Qx showed a significant decrease with age.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Old</th>
<th>Young</th>
<th>F statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td>142.1</td>
<td>140.1</td>
<td>F(1,27) = 0.2, not significant</td>
</tr>
<tr>
<td>SD(F0) (HZ)</td>
<td>2.63</td>
<td>1.31</td>
<td>F(1,27) = 2.3, not significant</td>
</tr>
<tr>
<td>Qx (%)</td>
<td>42.5</td>
<td>46.1</td>
<td>F(1,27)=5.05 (p&lt;0.05)</td>
</tr>
<tr>
<td>SD(Qx) (%)</td>
<td>1.78</td>
<td>1.42</td>
<td>F(1,27) = 1.1, not significant</td>
</tr>
</tbody>
</table>
When examining the target pitch, there was only a significant effect for F0 (F(1.1,29.9)=50.6, p(GGe)<0.001). The differences can be seen in Figure 4.2. Post-hoc t-tests showed that all of the F0 differences were significant, as expected given the nature of the task. SD(F0) and Qx both decreased with increased target pitch, but not significantly.

![Figure 4.2](image)

*Figure 4.2: Target pitch differences for sustained vowels, for (a) F0, (b) SD(F0), and (c) Qx. There were no visible differences for SD(Qx).*

For target loudness, there were significant effects for F0 (F(1.3,35.6)=14.8, p(GGe)<0.001), SD(F0) (F(1.01,27.3)=4.25, p(GGe)<0.05), Qx (F(2.54)=103, p<0.001), and SD(Qx) (F(2,54)=7.14, p<0.01), which can be seen in Figure 4.3. Post-hoc t-tests showed that the loud sustained vowels had significantly higher F0 than the normal or soft vowels. Qx was significantly different across all three conditions, with Qx increasing with loudness. The soft sustained vowels also had a significantly higher SD(Qx) than the normal sustained vowels. The post-hoc t-tests showed no significant differences for SD(F0).

![Figure 4.3](image)

*Figure 4.3: Target loudness differences for sustained vowels, for (a) F0, (b) SD(F0), and (c) Qx. There were no significant effects on SD(F0).*

The results for F0, Qx, SPL, jitter, and shimmer were calculated using the Speech Studio software. Jitter calculations were based on the EGG waveform while shimmer calculations were based on the acoustic waveform. Statistical analysis was then performed in R. Of the 522 sustained vowels in the analysis, seven were unable to be tracked by the laryngograph software,
and were removed from the analysis. Of these seven segments, five were from ‘old’ speakers and all of them had a ‘quiet’ target loudness. The fact that more ‘old’ segments failed to be tracked hints at greater instability for older speakers, which could be a pertinent example of age related difference in voice quality.

Analysis was performed using a repeated measure ANOVA design. Target pitch and target loudness were analysed within speakers, while age was analysed between speakers. Mauchly’s Sphericity test was used to test for sphericity of the ‘within’ measures, and Greenhouse-Geisser adjustments were made for variables that did not exhibit sphericity. In this study the emphasis is on the independent variables age and target loudness, but more detail about the target pitch and vowel can be found in a previous study in Bier and Watson, 2012.

Mean F0 showed significant differences with target loudness ($F(2,54)=14.9, p(GGe)<0.001$), but failed to show a significant difference with age. Mean Qx showed significant differences with both age ($F(1,27)=5.0, p<0.05$) and target loudness ($F(2,54)=103, p<0.001$). Mean SPL showed significant differences with target loudness ($F(2,54)=377, p<0.001$). Jitter failed to show any significant differences with any of the independent variables, but shimmer did show significant differences with target loudness ($F(2,54)=20.2, p(GGe)<0.001$). Mean values of each of these parameters can be seen in Tables 4.1 and 4.2, along with significant results from post-hoc t-tests.

| Table 4.2: Mean values of the Speech Studio parameters for each age group, with post-hoc analysis results. |
|-----------------------------------|-----------------------------------|-----------------------------------|
| F0 (Hz)                           | Qx (%)                            | SPL (dB)                          |
| 140.1 (SD=64)                     | 46.1 (SD=8.2)                     | 82.9 (SD=7.4)                     |
| Young                             | Old                               | Post hoc analysis                 |
| 142.1 (SD=51)                     | 42.5 (SD=7.9)                     | Old>Young, not significant        |
|                                   | 83.2 (SD=7.8)                     | Young>Old (t(27)=2.3, p<0.05)     |
| Jitter (%)                        | Shimmer (%)                       | Old>Young, not significant        |
| 0.87 (SD=2.3)                     | 7.7 (SD=6.2)                      | Old>Young, not significant        |
|                                   | 2.3 (SD=5.5)                      |                                  |
|                                   | 9.6 (SD=9.0)                      |                                  |

Examination of the effect of age on the Qx indicates that it decreases with age, seen in Table 4.2 (above). It also increases with increased loudness, as seen in Table 4.3 (below). Neither the F0 nor the SPL demonstrated the significant age differences expected from Schötz 2007.
Table 4.3: Mean values of the Speech Studio parameters for each target loudness, with post-hoc analysis results.

<table>
<thead>
<tr>
<th></th>
<th>Quiet</th>
<th>Normal</th>
<th>Loud</th>
<th>Post hoc analysis</th>
</tr>
</thead>
</table>
| F0 (Hz)        | 135.7 (SD=57) | 135 (SD=50) | 152.3 (SD=65) | Loud>Quiet (t(28)=3.7, p<0.01)  
Loud>Normal (t(28)=4.1, p<0.01)  
Quiet>Normal, not significant |
| Qx (%)         | 39.7 (SD=7.0) | 44.4 (SD=7.3) | 48.8 (SD=7.8) | Normal>Quiet (t(28)=7.4, p<0.001)  
Loud>Quiet (t(28)=12, p<0.001)  
Loud>Normal (t(28)=7.8, p<0.001) |
| SPL (dB)       | 76.9 (SD=5.0) | 82.4 (SD=5.4) | 89.5 (SD=6.1) | Loud>Normal (t(28)=18, p<0.001)  
Loud>Quiet (t(28)=24, p<0.001)  
Normal>Quiet (t(28)=13, p<0.001) |
| Jitter (%)     | 2.3 (SD=6.1) | 0.96 (SD=1.5) | 1.5 (SD=3.7)  | Loud>Normal, not significant  
Quiet>Loud, not significant  
Quiet>Normal, not significant |
| Shimmer (%)    | 13.1 (SD=10.8) | 7.3 (SD=4.7) | 5.7 (SD=2.9)  | Quiet>Loud (t(28)=4.9, p<0.01)  
Quiet>Normal (t(28)=5.9, p<0.001)  
Normal>Loud, not significant |

While the sustained vowels were spoken, by virtue of being sustained they took on a songlike nature. The lack of significant age differences in F0 or SPL may be due to the sung nature of sustained vowels. The recording task required participants to consciously control both the F0 and SPL produced, which is not the case in spontaneous speech. The F0, SPL, and shimmer all showed significant differences with the target loudness, with shimmer decreasing with increased target loudness, and F0 being higher for the “loud” sustained vowels. There was no significant interaction between age and target loudness; all of the target loudness based effects were present for both age groups.

Table 4.2 shows that the jitter and the shimmer are both higher in the older speakers, although this difference is not significant. This suggests an age dependent decrease in the stability of the voice. Both the jitter and shimmer are generally based on acoustic features, but their generation is due to the underlying behaviour of the vocal folds. With jitter, the link between the acoustic feature (F0) and the underlying physiology (rate of vocal fold vibration) is clear, but with shimmer the connection is less obvious. If there is decreased vocal stability with age evidenced by acoustic measures of jitter and shimmer, there should also be evidence of decreased vocal stability in a more physiologically based perturbation measure. Following this reasoning, we hypothesise that a perturbation measure of the Qx would demonstrate decreased vocal stability with age. To test this, it is necessary to develop a method of calculating the Qx perturbation.
4.3.1 Perturbation measures

The contact quotient perturbation (CQP), F0, Qx and jitter were all calculated using the threshold crossing based algorithms in R. The start and end points of the speech segments used were exactly the same as those used in the earlier analysis. There were four segments where the new algorithm was not able to correctly track F0, so these were removed, leaving 518 segments in the analysis. Of these four segments, two were from the seven removed in the analysis using the measures from Speech Studio, and all four were ‘old’ speakers. Following this, a similar analysis of variance was performed.

As with the Speech Studio analysis, F0 failed to show a significant difference with age, but had a significant difference with target loudness (F(2,54)=16.2, p(GGe)<0.001). While the mean Qx for old speakers remained lower than for the young speakers, this was no longer significant. However, the significant differences with target loudness remained (F(2,54)=72.9, p(GGe)<0.001). The jitter calculated with the new algorithm had a significant difference with age (F(1,27)=6.5, p<0.05). CQP showed a significant difference with age (F(1,27)=9.37, p<0.001), with the old speakers having a higher mean (3.54%) than the young speakers (1.55%).

CQP also showed significant differences with target loudness (F(2,54)=6.86, p(GGe)<0.01). This relationship was consistent with the age effect, with the contact quotient perturbation for all three loudness levels being greater for the older speakers (6.0, 2.6, and 2.0 for quiet, normal and loud, respectively) than the young speakers (1.9, 1.4, and 1.4 for quiet, normal and loud, respectively). Mean values of each of the new measures can be seen in Tables 4.4 and 4.5 (below), along with significant results from post-hoc t-tests.

A Pearson’s correlation test was used to compare the F0, Qx, and jitter calculated with the new algorithms to the values calculated using Speech Studio, and there were highly significant strong correlations for each F0 (r = 0.96, p<0.001) and Qx (r = 0.88, p<0.001), and a significant correlation for jitter (r = 0.45, p<0.001).

Table 4.4: Mean values of the measures calculated in R for each age group, with post-hoc analysis results.

<table>
<thead>
<tr>
<th></th>
<th>Young</th>
<th>Old</th>
<th>Post hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 R (Hz)</td>
<td>140.3 (SD=64)</td>
<td>143.5 (SD=50)</td>
<td>Old&gt;young, not significant</td>
</tr>
<tr>
<td>Qx R (%)</td>
<td>52.0 (SD=9.2)</td>
<td>49.3 (SD=9.7)</td>
<td>Young&gt;Old, not significant</td>
</tr>
<tr>
<td>CQP (%)</td>
<td>1.55 (SD=1.2)</td>
<td>3.54 (SD=8.2)</td>
<td>Old&gt;young (t(27)=2.1, p=0.05)</td>
</tr>
<tr>
<td>Jitter R (%)</td>
<td>0.55 (SD=0.65)</td>
<td>2.35 (SD=8.47)</td>
<td>Old&gt;young (t(27)=2.3, p&lt;0.05)</td>
</tr>
</tbody>
</table>
Table 4.5: Mean values of the measures calculated in R for each target loudness, with post-hoc analysis results.

<table>
<thead>
<tr>
<th></th>
<th>Quiet</th>
<th>Normal</th>
<th>Loud</th>
<th>Post hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 R (Hz)</td>
<td>137.8</td>
<td>136.6</td>
<td>151.0</td>
<td>Loud&gt;Quiet (t(28)=4.4, p&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td>(SD=57)</td>
<td>(SD=50)</td>
<td>(SD=64)</td>
<td>Loud&gt;Normal (t(28)=3.8, p&lt;0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quiet&gt;Normal, not significant</td>
</tr>
<tr>
<td>Qx R (%)</td>
<td>46.2</td>
<td>50.3</td>
<td>55.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SD=9.1)</td>
<td>(SD=8.4)</td>
<td>(SD=8.7)</td>
<td>Loud&gt;Quiet (t(28)=6.2, p&lt;0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Loud&gt;Normal (t(28)=7.4, p&lt;0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Loud&gt;Normal (t(28)=10.1, p&lt;0.001)</td>
</tr>
<tr>
<td>CQP (%)</td>
<td>3.84</td>
<td>1.98</td>
<td>1.70</td>
<td>Quiet&gt;Loud (t(28)=2.7, p&lt;0.05)</td>
</tr>
<tr>
<td></td>
<td>(SD=9.1)</td>
<td>(SD=3.3)</td>
<td>(SD=2.1)</td>
<td>Normal&gt;Loud (t(28)=3.9, p&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quiet&gt;Normal, not significant</td>
</tr>
<tr>
<td>Jitter R (%)</td>
<td>2.27</td>
<td>1.09</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SD=8.1)</td>
<td>(SD=4.5)</td>
<td>(SD=4.4)</td>
<td>Quiet&gt;Loud, not significant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal&gt;Loud, not significant</td>
</tr>
</tbody>
</table>

This verifies that the new algorithm correctly determines the vocal fold periods, and as such there are still no significant differences in F0 between the two age groups. This is likely to be due to the sung nature of the sustained vowels, with attempted holding of a target pitch, as preliminary results from the connected speech and hVd wordlist recordings do show an increase in F0 with age (Bier and Watson, 2012). The recalculated jitter is lower than that seen in Tables 4.1 and 4.2, and now shows significant differences with age.

The CQP shows significant differences with both age and target loudness. The CQP is increased for older speakers, which is consistent with the idea of decreased vocal stability in older speakers discussed in Dehquan et al. (2013). CQP also decreases with increased target loudness, indicating greater vocal stability with increased loudness. This is consistent with the decreased shimmer with target loudness seen in Table 4.2. The jitter values calculated using time point interpolation are lower than the jitter values calculated with Speech Studio. Both jitter measures did have a higher mean value for older speakers, even if this was only statistically significant for the measure calculated in R.

### 4.4. Discussion

The CQP exhibited significant differences with both age and target loudness. The increased CQP with age provides further evidence that vocal stability is decreased with age (Dehquan et al., 2013). The interaction with target loudness implies that vocal effort has a strong effect on this perturbation measure, with the extent of perturbation decreasing with increased target
Analysis of sustained vowels using the contact quotient perturbation

loudness. This decrease was accompanied by a decrease in the recorded shimmer, which further strengthens the connection between Qx and loudness. There is also an implication that the participants had greater vocal stability when phonating at higher loudness levels, which is reinforced anecdotally by many participants finding quieter sustained vowels the most difficult to produce during the recordings.

An important difference between the CQP and the other perturbation measures, such as jitter and shimmer, is that CQP is a physiological measure rather than an acoustic measure. The jitter and shimmer provide some of the acoustic consequences of the behaviour of the physiology of voice production, including the behaviour of the vocal folds. The CQP has the potential to explain some of the acoustic differences between young and old speakers, and the most likely acoustic measures to be explained are the jitter and shimmer. However, if the relationship between the Qx and the other perturbation measures is directly causational, then the CQP should be predictable on the basis of the jitter and shimmer, circumventing the need for EGG. This raises the question of whether the CQP gives us any new information that is not covered by the other perturbation measures.

By examining Equations 4.1 and 4.2, we can see that the period of each vocal fold cycle is part of both calculations, which provides a definite connection between the two. They also both show an increase with age and a decrease with loudness in Tables 4.3 and 4.4, respectively. However, the CQP is greater than the jitter and shows more significant differences with both age and target loudness, thus indicating that there is more to the CQP than variation in the overall time period for each time period. A Pearson’s correlation test showed a strong positive correlation between the jitter calculated in R and the CQP ($r = 0.63$, $p<0.001$). A certain extent of correlation is expected, as the calculation of both measures use the same start and end points for each vocal fold cycle, but given that the CQP has a significant interaction with target loudness and jitter does not, the CQP is not predictable based on jitter alone.

The shimmer also shows similar trends to the contact quotient perturbation for both age differences and target loudness differences, although the shimmer lacked the significant difference between age groups evident with the CQP. When the CQP was tested for correlation with the shimmer, there was a weak, yet significant correlation ($r=0.20$, $p<0.01$). The weakness of this correlation does detract from the possibility that perturbation of the Qx is part of the mechanism leading to the shimmer observed in voice recordings. However, the decrease in
both CQP and shimmer with increased loudness seen in Table 4.4 shows a definite impact of target loudness on vocal stability.

The CQP is providing additional information over and above the traditional perturbation measures of jitter and shimmer. Being a physiologically-based measure, this additional information provided by the CQP is more directly related to the behaviour of the vocal folds, and will provide useful information for the development of models of voice production.

4.5. Conclusions

The stability of the vocal folds during phonation can be assessed by measuring the perturbation of the Qx. This perturbation measure increases with age, indicating decreased vocal stability with age. The CQP and the more commonly reported perturbation measures of jitter and shimmer, all show a decrease with increased target loudness. The CQP provides a more physiologically-based measure of the decrease in vocal stability observed with age, resulting in a quantifiable change in vocal fold behaviour. This measure provides more information about the vocal fold behaviour than can be acquired from traditional perturbation measures such as jitter. These results can then be used to build more realistic models of the vocal fold behaviour that can account for age. Normative models of aged speech would then be able to be used for comparison in the study of pathological speech in aged speakers.
Chapter 5. Development and comparison of dynamic measures of voice stability

When examining voice quality in particular, it is common to use measures relating to the short term (cycle to cycle) variation of the voice. Typical measurements of stability take the form of standard deviations or perturbation measures. Separating short term variation in voice features from longer term variation due to prosodic and phonetic variation can be difficult. A common way to avoid this problem is to examine sustained vowels where the speaker is producing a constant target pitch and target loudness, as analysed in Chapter 4. When examining such a sustained vowel, both the standard deviation of F0 and the perturbation of F0 (jitter) provide an indication of a speaker’s ability to hold a stable F0. Likewise, the standard deviation of SPL and the perturbation of SPL (shimmer) give an indication of the stability of SPL. Similar measurements can be made regarding the vibratory characteristics of the vocal fold behaviour, with measures of the standard deviation of the Qx or its perturbation as seen in Chapter 4.

These stability measures work well in the context of sustained vowels where the features of interest are held relatively constant for the duration of phonation. However, they become more difficult to implement or interpret in connected speech, where prosodic and phonemic effects need to be accounted for. Standard deviation measures cease to indicate stability around a target mean, and begin to encompass the prosodic and phonemic variance within the speech. Perturbation measures are often calculated via equations that don’t account for transient behaviour in F0 or SPL, operating under an assumption that the mean value remains constant throughout the utterance. Some calculations counter this by using a running average from (typically) five cycles as part of the calculation, but this can lead to a systemic underestimation of jitter (Schoentgen & De Guchteneere, 1995). Any measurements based on sustained vowels still need to be related to connected speech, as ultimately, this is the context in which speech is usually heard.

In a sense, spontaneous speech is the ideal form of elicitation to analyse voice quality, because it is most indicative of “real life” contexts. However, it lacks the experimental convenience of repeatability, and therefore limits comparisons across speakers. Read passages are a useful compromise as they allow easy comparison between speakers, but care needs to be taken when selecting a passage in order to ensure consistent phonetic and prosodic environments for the phonemes under investigation. As Mooshammer (2010) demonstrates, word stress has an effect
on measures of F0, SPL, and contact quotient. Word lists can provide consistent phonetic and prosodic environments, but lose the flow of connected speech. Ultimately, a combination approach allows the strengths and weaknesses of the different recording paradigms to emerge, providing a more complete picture of voice quality.

5.1. Hypothesis

This study aimed to investigate age-related changes in the behaviour of F0 and Qx using a combination of hVd word lists (CVC words with the consonants h and d and all eleven monophthongs vowels. See Table 5.1), and connected speech. The expectation was that F0 would be higher for the older speakers in both elicitations while Qx measures would be lower for older speakers in both as well. Standard deviation measures of both F0 and Qx were expected to be higher for older speakers, as an indication of decreased voice stability. Dynamic measures of F0 and Qx in both the word lists and connected speech recordings would provide more information about how standard deviation measures of F0 and Qx relate to the voice quality changes that occur with age. Overall, we expected the age-related differences in voice quality to show similar quantified differences across all three elicitations, and to have a greater impact than the differences between the elicitation methods.

5.2. Methods and materials

This study uses the same set of recordings as in chapter 4, but focuses on the hVd word lists and the connected speech passage. The recordings for one of the old speakers failed to yield usable results in the hVd word recordings, resulting in a total of 29 speakers this section.

For the connected speech, participants read aloud the Rainbow Passage (Fairbanks, 1960), which took approximately 2 minutes for each speaker. For the word lists, participants read five different hVd word lists which cycled through all 11 English monophthongs (/i:, I, e, æ, a:, ʌ, ɒ, ɔ, u, u:/) in a randomised order. Each list had an additional final word which was removed from analysis to limit list effects. This task took each speaker approximately 1 to 2 minutes. There was no observed or reported speaker fatigue over the duration of the experiment.

For the Rainbow Passage, the second, third and fourth sentences were broken into seven natural phrases which are shown in Table 5.2. These phrases were based on the common phrasing within the set of recordings. The first sentence was not used because this was the first part of
the recording and it is important to avoid initial effects. Divisions were made based on pauses in the spoken speech, and audio assessment showed only minor deviations from the transcribed text.

For the hVd words, the entire voiced segment of each word was selected. Vertical striations in a 300 Hz wide band spectrogram were used to identify voicing, as in a wide band spectrogram, the time resolution allows the successive closures of the vocal folds to manifest themselves as these striations (Lamel, 1988). The labelled segments can be seen in Table 5.1.

A random selection of approximately 50% of the speech segments was checked by supervising researchers in order to ensure that the segmentation was performed with sufficient accuracy.

Table 5.1: Phonetic symbols for hVd words analysed.

<table>
<thead>
<tr>
<th>phonetic</th>
<th>hod</th>
<th>hard</th>
<th>hoard</th>
<th>had</th>
<th>head</th>
<th>herd</th>
<th>hid</th>
<th>heed</th>
<th>hood</th>
<th>who’d</th>
<th>hud</th>
</tr>
</thead>
<tbody>
<tr>
<td>symbol</td>
<td>/ɒ/</td>
<td>/ɑː/</td>
<td>/ɔː/</td>
<td>/æ/</td>
<td>/æ:/</td>
<td>/ɛ:/</td>
<td>/ɜː/</td>
<td>/ɪ:/</td>
<td>/ʊ:/</td>
<td>/u:/</td>
<td>/ʌ:/</td>
</tr>
</tbody>
</table>

Table 5.2: ID codes for the seven phrases analysed in the continuous speech recordings

<table>
<thead>
<tr>
<th>ID</th>
<th>Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a</td>
<td>‘The rainbow is a division of white light’</td>
</tr>
<tr>
<td>2b</td>
<td>‘into many beautiful colours’</td>
</tr>
<tr>
<td>3a</td>
<td>‘These take the shape of a long round arch’</td>
</tr>
<tr>
<td>3b</td>
<td>‘with its path high above’</td>
</tr>
<tr>
<td>3c</td>
<td>‘and its two ends apparently beyond the horizon’</td>
</tr>
<tr>
<td>4a</td>
<td>‘There is according to legend’</td>
</tr>
<tr>
<td>4b</td>
<td>‘a boiling pot of gold at one end’</td>
</tr>
</tbody>
</table>

5.3. Analysis tools

The Speech Studio (version 3.5.6) software provided with the Laryngograph includes an analysis tool called Multidimensional Voice Profile (MDVP). This provides a number of measures calculated from the EGG and acoustic waveforms, including the fundamental
Development and comparison of dynamic measures of voice stability

frequency (F0), contact quotient (Qx), sound pressure level (SPL), jitter, and shimmer. The MDVP tool is designed to be used on continuously voiced segments, and is used to analyse the sustained vowel and hVd word recordings. For analysing the connected speech passages, the Speech Studio Quantitative analysis (QA analysis) tool was used. This provided the means and standard deviations for F0 and Qx. Statistical analysis was then performed using R 3.0 (R Core Team, 2013), an open source statistical analysis software package.

The MDVP tool only provides averaged F0 and Qx from each segment analysed. In order to provide dynamic measurement of how F0 and Qx vary over a segment, an algorithm was implemented in R to calculate F0 and Qx from the EGG waveform, as described in Section 3.2. This was used to provide F0 and Qx contours for the connected speech analysis and the hVd word analysis.

5.4. hVd word list analysis

Initial analysis of the hVd word list MDVP data comprised of a series of repeated measure ANOVAs where age was a ‘between subject’ independent variable and vowel was a ‘within subject’ independent variable. As with the connected speech analysis, corrections for sphericity were made when appropriate. The results of the ANOVAs can be seen in Table 5.3. F0 increased with age and Qx decreased with age, which is consistent with our hypothesis in both cases. While not significantly different, the decrease in the standard deviation measures with age runs counter to what is expected. Potential reasons for this discrepancy will be explored in the contour analysis in Section 5.6.

| Table 5.3: Mean values of the Speech Studio parameters for hVd words, with post-hoc analysis results. |
|-------------------------------------------------|-----------------|-----------------|-----------------|
|                    | Young           | Old             | Post hoc analysis |
| F0 (Hz)             | 113.6 (SD=17.8) | 126.4 (SD=18.7) | Old>Young (t(26.4)=2.4, p<0.05) |
| SD(F0) (Hz)         | 6.6 (SD=5.1)   | 5.8 (SD=6.5)    | Young>Old, not significant |
| Qx (%)              | 49.5 (SD=5.0)  | 44.4 (SD=6.5)   | Young>Old (t(23.0)=2.7, p<0.05) |
| SD(Qx) (%)          | 3.6 (SD=2.2)   | 3.2 (SD=2.4)    | Young>Old, not significant |
| SPL (dB)            | 86.2 (SD=4.5)  | 87.0 (SD=4.5)   | Old>Young, not significant |
Development and comparison of dynamic measures of voice stability

Figure 5.1: (a) F0 and (b) Qx for each vowel from the hVd words, for old (o) and young (y) speakers.

The significant differences between vowels provide evidence of phonemic effects for F0 and Qx, and the lack of an interaction with age group suggests that these phonemic effects occur independently of any age-related changes. The overall differences in F0 and Qx for each vowel (between old and young speakers) are illustrated in Figure 5.1. Post-hoc t-test analysis showed several significant vowel differences for F0 and Qx, which can be seen in Table 5.4 and Table 5.5.
Table 5.4: Significant vowel differences in F0. Vowels are ordered from highest F0 to lowest. Significant differences are shaded.

<table>
<thead>
<tr>
<th>F0</th>
<th>/u/</th>
<th>/ɪ/</th>
<th>/u:/</th>
<th>/i:/</th>
<th>/i:/</th>
<th>/ɛ/</th>
<th>/ɛ:/</th>
<th>/æ/</th>
<th>/a:/</th>
<th>/æ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ʊ/</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ/</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u:/</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i:/</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɛ:/</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɛ:/</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/æ:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a:/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>/æ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Similar to what was noted by O'Shaunessy and Allen (1983), the F0 was mostly higher for vowels with greater tongue height. Qx is highest in the vowels with lip rounding (in New Zealand English, /ɔː, u:, u, ʊ:/ are lip rounded). This may suggest that the speakers were compensating for the natural decrease in sound propagation caused by lip rounding. Speakers compensate for differences in articulation during speech (Maeda, 1990). Raising vocal effort could help to bring loudness to an equivalent level to non-rounded vowels, and show itself in the form of raised Qx. The standard deviations of F0 and Qx did not show a significant pattern in their variation.

While the vowel was not considered as a factor in the sustained vowel analysis, it is pertinent here to compare the results of normal target loudness and mid target pitch sustained vowels to
those for the same two vowels in the hVd word analysis, as seen in Table 5.6. These conditions allow comparison of what could be considered “normal” for each of these two recordings. Comparing the two vowels, /i:/ had higher F0 and Qx than /a:/ which was consistent with the hVd word findings. The younger speakers also exhibited lower mean Qx for their sustained vowels than for their hVd word lists, but there was no apparent difference for the older speakers.

Table 5.6: Comparing /a:/ and /i:/ between the hVd word analysis and the sustained vowel analysis.

<table>
<thead>
<tr>
<th></th>
<th>Old /a:/</th>
<th>Young /a:/</th>
<th>Old /i:/</th>
<th>Young /i:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 hVd</td>
<td>118</td>
<td>107</td>
<td>123</td>
<td>113</td>
</tr>
<tr>
<td>F0 sustained</td>
<td>116</td>
<td>110</td>
<td>125</td>
<td>119</td>
</tr>
<tr>
<td>QxhVd</td>
<td>42.6</td>
<td>47.9</td>
<td>42.4</td>
<td>50.4</td>
</tr>
<tr>
<td>Qx sustained</td>
<td>42.1</td>
<td>45.2</td>
<td>42.2</td>
<td>48.4</td>
</tr>
</tbody>
</table>

5.5. Rainbow passage analysis

For the phrasal analysis, seven phrases were taken from each speaker. The MDVP algorithm failed to track Qx for three phrases from older speakers, leaving data from 207 phrases in the analysis.

Repeated measure ANOVAs were performed on the connected speech recordings, with age being a between subject factor and phrase being a within subject factor. Separate ANOVAs were performed for each of the dependent factors; F0, SD(F0), Qx, and SD(Qx). The sphericity of within factors was tested using Mauchly's Test for Sphericity, and Greenhouse-Geisser adjustments were made for variables that did not exhibit sphericity. There were significant differences between age groups for Qx (F(1,28)=9.3, p<0.01), and significant differences between phrases for F0 (F(2.2,60.9)=11.9, p(GGe)<0.001), Qx (F(4.2,117.3)=9.4, p(GGe)<0.001), and SD(Qx) (F(6,168)=2.5, p<0.05).

The overall age and phrase differences are illustrated in Figure 5.2. Both F0 and Qx showed some significant effects between the different phrases. However, SD(Qx) did not show any significant differences between phrases in post hoc analysis.
Development and comparison of dynamic measures of voice stability

Figure 5.2: Phrase by phrase comparison of young and old speakers for (a) F0, (b) SD(F0), (c) Qx, and (d) SD(Qx). Each box represents the range of values across the 15 speakers in each age group, for a given phrase. The corresponding phrase for each code can be seen in Table 5.2.

Table 5.7: Mean values of the Speech Studio parameters for analysed phrases, with post-hoc analysis results.

<table>
<thead>
<tr>
<th></th>
<th>Young</th>
<th>Old</th>
<th>Post hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td>108.6</td>
<td>116.2</td>
<td>Old&gt;Young, not significant</td>
</tr>
<tr>
<td>SD(F0) (Hz)</td>
<td>21.6</td>
<td>29.3</td>
<td>Old&gt;Young, not significant</td>
</tr>
<tr>
<td>Qx (%)</td>
<td>46.9</td>
<td>42.4</td>
<td>Young&gt;Old (t(28)=3.05, p&lt;0.01)</td>
</tr>
<tr>
<td>SD(Qx) (%)</td>
<td>4.75</td>
<td>4.56</td>
<td>Young&gt;old, not significant</td>
</tr>
</tbody>
</table>

Qx was significantly lower for the older speakers. F0, SD(F0), and SD(Qx) showed no significant difference in age. Values averaged across all seven phrases for the two age groups can be seen in Table 5.3. The results for Qx and the trends for F0 and SD(F0) are consistent with our hypothesised results, although the results pertaining to F0 and SD(F0) did not achieve statistical significance. Even if it were significant, the increased SD(F0) for older speakers would not necessarily correspond to changes in vocal stability, instead pointing toward older
speakers using a larger prosodic range within the recordings. Following from this, the question of whether Qx and its standard deviation are largely determined by prosodic features needs to be investigated.

The standard deviations of F0 and Qx show how much F0 and Qx both vary within each phrase. Most of the variations can be explained by prosodic variation, but it is possible that there are other factors causing the variation of F0 and Qx. To some extent, both the phonetic content and voice quality differences between speakers are likely to produce some variation. Regarding phonetic effects, vowels have different inherent F0 and loudness features (Black, 1949, Ewan, 1975, Shadle et al., 1979, O'Shaughnessy and Allen, 1983, Whalen and Levit, 1995, Maddieson, 2006, Fox and Jacewicz, 2014), and as such, might also have inherent Qx features. The sustained vowel analysis shows that Qx varies with F0 and loudness, and it is entirely possible that other aspects of voice production relate to differences in the Qx.

5.6. Contour analysis

The analyses in the previous three sections deal with static measures derived from each of the segments analysed. These static measures are essentially providing simplified representations of the dynamic behaviour occurring during phonation. Both the F0 and Qx change over the duration of a voiced speech segment, and while the extent of these changes is captured in part by the standard deviation, a more direct measure of the dynamic behaviour should tell us more.

To this end, the shape of the F0 and Qx contours for the segments from each of the three recording types were described using discrete cosine transform (DCT) coefficients. The DCT coefficients calculated provide a description of the overall shape of a given waveform by describing it in terms of a series of cosine waveforms of increasing frequency. DCT coefficients can be used to describe contours of speech features, such as in Watson and Harrington (1999), Teutenberg et al. (2008), and Bier et al. (2011). Lower order coefficients provide broad descriptors of the shape of a contour; the 1st coefficient corresponds to the mean of a contour, the 2nd coefficient corresponds to the overall slope of the contour, and the 3rd coefficient corresponds to the overall peakiness of the contour. Higher order coefficients provide more detail of higher frequency oscillations, but these low order coefficients essentially provide a low passed representation that gives the general shape. The contours analysed in this paper are well described using low order DCT coefficients, as they are simple curves with no oscillation. Watson and Harrington (1999) used two DCT coefficients to describe formant trajectories,
whereas Bier et al (2011) used five DCT coefficients to describe phrasal F0 contours. The contours analysed in the current paper each correspond to a single vowel, so the descriptions analysed are based on the first three DCT coefficients. We expect differences in the first coefficient to correspond to expected differences in mean F0 and Qx. The second and third coefficients should provide information about the trajectories of F0 and Qx within each vowel. Given that larger standard deviations of F0 and Qx were observed for younger speakers in the earlier hVd vowel analysis, we would also expect the second and third coefficients to be larger for younger speakers. Contour analysis was applied to the hVd words and selected words from the rainbow passage.

5.6.1 Contour analysis of hVd words

Calculation of the F0 and Qx contours was performed using the same vowel boundaries as in Section 5.4. Normalised contours for each vowel can be seen in Figure 5.3. The x axis shows normalised time. As with the earlier hVd word analysis, the vowels with more lip rounding have higher Qx values than more open vowels. The original contours ranged from 6 to 54 vocal fold cycles, with a mean of 21.95 and a median of 20. Each contour was interpolated to 20 data points, and then averaged across hVd vowel. Figure 5.5 shows a clear difference in the shape of the F0 and Qx contours for the young and old speakers. The younger speakers exhibited greater mean Qx and lower mean F0, and also exhibited a greater fall in F0 and rise in Qx over the duration of the vowels.

Two way ANOVA tests were performed on the first three DCT coefficients for the F0 and Qx contours, to test for differences in the means and slopes of these contours. The first DCT coefficient for F0 showed a significant difference with age (F(1,27)=5.5, p<0.05) and vowel (F(3.8,102.6)=32, p(GGe)<0.001). The second DCT coefficient for F0 failed to show a significant age difference, but did show a significant vowel difference (F(5.2,140.4)=3.8, p(GGe)<0.01). The third DCT coefficient for F0 showed no significant difference between age or vowel.
The first DCT coefficient of the Qx contour showed significant differences for both age (F(1,27)=5.9, p<0.05) and vowel (F(4.4, 118.8)=12.2, p(GGe)<0.001). The second DCT coefficient for the Qx showed significant differences for both age (F(1,27)=16.9, p<0.001) and vowel (F(4.6, 124.2)=8.1, pGGe<0.001). The third DCT coefficient of the Qx contour only showed a significant difference with age, (F(1,27)=6.14, p<0.05).

5.6.2 Contour analysis of rainbow passage words

For the examination of single vowels from the Rainbow Passage, initial phonetic segmentation of the recordings was performed using the Munich Automatic Segmentation System (MAUS) web interface (http://www.bas.uni-muenchen.de/Bas/BasMAUS.html, Schiel,(1999)), followed by hand corrections where necessary. MAUS has a specific database adapted to New Zealand English which was used for the segmentation.

In order to perform analysis with reduced coarticulation variance, eight instances of the syllable “rain” were identified within the Rainbow Passage for each speaker (see Appendix C). While the rain vowel may not be directly comparable to the hVd word lists as it is a diphthong, it has the advantage of being replicated many times throughout the Rainbow Passage with consistent phonetic surroundings. Given that /e/ and /i:/ were not significantly different in F0 or Qx in the
Development and comparison of dynamic measures of voice stability

hVd analysis (see Section 5.4), there should not be any inherent contour within the /ei/ vowel that could cloud contour differences due to age.

F0 and Qx contours were calculated for analysis. In order to limit the effect of prosodic differences, the second DCT coefficient of the F0 contour was used to determine whether each segment was positively or negatively sloped, and the positive and negatively sloped segments were treated as a within factor in the statistical analysis of the Qx contours. Of the 240 segments analysed, 46 were downward sloped and 194 were upward sloped. Both the upward and downward sloped segments were evenly split between young and old speakers. However, the split of upward and downward sloped segments within speakers was uneven, so robust statistical assessment based on F0 slope was not possible. As such, the split is maintained here for viewing purposes only. Average normalised contours for these segments can be seen in Figure 5.4. The x axis shows normalised time. As with Figure 5.5, each contour was interpolated to 20 data points before being averaged according to the F0 slope. The younger speakers exhibited the higher Qx we would expect in both the upward and downward sloped segments. The F0 slope also appears to have little effect on the shape of the Qx contours.

A series of analysis of variance tests were performed on the first three DCT coefficients for each of the F0 and Qx contours. Age was a between subject variable for each of these tests. The first DCT coefficient for Qx showed the expected significant difference with age (F(1,28) = 8.8, p<0.01). Examination of Figure 5.6 suggests that younger speakers tend to have a greater

![Figure 5.4: Normalised F0 and Qx contours for upward and downward sloped F0 samples of the syllable “rain.”](image-url)
Development and comparison of dynamic measures of voice stability

downward slope in Qx for rain vowels, and analysis of variance with the second DCT coefficient for the Qx contour showed this to be significant ($F(1,28) = 10.5$, $p<0.01$). Higher DCT coefficients showed no differences between the age groups. None of the DCT coefficients describing the F0 contours showed any significant age differences.

Table 5.8: F0 DCT coefficient comparison between age groups

<table>
<thead>
<tr>
<th>DCT coefficient</th>
<th>Young</th>
<th>Old</th>
<th>Post hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 DCT1</td>
<td>155 (SD=26)</td>
<td>167 (SD=37)</td>
<td>Old&gt;Young, not significant</td>
</tr>
<tr>
<td>F0 DCT2</td>
<td>-2.45 (SD=5.2)</td>
<td>-2.3 (SD=5.4)</td>
<td>Old&gt;Young, not significant</td>
</tr>
<tr>
<td>F0 DCT3</td>
<td>0.5 (SD=3.7)</td>
<td>-0.8 (SD=3.4)</td>
<td>Young&gt;Old, not significant</td>
</tr>
</tbody>
</table>

Table 5.9: Qx DCT coefficient comparison between age groups.

<table>
<thead>
<tr>
<th>DCT coefficient</th>
<th>Young</th>
<th>Old</th>
<th>Post hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qx DCT1</td>
<td>78.0 (SD=9.3)</td>
<td>69.3 (SD=10.3)</td>
<td>Young&gt;old ($t(28)=3.0$, $p&lt;0.01$)</td>
</tr>
<tr>
<td>Qx DCT2</td>
<td>1.5 (SD=2.0)</td>
<td>0.14 (SD=1.7)</td>
<td>Young&gt;old ($t(26.4)=3.3$, $p&lt;0.01$)</td>
</tr>
<tr>
<td>Qx DCT3</td>
<td>0.07 (SD=1.3)</td>
<td>-0.002 (SD=0.9)</td>
<td>Young&gt;Old, not significant</td>
</tr>
</tbody>
</table>

5.7. Discussion

It is clear that the different elicitation methods provide varying results when comparing young and old speakers. The sustained vowels in this study failed to show the hypothesised rise in F0 with older speakers, while both the hVd words and connected speech did exhibit the hypothesised increase. Variation between age groups for SD(F0) and SD(Qx) was inconsistent between elicitation types and lacked statistical significance. The most consistent measure across this study was Qx, which was lower in older speakers for all three sets of recordings. This shows that Qx is a more consistent indicator than F0 of the age-related voice differences evident in this cohort of speakers. The sustained vowel recordings also suggest that Qx is related to loudness, although this relationship appears to be dependent on age, as the older speakers have significantly lower Qx without also having any significant difference in SPL in the analysis in Chapter 4.
The differing results for SD(F0) as calculated by Speech Studio in the three different elicitation methods show that it is really relating to different things for each of the elicitation methods. The phrasal analysis produced SD(F0) values mostly in the range of 10 to 40Hz, whereas the sustained vowel analysis produced SD(F0) values in the region of 1 to 2Hz. For the sustained vowels, participants were trying to maintain a constant pitch, so the SD(F0) is a measure of how stable they were at maintaining that pitch. These SD(F0) values were lower than for the other elicitation methods (1-2Hz), showing some degree of ability to hold pitch within the participants. For the word lists, there was no specified pitch that participants were attempting to hold, so each word had a pitch range. The SD(F0) becomes a measure of that pitch range over each word, and these were in the order of 5-8 Hz. The connected speech, understandably, provides the largest SD(F0) measures (20-30 Hz) as it is measuring pitch variation over the course of an prosodic phrase. The same measure is effectively providing three different pieces of information in the three different elicitations. The SD(Qx) has similar issues, with the sustained vowel results being the only analysis in which it is useful as a stability measure.

Ultimately, the standard deviations of F0 and Qx are of limited use when assessing speech stability, given how dependent they are on elicitation type. Any age-related differences observed need to be tempered by understanding how the elicitation method is affecting the results. More direct examination of cycle to cycle variation of each of these measures would be more informative regarding changes in vocal stability that occur with age. Both the jitter (F0 perturbation) and a perturbation measure of Qx were examined in Chapter 4 for the sustained vowel portion of these recordings, and the Qx perturbation showed an increase for the older speakers. Unfortunately, perturbation measures such as these are less useful in connected speech where the voiced segments are broken up, subject to prosodic variation, and are often relatively short. Many of the vowels in the hVd analysis contained fewer than 10 vocal fold cycles, and short vowels within the Rainbow Passage provided similarly small numbers of vocal fold cycles. Even if a running average were incorporated into the jitter calculation as in Schoentgen and Guchtenee (1995), the limited length of many voiced segments would prevent this from compensating for the prosodic variation. Coarticulation effects such as that evident with the hVd vowel analysis would also cloud any ability to assess stability with perturbation measures for short vowels.

This is where the value of contour based measurements becomes relevant. They allow for greater understanding of what the standard deviation measures represent. For instance, in Figure 5.3, it is clear that the younger speakers cover a greater range of Qx values in their hVd
Development and comparison of dynamic measures of voice stability

vowels, but it is also clear that this is related to the underlying shape of the Qx contour, rather than the stability of Qx. The contour shape was also somewhat unaffected by vowel length; there was no clear pattern regarding the contours of long vowels as opposed to short vowels. Vowel duration is also unlikely to have affected the differences between age groups, as the duration varied much more between short and long vowels than between age group.

Given that lower Qx values are associated with breathiness, it is possible that the particular contour exhibited here is the result of a coarticulation effect, with the breathiness of the /h/ requiring a lower Qx, and the younger speakers having a higher target Qx for the voiced portion of the word. We can also see that the targets for the different vowels have an effect on the contours. The same general underlying shape was evident for both the long and short vowels, suggesting that the speed of convergence to each given vowel’s target Qx was dependent on the length of the vowel. If the convergence were in the same amount of time for all vowels, a distinct difference between long and short vowels would be evident in the time normalised contours shown in Figure 5.3.

Examining the contours for “rain,” the transition in Qx over the course of the vowel, is not as large, which may be because the consonants before and after the vowel are both voiced. If there were any coarticulation effect, one could speculate that the /r/ sound requires greater Qx, and that Qx then moves toward some form of target over the course of the vowel, based on the declining contour for the younger speakers. There is less to suggest this than with the hVd words, given that there is no apparent shift in Qx for the older speakers. Rather than being a coarticulation effect, the downward slope in Qx for the young speakers may be a behavioural quirk of the young speakers in this analysis. When comparing Figure 5.3 and Figure 5.4, there is some similarity in that the older speakers reach a Qx of approximately 50% at the end of each vowel, and the young speakers reach a Qx of approximately 55% at the end of each vowel. This does suggest some form of comfortable phonation target for Qx, although in both the hVd word and Rainbow Passage contours the apparent target is higher than the mean Qx values for normal loudness, normal pitch sustained vowels.

This indicates perhaps that the normal loudness, normal pitch sustained vowels are not necessarily as representative of normal phonation as one might hope. While there is still a lot that can be learned about the stability of voice production from sustained vowels, there may be some differences in the stability observed with sustained vowels to the stability present in connected speech. However, given the findings that the SD(Qx) in sustained vowels is lower
when the mean of Qx is higher, we can speculate that the stability present in connected speech will be greater than that of normal pitch, normal volume sustained vowels. Combining the results of all three elicitation types does allow us to better understand what is going on within each set of recordings, and as such, using all three sets of recordings has added to the picture in terms of helping describe the changes in voice production that occur with age.

5.8. Conclusions and future development

A combined approach of using multiple elicitation types is beneficial when examining voice quality changes with age. Sustained vowels provide useful measures of voice stability, while connected speech provides averages that are representative of a speaker’s natural speaking. The sustained vowel analysis showed greater standard deviations of F0 and Qx for older speakers, which was indicative of decreased vocal control with age. The connected speech and hVd vowels showed increased mean F0 and decreased mean Qx for older speakers, showing a change in vocal behaviour with age. Further examination using dynamic measures of F0 and Qx for vowels taken from word lists and connected speech provide more information about the changing behaviour of the vocal folds during phonation. Qx contours for the hVd vowels showed both old and young speakers exhibited a coarticulation effect which was more pronounced in the younger speakers than the older speakers. The younger speakers exhibited a greater transition of Qx over the hVd vowels to reach their naturally higher Qx values, expressing itself as a larger interaction between the phonemes. The tendency of older speakers to have lower mean Qx had the potential to diminish the visibility of this coarticulation effect, but was not enough to remove it.

In the future, examination of the effects of different phonemes on the Qx contour will allow further investigation into the prosodic and phonetic effects on age-related quality differences. This would be aided by a study using consistent framing to allow repeated analysis of the vowels and consonants of interest. Overall, the Qx has proven to be a useful metric for analysing age-related differences in voice quality, and research into how Qx interacts with age should be continued.
Chapter 6. Analysis considerations and future directions

While the electrographic data analysed in Chapter 4 and Chapter 5 have both shown useful results, there are limitations in the information that the EGG signal can provide. Section 3.1.3 identified some important considerations when discussing a contact quotient rather than a closed quotient when analysing EGG data, as the EGG signal itself does not provide certainty of complete closure of the vocal folds. To ascertain information pertaining to complete closure of the vocal folds, video laryngoscopy is a better method. It also provides further information about the vibratory patterns over EGG, particularly in the open phase when the extent of vocal fold contact is not changed but the glottal area is. However, video laryngoscopy is not without its limitations. It requires insertion of a camera via either the oral or nasal passages, which is not typical of normal speaking conditions. The camera used has much lower frame rates than those available in EGG, ranging from 24 Hz (used with stroboscopy) to typical frame rates of 5kHz (for high speed video). These lower sampling rates compromise the accuracy of perturbation based measurements (see Section 3.2.3) at best, and at worst make it impossible in the case of stroboscopy where the individual frames in the video are taken from separate vocal fold cycles. Stroboscopy also requires sustained vowels to allow the strobe to synchronise its speed with the F0 of the voice.

Both EGG and video laryngoscopy have immense value in voice analysis; EGG has the advantage of being less invasive and having higher sampling rates while video laryngoscopy has the advantage of directly displaying the vibratory behaviour of the vocal folds.

6.1. Electroglottography related to laryngoscopy

In order to highlight the strengths and limitations that each of these two recording methods bring to the table, two of the participants from the EGG study also provided stroboscopic recordings of sustained vowels, in a similar manner to those analysed in Chapter 4.

6.1.1 Procedure

Stroboscopic examination was performed by a qualified ENT surgeon, specialising in laryngology, according to standard procedure. The laryngoscope was inserted through the nose, so the vocal folds were best viewed with tongue fronted vowels. The young participant produced the vowels /a:/ and /i:/ at three pitches (low, normal, and high) from within his
comfortable speaking range and at three volume levels (quiet, normal and loud). This amounts to 18 sustained vowels for the young participant. The old participant only produced the /i:/ vowels, providing nine sustained vowels. As such the focus in this chapter is on the /i:/ vowels.

For each vowel there was a target of at least three seconds of recorded vibration. For each vowel, it was necessary for the camera to move closer after the onset of phonation in order to gain a clear view of the vocal folds. The video was recorded with concurrent audio. The video frame rate was 24 Hz.

6.1.2 Qualitative assessment of stroboscopic recordings

The stroboscopic nature of the recordings means that all the videos are aliased, and this aliasing is reliant on the accuracy of the strobing frequency. With that in mind, it is still possible to make qualitative assessments of the vibratory patterns and some simple estimation based on the visible apparent vibratory pattern. While there may not be enough quantitative data for any statistical analysis, it is still possible to use the estimations that can be made to reaffirm the results from the electroglottographic studies.

Analysis of the laryngoscopic video was performed by extracting individual frames using VLC media player (VideoLan, 2016). In recording a vowel, there is an initial period after phonation commences in which the camera is moved into an appropriate position, while the strobing also starts to correctly track the pitch of the voice. As such, the first stably tracked full cycle of vibration for each vowel was used for analysis.

The apparent vibratory frequency of the strobbed video was 1 Hz, giving 24 frames per cycle. A series of still shots from the young speaker showing an apparent vibratory cycle of the vowel /i:/ at varying pitch and loudness levels can be seen in Figure 6.1 through to Figure 6.6. Each of these figures features a vibratory cycle with the images proceeding in a clockwise direction from the top right image, and every third frame was used in these figures to provide a total of 8 frames to cover a vocal fold cycle. The images are related to an EGG cycle for the same speaker producing the vowel /i/ with the same target pitch and loudness, showing the approximate position in the vocal fold cycle for each laryngoscopic image. These are only approximate positions as each laryngoscopic image is actually from a separate vocal fold cycle. The EGG cycle used was from the recordings analysed in Chapter 4, and a 40 ms excerpt of each of these recordings can be seen in Figure 6.8.
Analysis considerations and future directions

Only the normal pitch and high pitch vowels are shown here; at low pitch, the strobe had greater difficulty latching on to the speaker’s F0 so there were not any complete apparent vocal fold cycles.

There are some clear differences apparent in these figures. Both of the loud vowel productions seen in Figure 6.3 and Figure 6.6 show fewer frames with an open glottis than the quieter vowel productions. Relating this to EGG recordings in Figure 6.8, we can see a greater contact quotient with increased vocal loudness. The maximum width of the glottal opening is also larger than seen with the quiet vowel productions in Figure 6.1 and Figure 6.2. These two quiet vowel productions also exhibit the longest open phases.

In Figure 6.1 and Figure 6.2, there is a slight posterior chink evident, with full closure of the vocal folds not being achieved. This would suggest that full closure is more likely to be obtained with both higher target loudness and higher target pitch. For this to be apparent with stroboscopy, the posterior chink would have needed to be consistent across the duration of the phonation. The speaker in this particular set of recordings also reported feeling that they had greater vocal control with higher target loudness and pitch, so the completeness of vocal fold closure may have some correlation with vocal control.

The posterior chink evident in these images also nicely demonstrates the difficulty with choosing appropriate naming conventions for EGG derived measures of the different phases of vocal fold vibration discussed in Section 3.1.3. If the term closed quotient were used rather than contact quotient for these three examples, it would not strictly be accurate, because the vocal folds are never fully closed. The EGG does, however, provide some information not seen in these figures; because the view of the vocal folds in laryngoscopy is purely top down, it provides no measure of the full extent of contact below the top of the vocal folds during the contact phase of the vibratory cycle.
Figure 6.1: One vocal fold cycle from the young speaker producing the vowel /i:/ at normal pitch and low intensity. The images proceed in a clockwise manner.

Figure 6.2: One vocal fold cycle from the young speaker producing the vowel /i:/ at normal pitch and normal intensity. The images proceed in a clockwise manner.
Analysis considerations and future directions

Figure 6.3: One vocal fold cycle from the young speaker producing the vowel /i:/ at normal pitch and high intensity. The images proceed in a clockwise manner.

Figure 6.4: One vocal fold cycle from the young speaker producing the vowel /i:/ at high pitch and low intensity. The images proceed in a clockwise manner.
Analysis considerations and future directions

Figure 6.5: One vocal fold cycle from the young speaker producing the vowel /i:/ at high pitch and medium intensity. The images proceed in a clockwise manner.

Figure 6.6: One vocal fold cycle from the young speaker producing the vowel /i:/ at high pitch and high intensity. The images proceed in a clockwise manner.
When making similar observations for the older speaker, there were a few drawbacks. Unfortunately, the ends of the vocal folds were obstructed for the older speaker, preventing detection of the presence of a glottal chink, and limiting how close the camera was able to get. The lower strength and vocal control of the older speaker’s voice also resulted in greater difficulty for the strobe to latch on to the produced F0; given that the strobing attempts to produce an apparent F0 of 1 Hz, to obtain a full apparent vocal fold cycle in the video it is necessary to successfully track the F0 for at least one second.

As such, a full cycle is shown here for only one of the older speaker’s vowels, in Figure 6.7. This shows the vowel /i:/ at normal pitch and loudness, making it comparable to Figure 6.2. As the camera was not able to get as close for the older speaker, a zoomed in view was necessary, giving us less resolution of the vocal folds themselves. There was a slightly longer open phase in the laryngoscopic video for the older speaker in this instance, which is consistent with the EGG examples.
Another notable difference between the young and old speaker in this instance is the magnitude of the vocal fold movement. While the different distance between the camera and the vocal folds for the two speakers makes direct comparison difficult, the older speakers’ vocal folds appear to open wider than the vocal folds of the younger speaker, relative to the overall length of the vocal folds. Watching the vibration in motion, the older speaker exhibited looser vibration indicative of decreased muscle tone. There are also visible differences in the surrounding tissue, with the younger speaker’s tissue being smoother, with less swelling.

Figure 6.8: Electroglottographic waveforms for target pitch and loudness levels examined with laryngoscopy
6.1.3 Quantitative analysis of stroboscopic

The source video had a frame rate of 24 frames per second, and the strobing produced an aliased apparent vibration frequency of 1 Hz. While only every third frame from an apparent vocal fold cycle is shown here, allowing the entire cycle to be visualised in 8 frames, some numerical analysis was performed on the full 24 frames of each of these cycles. For each of the cycles analysed, the number of frames in which the vocal folds were in contact was divided by the total number of frames to give an estimate of Qx, which could then be compared to the results from Chapter 4 for each of these two speakers. These two sets of data are shown in Table 6.1 and Table 6.2.

Table 6.1: Estimated Qx from stroboscopic video analysis

<table>
<thead>
<tr>
<th>Vowel condition</th>
<th>Young /a:/</th>
<th>Young /i:/</th>
<th>Old /i:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid/soft</td>
<td>37.5%</td>
<td>33.3%</td>
<td>29.2%</td>
</tr>
<tr>
<td>Mid/normal</td>
<td>58.3%</td>
<td>54.2%</td>
<td>45.8%</td>
</tr>
<tr>
<td>Mid/loud</td>
<td>66.7%</td>
<td>66.7%</td>
<td>62.5%</td>
</tr>
<tr>
<td>High/soft</td>
<td>33.3%</td>
<td>37.5%</td>
<td>33.3%</td>
</tr>
<tr>
<td>High/normal</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>High/loud</td>
<td>66.7%</td>
<td>62.5%</td>
<td>58.3%</td>
</tr>
</tbody>
</table>

Table 6.2: Mean EGG derived Qx values for speakers in stroboscopic analysis.

<table>
<thead>
<tr>
<th>Vowel condition</th>
<th>Young /a:/</th>
<th>Young /i:/</th>
<th>Old /i:/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid/soft</td>
<td>39.4%</td>
<td>50.94%</td>
<td>36.58%</td>
</tr>
<tr>
<td>Mid/normal</td>
<td>39.8%</td>
<td>53.36%</td>
<td>46.27%</td>
</tr>
<tr>
<td>Mid/loud</td>
<td>36.63%</td>
<td>54.11%</td>
<td>61.59%</td>
</tr>
<tr>
<td>High/soft</td>
<td>40.39%</td>
<td>41.87%</td>
<td>35.93%</td>
</tr>
<tr>
<td>High/normal</td>
<td>36.79%</td>
<td>40.88%</td>
<td>41.6%</td>
</tr>
<tr>
<td>High/loud</td>
<td>44.19%</td>
<td>56.63%</td>
<td>49.93%</td>
</tr>
</tbody>
</table>

The estimated results based on the stroboscopy have a wider range than those from the EGG, with the trend of Qx rising with increased target volume being more pronounced. The older speaker also generally had lower mean Qx values, which reflected the age group difference seen in Chapter 4. The results from the EGG study itself for these two speakers were less pronounced, and in fact were not entirely consistent with the expected trends. It may be that the more invasive nature of the stroboscopic recordings enforced a bit more rigour on behalf of the speakers in achieving their target pitch and loudness levels. This does, however, indicate the need for larger numbers of speakers in a given study to compensate for the intraspeaker variance in voiced speech behaviour.
6.1.4 Comparison between electroglottographic and stroboscopic analysis

Based on the quantitative comparisons of estimated Qx between the electroglottographic and laryngoscopic recordings the overall patterns for Qx remained consistent. There was a clear demonstration in some of the young speaker’s laryngoscopic footage that complete glottal closure was not always evident.

6.1.5 Physiological reasons for observed differences between age groups

The stroboscopic video gives a visual indication of the physiological differences in the vocal between the two age groups analysed in this thesis. However, these physiological differences will still be manifesting themselves in the data from the EGG recordings.

The video footage made it clear that there was decreased elasticity in the vocal folds for the old speaker, leading to a much quicker transition between “fully closed” and “fully open” and a larger proportion of the time spent at the extremes of the vocal fold positions. In terms of how this is likely to affect EGG based measurement of the contact quotient, the duration initial opening of the glottis and final closing of the glottis are both shortened with age, which will contribute to a shorter overall contact phase as observed in the EGG results.

Decreased muscle tone and elasticity is also likely to contribute to an increase in variation of the duration of the contact phase, which would manifest itself in the increased CQP observed in the old speakers in the sustained vowel analysis.

6.2. Continued development of Qx as a voice behaviour measure

The studies in Chapters 4 and 5 both demonstrated ways in which EGG derived Qx was able to be used to quantify differences in vocal fold behaviour, both within and between the two speaker groups. The sustained vowel study in Chapter 4 used Qx and its perturbation to quantify differences between different target loudness levels, as well as revealing an age-related difference in vocal fold behaviour. The contour analysis in Chapter 5 used contours of the Qx to show differences in vocal fold behaviour over the duration of different vowels, as well as revealing a likely coarticulation effect at the onset of each hVd vowel.

In both of these cases, the Qx and its variation over the course of a voice segment provided information that could not be gleaned from the more common measure of F0 alone. The use of EGG in addition to the acoustic recording was minimally invasive for the participants in this research, but allowed for a much richer dataset using the analysis tools developed in Chapter.
3. While this dataset was collected with age group comparison in mind, the use of EGG derived Qx could just as easily be applied to other speaker comparisons. Its sensitivity to loudness and dynamics along with phonetic variation should make it an attractive measure for future use.

The Qx analysis methods presented in this thesis should also be useful for studies that do not incorporate EGG recordings. It is possible to develop an estimation of the open quotient from the acoustic waveform using linear predictive coding. The open quotient can be treated a bit like the inverse of the Qx, and as such, measuring its perturbation or transient change over the course of a voiced speech segment should provide similarly useful descriptions of vocal fold behaviour, provided it can be calculated accurately.

This examination of stroboscopic video in this chapter allowed visualisation of the vocal fold vibratory behaviour. Even with a rather limited dataset, there were observable differences between the vocal fold behaviour of the young and old speakers, as well as observable differences between phonations at varying target pitch and loudness levels. While these differences do not have statistical rigour to stand on their own, they do support the findings of the two EGG based studies in Chapters 4 and 5, and illustrate how the two different analysis approaches have different strengths. This helps to show the value of continued development of Qx based measures in analysis of voiced speech.

6.2.1 Expansion into studies with female speakers

The studies presented in this thesis focused on male speakers as a testing ground for the analysis methods developed. However, the methods will also be applicable to female speakers, and this is a good focus for further research.

Other studies that have investigated how Qx changes with age in both male and female speakers have found that the trend for the mean Qx is the opposite in female speakers compared to male speakers (Ma & Love, 2010, Higgins & Saxman, 1991), with mean Qx increasing with increased age in female speakers. This reversal in trend between male and female speakers for the mean Qx is likely to repeat itself, but the possible patterns for CQP and contour analysis as examined in Chapter 4 and Chapter 5 are less certain.

In terms of the physiological difference in ageing between men and women, there are different hormones involved that affect how the vocal folds change. That the ageing process affects the vocal folds differently in men and women is already evident in the differing patterns for age-related changes in mean Qx and F0. Perhaps the larger dimensions of the vocal folds in male
speakers would allow for greater variation in how the vocal folds come into contact each cycle, compared to female speakers. If this were the case, then we would expect to see greater CQP in old male speakers than in old female speakers, as well as larger slopes when analysing Qx contours. The methods developed within this thesis should function with female speakers as well as with male speakers. Judging by the fact that prior research has successfully analysed mean values of Qx for female speakers, it simply is a matter of pulling out the individual vocal cycles over the course of phonation. Once they have been obtained, the CQP and contour analysis will both be viable methods for analysing Qx.

Age is only one of many factors that influences voice quality, and the methods developed in this thesis should be applicable when examining the effects of other factors as well. Voice pathologies that alter the shape of the vocal folds (such as vocal nodules or cysts) force changes in how the vocal fold contact will vary over vocal fold cycles, and as such could benefit from analysis using EGG-derived measures. The CQP would be useful for investigating whether the altered shape of the vocal folds lead to a consistent altered vibratory pattern or introduces more randomness.

6.3. Continued research using recorded data set

In addition to providing a data set to test and develop methods for analysing voice quality, the recordings made for the research in this thesis were retained to provide a normative set of data for continued research into the impact of age on voice quality. This further research is already underway, and includes research into the perceptual identification of speaker age (Herath et al, 2017, under review). The study's primary aim was to investigate the extent to which a listener can perceive speaker age, with a secondary aim of identifying the factors (perceptual or acoustic) which play a role in determining the accuracy of listener estimation. In this study, an online survey was completed by 213 respondents, in which they listened to recorded samples and estimated the age of the speaker. The samples included the recorded hVd words and sample phrases from the Rainbow Passage. The samples chosen for this study were taken from five young and five old speakers from the wider dataset. The majority of the respondents were able to estimate the speakers' age within two decades of the chronological age of the speaker. The mean absolute error of the estimations was 9.2 years, taken across all respondents for all speakers. While previous training and/or employment in the field of voice were not found to play a role, speaker and listener age, the number of years lived in New Zealand, and the first language of the listener were all found to influence the accuracy of listeners. Acoustic measures
of the voice were also found to be good predictors of perceived age, with increased F0 and decreased Qx both tending to correspond to increases in the perceived age of the speakers.

The recordings have also been used in continued development of tools for the acoustic analysis of senescence in speech (Ben Dom, 2017). This study utilised the hVd word lists as a data set for analysis of glottal flow waveforms derived from the acoustic waveforms using LPC analysis. The glottal flow waveforms were then analysed using open quotient based measurements similar in nature to the Qx.

The open quotient measurements in this study showed an increase with age, which corresponds well to the decrease in Qx with age observed in Chapter 4 and Chapter 5 of this thesis.

6.4. Conclusions

Each of the different methods of recording speech in order to analyse voice quality has merits and drawbacks. Both laryngoscopy and EGG are able to provide useful results though, and a cursory analysis of a subset of speakers from the wider EGG studies presented in Chapter four and Chapter five indicated that results from stroboscopy will support the findings from the analysis of EGG signals.

As such, the EGG analysis techniques presented in this thesis provide a useful toolkit which may then be developed for further use in analysing the changes in voice quality that occur with age, and potentially the analysis of voice quality in areas other than ageing as well.

The recordings and the techniques developed in this thesis will both continue to be used for further research.
Chapter 7. Conclusions

This thesis presented new methods of analysing voice quality using EGG signals, and applied them to the context of analysing changes in voice quality that occur with age.

There are multiple software options available to analyse voice quality metrics, however, there is inconsistency in how some of these metrics are applied, and as such, there is inconsistency in the results produced by different programs, particularly in the case of perturbation. As such, any comparisons of these measures should be made within the same framework.

Following on from this, the significant contribution of this research was the development of algorithms to calculate F0 and contact quotient from electroglottographic recordings, and then analysing these both in terms of their perturbations and also in terms of describing their contours in speech.

Chapters four and five drew on a set of normative data for male New Zealand English speakers, recorded for the research in this thesis. Chapter four honed in on measuring the perturbation of the contact quotient, and using it to compare voice quality in sustained vowels between young and old speakers. This new perturbation measure increased with age, and demonstrated this age difference more strongly than the traditional perturbation measures of jitter and shimmer. In addition to the differences observed with age, the contact quotient perturbation, jitter and shimmer all varied with target loudness, with lower perturbation values being found for higher loudness levels.

As a measure of vocal fold stability, the contact quotient perturbation had a stronger age-related change than the other perturbation measures. It also has the advantage of being more directly related to the vibratory behaviour of the vocal folds themselves, and as such could be considered more informative than the jitter and shimmer.

Perturbation measures have often been used in voice quality analysis as a way of measuring the instability of aspects of the voice, and they proved useful in the analysis of sustained vowels presented in this thesis. However, sustained vowels can be considered to be more of a singing task than a speaking task. There are also larger scale fluctuations in the voice present in continuous speech, and so we developed the contour analysis presented in Chapter Five.
Conclusions

Combining analysis of multiple elicitation types was helpful in order to improve understanding of the changes in voice quality with age. Sustained vowels provide useful measures of voice stability when analysed on their own, but they need the addition of connected speech to anchor it in measurements that are representative of a speaker’s natural speaking.

Examination of the F0 and contact quotient contour proved informative. By analysing hVd word lists in this manner, vowel dependent differences in F0 and contact quotient behaviour were revealed. There appeared to be an additional strong coarticulation effect, with the /h/ sound requiring a low contact quotient, which then quickly rose over the course of the vowel. The extent of this rise was greater for the young speakers which allowed consistency of the higher contact quotient exhibited by younger speakers across this thesis.

Using the discrete cosine transform allowed relatable quantification of these shapes that were observed in the F0 and contact quotient contours. Measurement of the first coefficient allowed an examination of the differences in mean values. This was consistent with the other mean values throughout this thesis, with contact quotient being higher for younger speakers and F0 being higher for older speakers. The second DCT coefficient for the contact quotient contours showed the increased slope evident for the younger speakers, and the third coefficient contributed toward the evident rise being earlier in the vowel.

There was less to be learned from analysing the word “rain” in the connected speech, but the contour analysis was still useful. The F0 and contact quotient contours both differed between young and old speakers, and there was consistency in the contact quotient contour despite variance in the F0 contour due to the varied prosody of the tokens measured. This method of analysis certainly seems to be a valuable tool, and further work will include applying it to a greater number of vowels with different phonemic contexts.

The results from the EGG analysis was related to stroboscopic recordings of two of the speakers from the EGG study, which indicated that mean results from the two methods of recordings supported each other. However, in terms of analysing how Qx changes over time, the EGG recordings provided much greater temporal resolution, and it was the analysis of the dynamic nature of Qx that proved to be the most useful aspect in comparing voice quality between young and old speakers in chapters four and five.

Throughout this thesis, perhaps the most important outcome is the usefulness of the contact quotient as a measure of voice quality. Its perturbation was a stronger indicator of age-related
voice quality changes than the other perturbation measures analysed, and analysis of the contact quotient contour proved to be informative when examining voice quality in differing phonemic contexts. There will be continued research into making use of the contact quotient as an indicator of voice quality.
References


References


References


References


References


References


## Appendix A: Participant questionnaire summary

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Age</th>
<th>Voice training</th>
<th>Current fitness</th>
<th>Smoking (current/respiratory)</th>
<th>Illness</th>
<th>Voice disorders</th>
<th>Hearing loss</th>
<th>Tinnitus</th>
<th>Other complaints</th>
<th>Head size</th>
</tr>
</thead>
<tbody>
<tr>
<td>y81</td>
<td>25</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y82</td>
<td>34</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y83</td>
<td>23</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>37 cm</td>
</tr>
<tr>
<td>y84</td>
<td>25</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>39 cm</td>
</tr>
<tr>
<td>y85</td>
<td>25</td>
<td>moderate</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>41 cm</td>
</tr>
<tr>
<td>y86</td>
<td>24</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>36 cm</td>
</tr>
<tr>
<td>y87</td>
<td>75</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>35.5 cm</td>
</tr>
<tr>
<td>y88</td>
<td>25</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>39 cm</td>
</tr>
<tr>
<td>y89</td>
<td>24</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y90</td>
<td>36</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y91</td>
<td>25</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y92</td>
<td>35</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>39 cm</td>
</tr>
<tr>
<td>y93</td>
<td>30</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>37 cm</td>
</tr>
<tr>
<td>y94</td>
<td>25</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>36 cm</td>
</tr>
<tr>
<td>y95</td>
<td>25</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>36 cm</td>
</tr>
<tr>
<td>y96</td>
<td>70</td>
<td>moderate</td>
<td>no</td>
<td>a little</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y97</td>
<td>75</td>
<td>moderate</td>
<td>no</td>
<td>a little</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>38 cm</td>
</tr>
<tr>
<td>y98</td>
<td>65</td>
<td>good</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>43 cm</td>
</tr>
<tr>
<td>y99</td>
<td>60</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>44 cm</td>
</tr>
<tr>
<td>y100</td>
<td>55</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>39 cm</td>
</tr>
<tr>
<td>y101</td>
<td>50</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>36 cm</td>
</tr>
<tr>
<td>y102</td>
<td>50</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>39 cm</td>
</tr>
<tr>
<td>y103</td>
<td>40</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>42 cm</td>
</tr>
<tr>
<td>y104</td>
<td>40</td>
<td>moderate</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>48 cm</td>
</tr>
<tr>
<td>y105</td>
<td>50</td>
<td>moderate</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>36 cm</td>
</tr>
</tbody>
</table>

98
Appendix B

Appendix B: Ethics Approval notifications

Initial ethics application:

UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE

27-Sep-2011

MEMORANDUM TO:

Dr Catherine Watson
Electrical & Computer Engineer

Re: Application for Ethics Approval (Our Ref. 7577)

The Committee considered your application for ethics approval for your project titled *Age related acoustic and electroglottographic characteristics of speakers of normal New Zealand English* on 23-Sep-2011.

Ethics approval was given for a period of three years.

The expiry date for this approval is 23-Sep-2014.

If the project changes significantly you are required to resubmit a new application to the Committee for further consideration.

In order that an up-to-date record can be maintained, you are requested to notify the Committee once your project is completed.

The Chair and the members of the Committee would be happy to discuss general matters relating to ethics approvals if you wish to do so. Contact should be made through the UAHPEC secretary at humanethics@auckland.ac.nz in the first instance.

All communication with the UAHPEC regarding this application should include this reference number: 7577.

(This is a computer generated letter. No signature required.)

Secretary
University of Auckland Human Participants Ethics Committee
c.c. Head of Department / School, Electrical & Computer Engineer
Mr Stephen Birr
Prof Allan Williamson
Clare McCann

Additional information:
1. Should you need to make any changes to the project, write to the Committee giving full details including revised documentation.
2. Should you require an extension, write to the Committee before the expiry date giving full details along with revised documentation. An extension can be granted for up to three years, after which time you
must make a new application.

3. At the end of three years, or if the project is completed before the expiry, you are requested to advise the Committee of its completion.

4. Do not forget to fill in the ‘approval wording’ on the Participant Information Sheets and Consent Forms, giving the dates of approval and the reference number, before you send them out to your participants.

5. Send a copy of this approval letter to the Manager - Funding Processes, Research Office if you have obtained funding other than from UniServices. For UniServices contract, send a copy of the approval letter to: Contract Manager, UniServices.

6. Please note that the Committee may from time to time conduct audits of approved projects to ensure that the research has been carried out according to the approval that was given.
Appendix B

Amendment to include laryngoscopy:

UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE

07-Jun-2012

MEMORANDUM TO:
Dr Catherine Watson
Electrical & Computer Engineer

Re: Application for Ethics Approval (Our Ref. 7983)

The Committee considered your application for ethics approval for your project entitled Age related laryngoscopic characteristics of speakers of New Zealand English.

Ethics approval was given for a period of three years.
The expiry date for this approval is 07-Jun-2015.

If the project changes significantly, you are required to submit a new application to UAHPEC for further consideration.

In order that an up-to-date record can be maintained, you are requested to notify UAHPEC once your project is completed.

The Chair and the members of UAHPEC would be happy to discuss general matters relating to ethics approvals if you wish to do so. Contact should be made through the UAHPEC Ethics Administrators at humanethics@auckland.ac.nz in the first instance.

All communication with the UAHPEC regarding this application should include this reference number: 7983.

(This is a computer generated letter. No signature required.)

UAHPEC Administrators
University of Auckland Human Participants Ethics Committee

c.c. Head of Department / School, Electrical & Computer Engineer
Prof Allan Williamson
Clare McCann
Mr Stephen Bier

Additional information:
1. Do not forget to fill in the 'approval wording' on the Participant Information Sheets and Consent Forms, giving the dates of approval and the reference number, before you send them out to your participants.
2. Should you need to make any changes to the project, write to the UAHPEC Administrators by email (humanethics@auckland.ac.nz) giving full details of the proposed changes including revised documentation.

101
Appendix B

3. At the end of three years, or if the project is completed before the expiry, please advise UAHPEC of its completion.

4. Should you require an extension, write to UAHPEC by email before the expiry date, giving full details along with revised documentation. An extension can be granted for up to three years, after which a new application must be submitted.

5. If you have obtained funding other than from UniServices, send a copy of this approval letter to the Manager - Funding Processes, UniA Research Office. For UniServices contracts, send a copy of the approval letter to the Contract Manager, UniServices.

6. Please note that UAHPEC may from time to time conduct audits of approved projects to ensure that the research has been carried out according to the approval that was given.
Appendix B

Allowing concurrent measurement of airflow

UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE

15-Feb-2013

MEMORANDUM TO:
Dr Catherine Watson
Electrical & Computer Engineer

Re: Application for Ethics Approval (Ref. 7577)
The Committee considered your request for change for your project titled Age related acoustic and electrophotographic characteristics of speakers of normal New Zealand English on 15-Feb-2013.
The Committee approved the following amendments:
1. To add an airflow measurement to the study

The expiry date for this approval is 23-Sep-2014.
If the project changes significantly you are required to resubmit a new application to the Committee for further consideration.
In order that an up-to-date record can be maintained, it would be appreciated if you could notify the Committee once your project is completed.
The Chair and the members of the Committee would be happy to discuss general matters relating to ethics approvals if you wish to do so. Contact should be made through the UAHEC secretary at humanethics@auckland.ac.nz in the first instance.

All communication with the UAHEC regarding this application should include this reference number: 7577.

(This is a computer generated letter. No signature required.)

Secretary
University of Auckland Human Participants Ethics Committee
c.c. Head of Department / School, Electrical & Computer Engineer
Prof Allan Williamson
Clara McCann
Mr Stephen Bier
Appendix C: Experiment reading materials

Rainbow Passage:

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows. Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of super-imposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.

hVd Word Lists:

<table>
<thead>
<tr>
<th>List 1</th>
<th>List 2</th>
<th>List 3</th>
<th>List 4</th>
<th>List 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hood</td>
<td>Hud</td>
<td>Hud</td>
<td>Hood</td>
<td>Herd</td>
</tr>
<tr>
<td>Hard</td>
<td>Hod</td>
<td>Head</td>
<td>Hod</td>
<td>Hood</td>
</tr>
<tr>
<td>Hud</td>
<td>Hid</td>
<td>Hoard</td>
<td>Herd</td>
<td>Hud</td>
</tr>
<tr>
<td>Hid</td>
<td>Hard</td>
<td>Had</td>
<td>Heed</td>
<td>Head</td>
</tr>
<tr>
<td>Herd</td>
<td>Herd</td>
<td>Heed</td>
<td>Hard</td>
<td>Head</td>
</tr>
<tr>
<td>Who'd</td>
<td>Had</td>
<td>Hod</td>
<td>Had</td>
<td>Had</td>
</tr>
<tr>
<td>Hoard</td>
<td>Who'd</td>
<td>Who'd</td>
<td>Hid</td>
<td>Hard</td>
</tr>
<tr>
<td>Hod</td>
<td>Head</td>
<td>Hid</td>
<td>Hoard</td>
<td>Hod</td>
</tr>
<tr>
<td>Hard</td>
<td>Hoard</td>
<td>Hard</td>
<td>Who'd</td>
<td>Hoard</td>
</tr>
<tr>
<td>Had</td>
<td>Heed</td>
<td>Herd</td>
<td>Head</td>
<td>Who'd</td>
</tr>
<tr>
<td>Head</td>
<td>Hod</td>
<td>Head</td>
<td>Hod</td>
<td>Herd</td>
</tr>
<tr>
<td>Heed</td>
<td>Hood</td>
<td>Hood</td>
<td>Hud</td>
<td>Hid</td>
</tr>
<tr>
<td>Hid</td>
<td>Hard</td>
<td>Hod</td>
<td>Who'd</td>
<td>Heed</td>
</tr>
</tbody>
</table>