COMPARING VOWEL FORMANT NORMALISATION PROCEDURES

NICHOLAS FLYNN

Abstract

This article compares 20 methods of vowel formant normalisation. Procedures were evaluated depending on their effectiveness at neutralising the variation in formant data due to inter-speaker physiological and anatomical differences. This was measured through the assessment of the ability of methods to equalise and align the vowel space areas of different speakers. The equalisation of vowel spaces was quantified through consideration of the SCV of vowel space areas calculated under each method of normalisation, while the alignment of vowel spaces was judged through considering the intersection and overlap of scale-drawn vowel space areas. An extensive dataset was used, consisting of large numbers of tokens from a wide range of vowels from 20 speakers, both male and female, of two different age groups. Normalisation methods were assessed. The results showed that vowel-extrinsic, formant-intrinsic, speaker-intrinsic methods performed the best at equalising and aligning speakers' vowel spaces, while vowel-intrinsic scaling transformations were judged to perform poorly overall at these two tasks.

1. Introduction

The process of normalising vowel formant data to permit accurate cross-speaker comparisons of vowel space layout, change and variation, is an issue that has grown in importance in the field of sociolinguistics in recent years. A plethora of different methods and formulae for this purpose have now been proposed.

Thomas & Kendall (2007) provide an online normalisation tool, "NORM", a useful resource for normalising formant data, and one which has opened the viability of normalising to a greater number of researchers. However, there is still a lack of agreement over which available algorithm is the best to use.

This article aims to add to the normalisation literature by comparing a large number of the normalisation procedures available, evaluating their effectiveness at neutralising variation in vowel formant data due to inter-speaker physiological and anatomical differences.

In section 2 the rationale behind normalising vowel formant data is explained, before section 3 presents the formulae of existing methods, and section 4 recounts the findings of other studies that have compared different procedures. Section 5 describes the methodology used for this study and provides information about how the effectiveness of procedures was tested. Section 6 gives the results of the comparisons between the different methods, before section 7 discusses the relative effectiveness of each procedure and compares the findings to those of previous comparative studies of normalisation methods.

Throughout this article the following notation is used: F_i represents a formant, (i = 1,2,3). A superscript ^{N} is used to denote a normalised value. For example, F_1^N is the normalised value of the first formant of a token.

2. The concept of vowel formant normalisation

A major problem faced by researchers in sociophonetic variation is that no two speakers' vowel tracts share the same dimensions. As a consequence, the "same" phonological vowel uttered by different speakers will show formants at different frequencies due to the different

sizes of the speakers' vocal tracts. For example, female speakers tend to display higher formant frequencies than male speakers, as their vocal tracts are shorter and thus their resonance frequencies are higher. It can be difficult, then, when comparing the positioning of vowels within speakers' vowel spaces, to identify whether differences in formant values are due to a linguistic change in the vowel system, or are merely due to the anatomical and physiological differences between speakers.

It has been acknowledged that the raw Hertz formant frequencies of different speakers are not directly comparable, and that it is not ideal to plot formant values in Hertz from different speakers on the same formant chart (Watt et al. 2010). This presents a problem for sociophonetic research that seeks to describe variation and change through the comparison of speech from different speakers.

The solution is, in principle, to remove as much of the inter-speaker formant value differences due to biological differences as possible. This would leave quantities unaffected by the size of a speaker's vocal tract, and so would be directly comparable. The process of transforming formant frequencies to make them directly comparable with those from other speakers is called Vowel Formant Normalisation.

A number of differing formulae have been put forward as normalising algorithms. The sheer number of proposed normalisation algorithms indicates a distinct lack of consensus about how best to normalise. However, those researchers who have considered the process of normalisation have collectively identified a number of goals of normalisation:

1. to minimise or eliminate inter-speaker variation due to inherent physiological or anatomical differences;
2. to preserve inter-speaker variation due to social category differences, including age, gender and dialect, or due to sound change;
3. to maintain vowel category and phonemic differences;
4. to model the cognitive processes that allow human listeners to normalise vowels uttered by different speakers.

(Hindle 1978; Disner 1980; Thomas 2002; Langstrof 2006; Thomas & Kendall 2007; Fabricius 2008; Clopper 2009; Watt et al. 2010).

Of course, such goals are somewhat idealistic, and the unlikelihood of any normalisation method perfectly fulfilling all the above criteria has been acknowledged (Thomas 2002; Adank et al. 2004; Bigham 2008; Thomas & Kendall 2007). For example, it has been observed that through reducing physiological variation, sociolinguistic variation can also be reduced (Adank et al. 2004).

Some researchers may place greater importance on one criterion over the others. This will largely depend on the nature of the study. For example, perception-based studies want normalisation to approximate the process of human vowel perception as closely as possible (Rosner & Pickering 1994; Syrdal & Gopal 1986), while sociophonetic studies are less concerned with this, but place greater importance on the maintenance of sociophonetically-relevant information, such as age-based variation (Fabricius 2008; Watt et al. 2010; Langstrof 2006; Thomas & Kendall 2007).

Thomas (2002) makes the excellent observation that

> all normalisation techniques have drawbacks, […] choosing which normalisation technique to use is a matter of deciding which drawbacks are tolerable for the study at hand.

Thomas (2002:174)

The onus, then, appears to be on the researcher to choose from the numerous posited methods, a normalisation procedure that is appropriate for the type of study and its research objectives.

3. Existing normalisation formulae

Normalisation procedures have traditionally been categorised according to whether they are vowel intrinsic or extrinsic, formant intrinsic or extrinsic, speaker intrinsic or extrinsic, or a combination of these six categories (Adank 2003; Adank et al. 2004; Thomas & Kendall 2007; Clopper 2009; Fabricius et al. 2009; Watt et al. 2010).

Vowel-intrinsic techniques use information from a single vowel token, while vowel-extrinsic techniques use information from multiple vowels, often across several vowel categories to normalise a formant value. Formant-intrinsic procedures normalise a formant value using information from occurrences of that formant only, using F_1 measurements to normalise an F_1 value, for example, while formant-extrinsic procedures use information from multiple formants, for example using F_1, F_2 and F_3 measurements to normalise an F_1 value. Speaker-intrinsic methods use information from a single speaker, while speaker-extrinsic methods use information from a population.

Speaker-extrinsic procedures are rarely used, due to their complexity, and the fact that by their very nature, adding more speakers into a dataset will alter the normalised values, meaning that any calculations that have already been made must be discounted and redone from scratch. Despite this disadvantage, Labov et al. (2006), a recent major piece of American English sociolinguistic work, used a speaker-extrinsic normalisation procedure as part of its methodology.

A number of speaker-intrinsic normalisation procedures will now be presented, beginning with vowel-intrinsic methods. These can be regarded as a rescaling of Hertz frequencies onto a different scale, and were originally developed to closer model human vowel perception by transforming frequencies onto a more perceptually-relevant scale (Adank 2003; Adank et al. 2004).

A number of scales have been proposed, including the Mel scale, obtained via (1) (Stevens & Volkmann 1940), Equivalent Rectangular Bandwidth (ERB), obtained via (2) (Glasberg & Moore 1990) and Bark, obtained using (3) (Traunmüller 1990).

(1) $$F_i^N = 1127 \ln.\left(1 + \frac{F_i}{700} \right)$$

(2) $$F_i^N = 21.4 \ln.(0.00437F_i + 1)$$

(3) $$F_i^N = 26.81 \left(\frac{F_i}{1960 + F_i} \right) - 0.53$$

Bladon et al. (1984) extended the Bark transformation to normalise female speakers relative to male speakers using (4). Clopper (2009) makes the justifiable criticism that subtracting an extra 1 Bark for female speakers appears an arbitrary choice.

$$(4)\qquad F_i^N = \begin{cases} 26.81\left(\frac{F_i}{1960 + F_i}\right) - 0.53 & \text{, speaker is male} \\ \left(26.81\left(\frac{F_i}{1960 + F_i}\right) - 0.53\right) - 1 & \text{, speaker is female} \end{cases}$$

Syrdal & Gopal (1986) used the Bark transformation to create a formant-extrinsic, vowel-intrinsic method called the "Bark-Distance Method". Their method is based on their observation that the distance between neighbouring formants is similar across speakers (Syrdal & Gopal 1986). The equations used to normalise using Syrdal & Gopal's Bark-Distance Method are:

$$(5)\qquad F_1^N = F_1^{BARK} - F_0^{BARK}$$

$$(6)\qquad F_2^N = F_3^{BARK} - F_2^{BARK}$$

The correlates of vowel height and vowel frontness for F_1 and F_2 are maintained using these measurements (Syrdal & Gopal 1986).

Miller (1989) proposed scaling formant frequencies to a scale better aligned with perceptual differences by taking the (natural) logarithm of the Hertz value. That is,

$$(7)\qquad F_i^N = \ln.(F_i)$$

He then extended this concept, by suggesting a formant ratio model, broadly given in (8). *SR* is defined as the "Sensory Reference", a speaker-specific value based on the average fundamental frequency of the speaker whose frequencies are being normalised, and of all speakers in the sample (Miller 1989).

$$(8)\qquad F_i^N = \begin{cases} \left(\frac{\ln.(F_i)}{\ln.(F_{i-1})}\right) & \text{, } i > 1 \\ \left(\frac{\ln.(F_i)}{SR}\right) & \text{, } i = 1 \end{cases}$$

Miller's formant ratio method is an example of a speaker-extrinsic, formant-extrinsic, vowel-extrinsic procedure, as *SR* is derived from measurements taken from a population of speakers. Another speaker-extrinsic, formant-extrinsic, vowel-extrinsic procedure has been proposed by Nordström (1977).

Nordström's method uses (9) to scale formant values based on using an estimation of the difference between male and female vocal tract length to transform female speakers' values relative to males'.

$$(9)\qquad F_i^N = \begin{cases} F_i & \text{, speaker is male} \\ \left(\frac{\mu_{F_3}^{male}}{\mu_{F_3}^{female}}\right) F_i & \text{, speaker is female} \end{cases}$$

Here, $\mu_{F_3}^{male}$ and $\mu_{F_3}^{female}$ are the mean F_3 of all tokens with $F_1 > 600$Hz for all male speakers and all female speakers in the sample respectively.

Many formulae exist that are speaker-intrinsic, formant-intrinsic, and vowel-extrinsic. An early example which is recounted in Lobanov (1971), was the linear compression and expansion method, which scales a speaker's formant values relative to their maximum value for that formant.

$$(10)\quad F_i^N = \frac{F_i}{F_i^{max}}$$

This allows comparability between speakers, as formant values are represented as a proportion of a speaker's maximum formant frequency. In effect, speakers' vowel spaces are aligned by anchoring them at the maximum values for individual formants. Gerstman's (1968) method builds on this methodology by aligning speakers' vowel spaces at both endpoints of their formant frequency range. Values are scaled so that the values of the extremities are 0 and 999 rather than 0 and 1. The equation for Gerstman's method is given in (11).

$$(11)\quad F_i^N = 999\left(\frac{F_i - F_i^{min}}{F_i^{max} - F_i^{min}}\right)$$

The maximum and minimum F_i in (10) and (11) are taken from all vowel tokens for a speaker.

The previous two normalisation procedures express values relative to the extremities of a speaker's formant frequency range. The next procedure, developed by Watt & Fabricius (2002) expresses values relative to the constructed centroid of a speaker's vowel space.

In this technique, a speaker's vowel space is thought of as a triangle, with the apices at points representing the minimum and maximum F_1 and F_2 for the speaker, and labelled [i], [a] and [u′]. The coordinates of [i] are the speaker's mean F_1 and F_2 for vowels taken from their FLEECE lexical set, which are used to represent their minimum F_1 and maximum F_2 respectively. The coordinates of [a] are the speaker's mean F_1 and F_2 for their TRAP lexical set, with the F_1 value here representing the speaker's maximum observed F_1. [u′] is constructed so that F_1[u′] = F_2[u′] = F_1[i], and represents the minimum F_1 and F_2 possible for a speaker.

The centroid, *S*, of this triangle is then found as the grand mean of points [i], [a] and [u′] using (12). Figure 1 illustrates this.

$$(12)\quad S(F_i) = \frac{F_i[i] + F_i[a] + F_i[u']}{3}$$

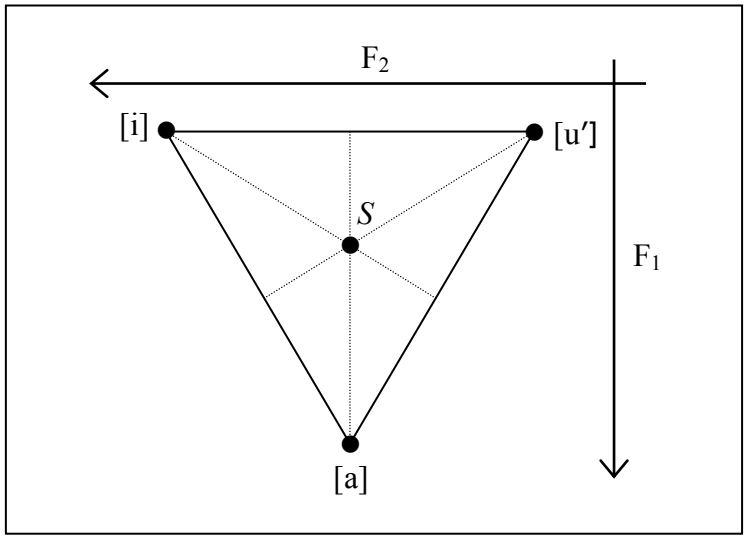


Figure 1: Construction of the centroid S as part of the Watt & Fabricius method of normalisation

Formant values are then expressed relative to the centroid.

(13) $F_i^N = \frac{F_i}{S(F_i)}$

It has been acknowledged that this procedure can skew values in the lower part of the vowel space (Thomas & Kendall 2007; Fabricius et al. 2009; Bigham 2008). As a result, Fabricius et al. (2009) offer a modified formula for calculation of the coordinates of the centroid.

(14) $S(F_i) = \begin{cases} \frac{F_i[i] + F_i[a] + F_i[u']}{3}, & i = 1 \\ \frac{F_i[i] + F_i[u']}{2}, & i = 2 \end{cases}$

Other variations of the Watt & Fabricius method have been used in research projects. For example, Kamata (2008) used mean formant values of the KIT and START lexical sets rather than the FLEECE and TRAP sets to construct the apices of the speakers' vowel triangles, because in the variety she was studying, FLEECE was subject to diphthongisation, and TRAP was suspected to be undergoing a shift (Kamata 2008).

Bigham (2008) used the centroid of a quadrilateral rather than a triangle for his research, as he believed a quadrilateral shape was a better reflection of the vowel space of American English than a triangle (Bigham 2008). The four apices of the quadrilateral used were the mean formant values for a speaker of the American English vowels [ɪ], [ʊ], [æ] and the average of [ɑ] and [ɔ], with tokens taken from word list items of the form /h*V*d/. As per the Watt & Fabricius method, to normalise, a speaker's formant values were expressed relative to their respective centroid, using (13).

A further normalisation procedure that expresses values relative to the hypothetical centre of a speaker's vowel space is that developed by Lobanov (1971). Using a method similar to that

in statistics to transform normally distributed data to a uniform normal distribution, formant values are normalised by subtracting a speaker's mean formant value across all vowel tokens, and then dividing by the standard deviation[1] for the formant across all vowels for that speaker.

$$(15)\quad F_i^N = \frac{(F_i - \mu_i)}{\sigma_i}$$

The final technique to be presented is attributed to Nearey (1978). It has two formulations, one formant-intrinsic and the other formant-extrinsic. In both cases, talkers' vowel spaces are made comparable by aligning them at speakers' mean formant frequencies (Clopper 2009).

In the formant-intrinsic formulation, sometimes referred to as Nearey's Single Log-Mean Method (Adank et al. 2004) or Nearey's Individual Formant Mean Method (Clopper 2009), the natural logarithm of a speaker's formant value is taken, and then the mean of the log-transformed formant frequency across all vowels for the speaker is subtracted.

$$(16)\quad F_i^N = \ln.(F_i) - \mu_{\ln.(F_i)}$$

In the formant-extrinsic formulation, sometimes called Nearey's Grand-Mean Method (Clopper 2009) or Nearey's Shared Log-Mean Method (Adank et al. 2004), the natural logarithm of a speaker's formant value is taken, and then the mean of the log-transformed formant frequency of all formants of all vowels for the speakers is subtracted.

$$(17)\quad F_i^N = \ln.(F_i) - \mu_{\ln.(F_j)} \quad , \quad \forall\, j = 1, \ldots, n$$

The results of studies which have compared the outcomes of normalising via different methods will now be considered.

4. Findings of previous comparative studies

Most researchers who devised their own normalisation methods, compared their normalised results to raw Hertz formant values, and often, values resulting from a transformation of the Hertz values onto a different scale, to evidence the improvement normalising via their formula offered in making formant values from different speakers comparable and giving weight to their argument that adoption of their normalisation formula is warranted.

For example, Watt & Fabricius (2002) showed their method dramatically improved the area ratio and degree of overlap of vowel spaces from two different speakers in comparison to raw Hertz values and Bark-transformed values.

Lobanov (1971) compared normalising using his formula to using Gerstman's method and to using the linear compression and expansion technique. He found that his own method performed the best of the three at reducing the spread of points of the same vowel spoken by different speakers while at the same time maximising the distances between adjacent phonemically-opposing vowels.

Fabricius et al. (2009) compared Watt & Fabricius' method with Nearey's individual log-mean method and Lobanov's method using very rigorous methodology and statistical testing of improvements of Hertz data and of differences between methods in their success at normalising. Their conclusion, was that the Watt & Fabricius method performed at least as

[1] In the original formulation, Lobanov (1971), RMS deviation rather than standard deviation is used.

well as Lobanov's and Nearey's methods, although the results actually show that Lobanov's method outperformed the other two methods overall to a statistically significant extent.

There is always the danger, when researchers compare their own methodology to that of others, that the experiment or results will be somewhat biased towards showing their own procedure in the best light. However, a number of independent comparisons of different normalisation methods have also been conducted.

One of the earliest was Hindle (1978), who compared normalising by Nordström's method, by Nearey's individual log-mean method, and by a six parameter regression method (see Hindle 1978:166 for details). Hindle (1978) concluded that Nearey's method performed the best of the three at normalising data overall, despite it not clustering formant values for the same vowel from different speakers as closely as the six parameter regression method, because Nearey's method was most successful at preserving known age-related formant differences.

Disner (1980) also found Nearey's method to be the best normalisation technique for English vowels out of the four she evaluated, although she used Nearey's grand-mean method.

More recently, Clopper (2009) evaluated both Nearey's individual log-mean and grand-mean methods along with seven other procedures, and found the two to be equally good at producing highly overlapping vowel spaces and aligning the vowel categories of different speakers. Clopper (2009) also found Lobanov's, Watt & Fabricius' and Gerstman's methods to be successful and effective techniques. Rather than singling out one specific procedure as performing the best, she concluded that it is vowel-extrinsic methods in general that are most effective, an opinion that is corroborated by Fabricius et al. (2009) and was earlier posited by Adank (2003) and Adank et al. (2004). Adank et al. (2004) further claimed that specifically, formant-intrinsic rather than extrinsic, vowel-extrinsic methods are the best to use for language variation research. They based this proposal on the findings of their comparison of seven procedures as well as four vowel-intrinsic scaling formulae.

Unlike Clopper (2009), Adank et al. (2004) did indicate a method that performed the best overall, namely, Lobanov's procedure, which they found to be most efficient at preserving phonemic variation, and joint best with Nearey's individual log-mean method at removing variation due to physiological differences (Adank et al. 2004). Langstrof (2006) also came to the conclusion that Lobanov's method was the best technique.

The evidence appears to be, then, that Lobanov's method and one or other of Nearey's methods are consistently found to be the most effective procedures when it comes to normalising vowel formant data.

In coming to their conclusions of which normalisation method is the best to use, studies varied as to how exactly a procedure was evaluated. Some studies looked for overall improvement of vowel space overlap, while some looked for a closer clustering of different speakers' realisations of the same vowels. Also, the purpose of normalising data varied, with the earlier studies such as Disner (1980) and Hindle (1978) seemingly angled more towards normalising to replicate human vowel perception rather than to aid presentation of vowel formant data in sociolinguistic research.

In addition, the range of techniques tested in each study was variable. Hindle (1978) did not test Lobanov's method, while Adank et al. (2004) omitted Watt & Fabricius' method from their otherwise notably thorough investigation. Langstrof (2006) and Fabricius et al. (2009) each only considered three procedures altogether. Clopper (2009) compared an impressive range of methods, but only used data from two speakers. Moreover, she came to her

conclusions without any statistical testing or rigorous mathematical procedure. For these reasons, the robustness of her results could be called into question.

A final point to make, is that very few of the existing comparative studies utilised British English data, the exception being Fabricius et al. (2009). Adank (2003) and Adank et al. (2004) used Dutch data, while Lobanov (1971) used Russian. Indeed, Lobanov's method was originally formulated to aid classification of Russian vowels.

Disner (1980) normalised data from six different Germanic languages by a variety of methods, and found that no one single procedure could, in her opinion, be considered the best at normalising for all the languages tested. For this reason, it could be argued that the normalisation procedure found most effective at normalising Dutch formant values by Adank et al. (2004), need not be the most effective procedure for normalising English data. Furthermore, the most effective procedure for normalising American English data need not be the most effective at normalising British English data (and vice versa).

The remainder of this article presents the methodology and results of a comparative study that used a sizeable dataset of British English data to compare a large variety of differing normalisation techniques, both older techniques that some newer comparative studies did not include due to the assumption that they are outdated, and more recently-proposed methods.

5. Methodology

The data that were normalised came from 20 speakers, 5 young (aged 18-22) and 5 older (aged 40-50) speakers of each sex, all resident in Nottingham. F_1 and F_2 measurements were extracted from monophthongal word list items, taken from the word list tasks of longer one-to-one sociolinguistic interviews collected for Flynn (fc). Mono tracks sampled at 22,050Hz were used, and formant measurements were taken in Praat using a Praat script[2]. A minimum of three adjacent points plotted by Praat's inbuilt formant-tracking tool were averaged for each formant of the vowel. Formant values were measured in Hertz, and then normalised using the algorithm of each procedure under comparison. As not all the procedures under consideration are available as part of NORM, the decision was made to use a specially-prepared Microsoft Excel spreadsheet to perform all the normalisation calculations. 3605 tokens were normalised altogether, giving an average of 180 tokens per speaker. Following normalisation, individual tokens were categorised according to their vowel keyword category, (Wells 1982), and a mean normalised value for each keyword category was calculated for each speaker.

20 procedures of vowel formant normalisation were compared, of which 6 were vowel-intrinsic scaling transformations, and the remaining 14 were vowel-extrinsic. Table 1 summarises the methods used. The equations for all 20 methods can be found in the appendix. Many of the procedures were introduced in section 3. Where modifications were made to established methods, these are described below, as are additional procedures not defined in section 3.

Two logarithmic transformations were computed. One was performed with base 10, the other was the natural logarithm function. Following the method of recent work by Adank (2003), Adank et al. (2004) and Clopper (2009), the equation given by Traunmüller (1990) was used to transform data from Hertz to Barks for all procedures using the Bark scale. A Bark-Difference measure was completed, following the idea of Syrdal & Gopal (1986), but using

[2] The Praat script used was devised by Phil Harrison, originally for use in forensic casework.

Procedure	**Hereafter referred to as**	**Topographical Classification**		
		Formant	**Vowel**	**Speaker**
Log transformation (uses log to the base 10)	*Log*	Intrinsic	Intrinsic	Intrinsic
Ln transformation (uses the natural logarithm)	*Ln*	Intrinsic	Intrinsic	Intrinsic
ERB transformation	*ERB*	Intrinsic	Intrinsic	Intrinsic
Mel scale transformation	*Mel*	Intrinsic	Intrinsic	Intrinsic
Bark scale transformation	*Bark*	Intrinsic	Intrinsic	Intrinsic
Bladon et al. method	*Bladon*	Intrinsic	Intrinsic	Intrinsic
Bark-difference method	*Bark-diff*	Extrinsic	Intrinsic	Intrinsic
Linear Compression/Expansion method	*LCE*	Intrinsic	Extrinsic	Intrinsic
Gerstman method	*Gerstman*	Intrinsic	Extrinsic	Intrinsic
Lobanov method	*Lobanov*	Intrinsic	Extrinsic	Intrinsic
Nordström method	*Nordström*	Extrinsic	Extrinsic	Extrinsic
original Watt & Fabricius method	*origW&F*	Intrinsic	Extrinsic	Intrinsic
Watt & Fabricius method modified as in Fabricius et al. (2009)	*1mW&F*	Intrinsic	Extrinsic	Intrinsic
Watt & Fabricius method modified as described below	*2mW&F*	Intrinsic	Extrinsic	Intrinsic
modified version of Bigham's method	*Bigham*	Intrinsic	Extrinsic	Intrinsic
lettER method	*Letter*	Intrinsic	Extrinsic	Intrinsic
Nearey's individual log-mean method	*Nearey1*	Intrinsic	Extrinsic	Intrinsic
Nearey's grand-mean method	*NeareyGM*	Extrinsic	Extrinsic	Intrinsic
Nearey's individual log-mean method as implemented in NORM	*exp{Nearey1}*	Intrinsic	Extrinsic	Intrinsic
Nearey's grand-mean method as implemented in NORM	*exp{NGM}*	Extrinsic	Extrinsic	Intrinsic

Table 1: The normalisation procedures included in this comparative study

the modification as applied by Thomas & Kendall (2007), namely, that $B_3 - B_1$ is substituted in place of $B_1 - B_0$ in Syrdal & Gopal's (1986) original methodology. (B_i represents Bark-transformed F_i.) The formula for computing the normalised values via this procedure can be expressed as (18).

(18) $F_i^N = B_3 - B_i \quad , \quad i < 3$

Nordström was implemented through the use of (9), with $\mu_{F_3}^{male}$ and $\mu_{F_3}^{female}$ defined as the mean F_3 calculated across all vowel tokens having $F_1 > 600$Hz from all male speakers and from all female speakers respectively. For the dataset used, $\mu_{F_3}^{male}$ was found to be 2494Hz, and $\mu_{F_3}^{female}$ was found to be 2847Hz.
The scale factor for female speakers, $\frac{\mu_{F_3}^{male}}{\mu_{F_3}^{female}}$ was therefore calculated to be 0.876 (to 3dp). [3]

As can be seen in Table 1, the Watt & Fabricius method was implemented in three different formulations. In each case, the general formula given in (19) was used to normalise the formant values.

(19) $F_i^N = \frac{F_i}{S(F_i)}$

S is defined as the centroid of a triangle with the apices of the triangle denoted [i], [a] and [u′] derived from the raw Hertz formant values of a speaker. (See Figure 1 in section 3 for an illustration of this methodology.) The construction of [i], [a] and [u′] differed for each of the three techniques.

OrigW&F used the original technique from Watt & Fabricius (2002). Under this method, $F_i[i] = F_i$[FLEECE], and $F_i[a] = F_i$[TRAP], using the mean points of the respective keywords. [u′] is then constructed so that

(20) $F_1[u'] = F_2[u'] = F_1[i]$

and the formant values of the centroid, *S*, are calculated using (21).

(21) $S(F_i) = \frac{F_i[i] + F_i[a] + F_i[u']}{3}$

1mW&F implemented the minor modification to the calculation of *S* introduced by Fabricius et al. (2009) in response to Thomas & Kendall's (2007) comment that the original Watt & Fabricius formula can distort the lower part of the vowel space. This adjustment places *S* equidistant between [i] and [u′] on the F_2 axis. The formula used to derive the formant values of *S* using this modified method is given in (22). Following Thomas & Kendall (2007), the following additional modifications were made to Watt & Fabricius' original formula in arriving at *1mW&F*. Firstly, rather than using F_1 and F_2 of mean FLEECE, $F_1[i]$ was set equal to the F_1 of whichever mean keyword vowel had lowest F_1, and $F_2[i]$ was set equal to the highest F_2 value of the mean keyword vowels. Similarly, $F_1[a]$ was taken from whichever keyword vowel category had the highest mean F_1. $F_2[a]$ was not computed, as it isn't used in the calculation of $S(F_2)$ under this methodology. [u′] was constructed using (20), as before.

[3] The scale factor used in the normalisation calculations was not rounded.

$$(22)\quad S(F_i) = \begin{cases} \dfrac{F_i[\mathrm{i}] + F_i[\mathrm{a}] + F_i[\mathrm{u}']}{3}, & i = 1 \\ \dfrac{F_i[\mathrm{i}] + F_i[\mathrm{u}']}{2}, & i = 2 \end{cases}$$

For *2mW&F*, [i] and [a] were constructed following identical procedure to that for *1mW&F*. However, [u′], rather than having formant values derived from the point [i], was constructed such that F_2[u′] was set equal to the lowest F_2 value of the mean keyword vowels, and F_1[u′] was set equal to the lowest F_1 value of the mean keyword vowels. The decision to derive [u′] in this way was made because it gives an arguably more realistic placement of [u′] in a speaker's vowel space. Using (20) results in [u′] having an F_2 value far lower than a vowel ever would have in reality. However, based on the results of Watt & Fabricius (2002), Fabricius et al. (2009) and Clopper (2009), deriving [u′] in this way appears to result in well-normalised data. Inclusion of *2mW&F*, with [u′] constructed as described, was intended to see whether this improved, worsened or had little or no effect on normalised data than when [u′] is constructed using (20).

A further derivation of the original Watt & Fabricius method that was included, was a modified version of Bigham's (2008) method. As per the three versions of the Watt & Fabricius method, a speaker's formant values were normalised using (19). However, under this method, *S* was equated from 4 points, that is, *S* is the centroid of a quadrilateral rather than a triangle. This method was included to see the effects on normalised data when *S* is derived from four points rather than three. For ease of notation, the points shall be denoted as [i′], [a′], [o′] and [u′].

The formant values of these constructed points were based on a speaker's mean formant values of the keyword vowel categories. F_2[i′] was set equal to the maximum F_2 value, while F_1[i′] was set equal to the minimum F_1 value, as was F_1[u′]. F_2[u′] was set equal to the minimum F_2 value, as was F_2[o′], while F_1[o′] and F_1[a′] were both set equal to the maximum F_1 value. F_2[a′] was set equal to F_2[TRAP].

The formant values of *S* were then calculated using (23), an equation equivalent to, but much simpler than, those suggested by Bigham (2008).

$$(23)\quad S(F_i) = \frac{F_i[\mathrm{i}'] + F_i[\mathrm{a}'] + F_i[\mathrm{o}'] + F_i[\mathrm{u}']}{4}$$

It could be argued that a quadrilateral shape better reflects the vowel space of British English speakers than a triangle, as speakers will have both front and back maximally-low vowels (for example TRAP and START respectively) rather than one central maximally-low vowel which a triangular representation of the vowel space could be viewed as implying.

A further innovative technique that was included was *Letter*. Like the Watt & Fabricius method and its derivations, this method fixes a speaker's vowel space at its midpoint and then expresses formant values in relation to this point. However, rather than constructing the formant values of the midpoint as the Watt & Fabricius method and derivations do, actual mean formant values of a speaker's *lett*ER vowel are used.

This is justifiable, as *lett*ER is typically realised as an unstressed schwa-like vowel in English (Wells 1982) and schwa has been described as defining the midpoint of a speaker's vowel space (Johnson 2003; IPA 1999).

It should be noted, that *lett*ER is not always realised as schwa. For example, a lowered [ɐ]-like pronunciation has been reported for speakers from Tyneside (Watt & Milroy 1999; Watt & Allen 2003; Beal 2008), London (Tollfree 1999; Trudgill 1986), and Birmingham and the Black Country (Clark 2008), while a lowered and retracted [ɒ]-like pronunciation has been recounted as appearing in the speech of speakers from Sheffield (Beal 2008), Manchester (Beal 2008) and Nottingham (Flynn 2007). Therefore, caution should be taken when normalising via this method, especially when speakers speak a variety where *lett*ER is not always schwa-like. As a result, care was taken to only include tokens perceived auditorily as [ə] in the calculations of the mean F_1 and F_2 of a speaker's *lett*ER vowel as used in this study.

To normalise the data relative to *lett*ER, the formula given in (24) was used.

$$(24)\quad F_i^N = \frac{F_i}{F_i[lett\text{ER}]} - 1$$

The similarity to (19) is apparent, with the added action of subtracting 1 from the resulting ratio of the raw formant value to corresponding *lett*ER formant. This serves to centre the normalised vowel space at (0,0) and to allow the relative positioning of vowels in relation to the vowel space midpoint to be more easily interpreted in acoustic-phonetic terms from the normalised values. Specifically, a negative F_1 indicates a position higher than centre, while a positive F_1 indicates a position in the vowel space lower than centre. Also, a positive F_2 value represents a fronter than central position in the vowel space, while a negative F_2 value is indicative of a backer than centre position.

The final methods that were included were algorithms based on Nearey (1978). It was noticed that when NORM normalises via either of Nearey's (1978) methods, it further takes the exponential[4] of each result calculated through the use of the original formulae. It could be argued that such an action might reverse some of the effects of the normalisation, since the exponential function is the inverse function of the natural logarithm. Certainly, additionally taking the exponential expresses values on a different scale.

The decision was therefore made to calculate the exponential of Nearey-normalised values to create two further sets of normalised data to compare with the other results, and in particular, to those of Nearey's (1978) original formulae, to see if taking the exponential improved, worsened or had no effect on the effectiveness of normalising via one of Nearey's original methods. The equations used can be expressed as (25) for *exp{Nearey1}*, which has *Nearey1* as a basis, and (26) for *exp{NGM}*, which has *NeareyGM* method as a basis.

$$(25)\quad F_i^N = \exp.\left\{ \ln.(F_i) - \mu_{\ln.(F_i)} \right\}$$

$$(26)\quad F_i^N = \exp.\left\{ \ln.(F_i) - \left(\frac{\mu_{\ln.(F_1)} + \mu_{\ln.(F_2)}}{2} \right) \right\}$$

To reveal the overall effectiveness of each procedure, a series of comparisons were performed to test the efficiency of techniques at satisfying the goals of normalisation. In this article, results are reported for the relative ability of each method in equalising and aligning the vowel space areas of the 20 speakers.

Equalisation of different speakers' vowel spaces could be seen as indicating the removal of variation attributable to anatomical differences between speakers, leaving directly

[4] Thomas & Kendall (2007) use the equivalent term 'anti-log' rather than 'exponential'.

comparable vowel spaces of similar dimensions. Taking inspiration from Fabricius et al. (2009), the equalisation of vowel space areas was quantified by examining the reduction of variance in the speakers' vowel space areas as calculated under each normalisation procedure. A speaker's vowel space was taken to be quadrilateral, with the apices constructed from mean keyword category formant values calculated under each method for each speaker, in a similar way to quadrilaterals involved in *Bigham*. That is, four points were defined, one with values (max. F_2, min. F_1), one (min. F_2, min. F_1), one (min. F_2, max. F_1) and one (F_2[TRAP], max. F_1). The lines connecting these four points were taken to represent the hypothetical outer limits of a speaker's vowel space. A graphical depiction of this is shown in Figure 2.

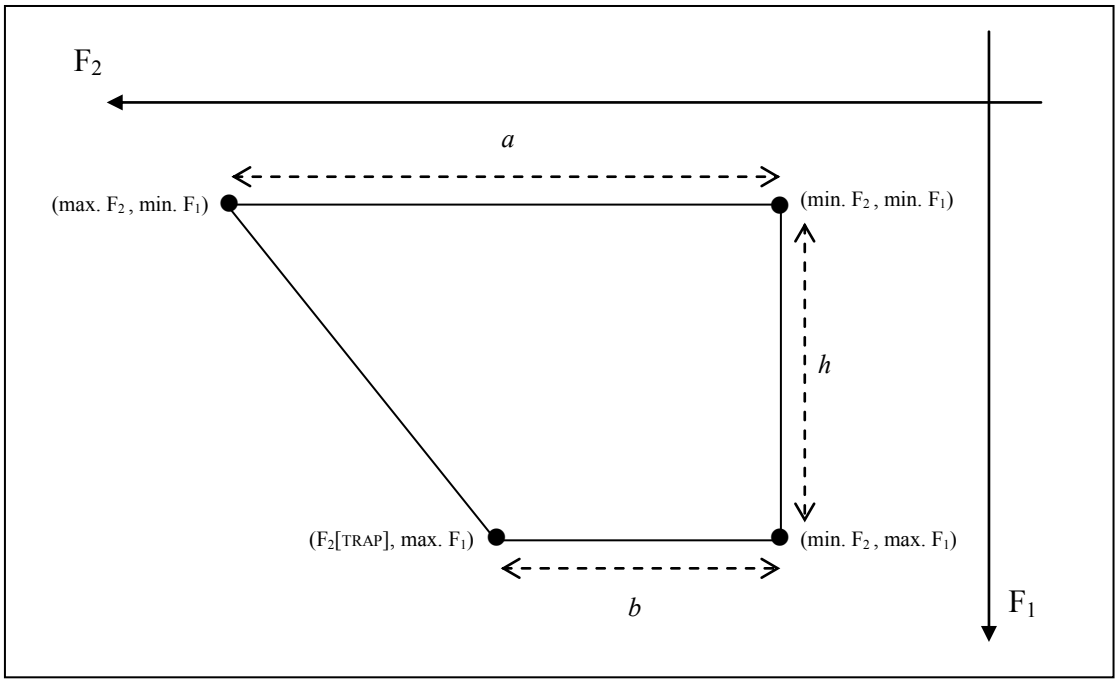


Figure 2: Construction of quadrilateral speaker vowel space areas

The generic formula for calculating the area of a trapezium given in (27) was used to calculate the area of each vowel space.

(27) $A = \frac{1}{2}(a+b)h$

The vertical height, *h*, as shown in Figure 2, was taken to be the F_1 range, while the parallel sides, *a* and *b*, were taken to be the F_2 range at the minimal and maximal extremities of the F_1 range respectively. Equations (28) to (30) present numerically the calculation of the dimensions used in the application of (27).

(28) $a = \text{max. } F_2 - \text{min. } F_2$

(29) $b = F_2(\text{TRAP}) - \text{min. } F_2$

(30) $h = \text{max. } F_1 - \text{min. } F_1$

As the different normalisation methods give normalised values in different units, the resulting vowel space areas also differed in terms of units. It was therefore not possible to immediately

compare directly the amount of variance existing in the sets of vowel space areas under different normalisation procedures to see which method best minimised variance and thus equalised areas to the greatest extent. Following the method of Fabricius et al. (2009), the squared coefficient of variance (SCV), a scale-invariant measure, was calculated for each procedure using (31). The SCVs were then compared.

$$(31)\quad \mathrm{SCV} = \left(\frac{\sigma}{\mu}\right)^2$$

A low SCV is indicative of a dataset having small variance, while a high SCV indicates a dataset has large variance. It is simple then to rank procedures in their effectiveness at equalising areas through examining the respective SCVs. Any normalisation method resulting in areas with a lower SCV than that of the SCV of the raw Hertz areas can be said to have reduced the variance of inter-speaker vowel space areas, and hence made different speakers' vowel space areas more similar. Furthermore, the method resulting in areas with the lowest SCV overall can be said to have reduced area variance the most, and hence equalised the vowel space areas of different speakers to the greatest extent.

In addition to considering the equalisation of vowel space areas, the alignment of speakers' vowel space areas was also taken into account. This is an important consideration to make, as two vowel spaces might have identical areas, but be different shapes, or show poor overlap. In either situation, the vowel spaces of the speakers could not be said to have aligned well, and consequently, it could be argued they still possess variation due to anatomical and physiological differences, and thus are unlikely to be directly comparable.

The alignment and overlap of vowel space areas under the different normalisation procedures were quantified and compared by the following method. Python v2.6.4 incorporating the Shapely v1.2.6 package was used to calculate and compute the areas of the intersection and union of all 20 speaker vowel space areas. Dividing the area of the intersection of all 20 speakers' vowel space areas by the area of the union of all 20 speakers' vowel space areas gave the percentage of area that overlapped. As the overall overlaps of the vowel spaces were calculated as percentages, they can be directly compared. A higher overlap percentage indicates better alignment of the vowel space areas by a normalisation procedure. Scale-drawn vowel space areas were also compared visually to confirm the results. Normalisation methods were again ranked according to how well they overall aligned the vowel space areas of the speakers.

6. Results

6.1. Equalising vowel space areas

Table 2 gives the SCV of speakers' hypothetical total vowel space areas under each normalisation method. The ranking of each method at effectiveness of equalising the vowel space areas based on these SCVs is also given.

As can be seen in Table 2, *Gerstman* displayed the smallest SCV of speaker vowel space areas, so can be said to have shown the least variance in vowel space area. It was thus most effective at equalising the vowel space areas. *LCE* was the next most effective, followed by *Lobanov* and *Bigham*.

1mW&F outperformed the original formulation, which in turn was more effective than *2mW&F* at equalising areas. *Nearey1* and *NeareyGM* performed the same with respect to equalising vowel space areas, giving no reason to favour one over the other based on this result alone. The exponential versions of Nearey's methods also gave identical results to one another, but were not as efficient at equalising the vowel space areas as the original formulations were.

Normalisation Method	**SCV**	**Rank**
Hertz	0.06212	N/A
Gerstman	0.01020	1
LCE	0.01487	2
Lobanov	0.02032	3
Bigham	0.02556	4
1mW&F	0.02587	5
Letter	0.02637	6
origW&F	0.02671	7
2mW&F	0.02818	8
ERB	0.03233	9
Nearey1	0.03250	=10
NeareyGM	0.03250	=10
Log	0.03250	=10
Ln	0.03250	=10
Bladon	0.03409	=14
Bark	0.03409	=14
Bark-diff	0.03549	16
Mel	0.03583	17
exp{Nearey1}	0.03798	=18
exp{NGM}	0.03798	=18
Nordström	0.03977	20

Table 2: SCVs of speakers' vowel space areas for each normalisation method.

All 20 normalisation techniques showed at least some improvement over raw Hertz values at equalising the vowel space areas of different speakers. *Nordström* performed the worst, followed by *exp{Nearey1}* and *exp{NGM}*, and then *Mel* and *Bark*. Although the three worst-performing methods were vowel-extrinsic, the results appear to imply that overall, vowel-intrinsic scaling formulae performed less well than vowel-extrinsic formulae.

6.2. Aligning vowel space areas

Table 3 presents the overlap percentages calculated and corresponding rankings awarded to each procedure based on their ability to align the speakers' vowel spaces. Impressionistically, the percentage overlaps fall into three distinct groups. Lines separating these groups have been drawn onto Table 3 to make the divisions more apparent.

Method	Percent Overlapping	Rank
Bigham	45.8%	1
2mW&F	43.8%	2
origW&F	43.4%	3
1mW&F	42.3%	4
Gerstman	30.0%	5
Lobanov	29.2%	6
Nordstrom	28.7%	7
exp{Nearey1}	27.6%	8
Nearey1	27.1%	9
exp{NGM}	26.9%	10
Bladon	25.9%	11
NeareyGM	25.7%	12
Letter	24.1%	13
LCE	23.1%	14
Bark-diff	13.5%	15
Bark	13.2%	16
Mel	13.1%	17
ERB	12.8%	18
Ln	12.2%	=19
Log	12.2%	=19
Hertz	12.6%	N/A

Table 3: Rankings for each normalisation method's ability to align speaker vowel spaces

Figure 3 displays the raw Hertz vowel space areas of the adult speakers. As can be seen, there is considerable difference in the sizes and shapes, with a clear distinction between the male and female speakers. As seen in Table 3, the percentage overlap was calculated as being just 12.6%.

Results for the vowel-intrinsic scaling methods were largely similar, with little if any overall improvement over raw Hertz in vowel space alignment, seen by overlap percentages increasing by less than 1%. As a visual illustration, Figure 4 gives the vowel space areas as normalised by *Bark*. The similarity to the raw Hertz areas is clearly evident.

The overlap percentages for *Log* and *Ln* were marginally smaller than for Hertz, suggesting that normalisation by these methods actually further misaligns speaker vowel spaces, albeit to a very minimal extent. *Bladon* was the best-performing vowel-intrinsic procedure, appearing mid-table. The total overlap of vowel spaces normalised by *Bladon* was nearly double that of those normalised using just *Bark*. *Letter* finished relatively low in the rankings, with an overlap percentage of 24.1%. This is quite surprising, as, visually, the vowel spaces appeared to align reasonably well, with one notable exception, as seen in Figure 5.

All four formulations of the Nearey method showed noticeable improvement over raw Hertz in aligning speaker vowel space areas, with *Nearey1* performing better than *NeareyGM*, and the exponential version of each method performing better than the original versions in each case, although there were only marginal differences in overlap percentages between the four renderings.

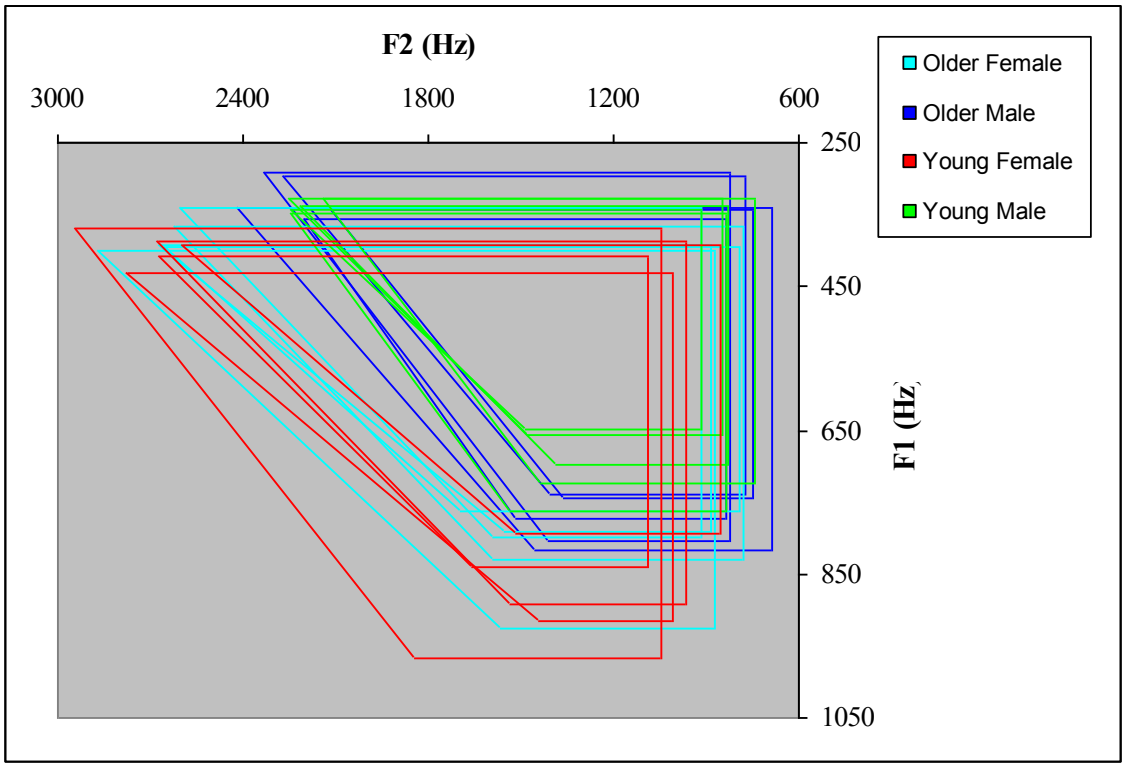


Figure 3: Raw Hertz vowel spaces of the 20 speakers

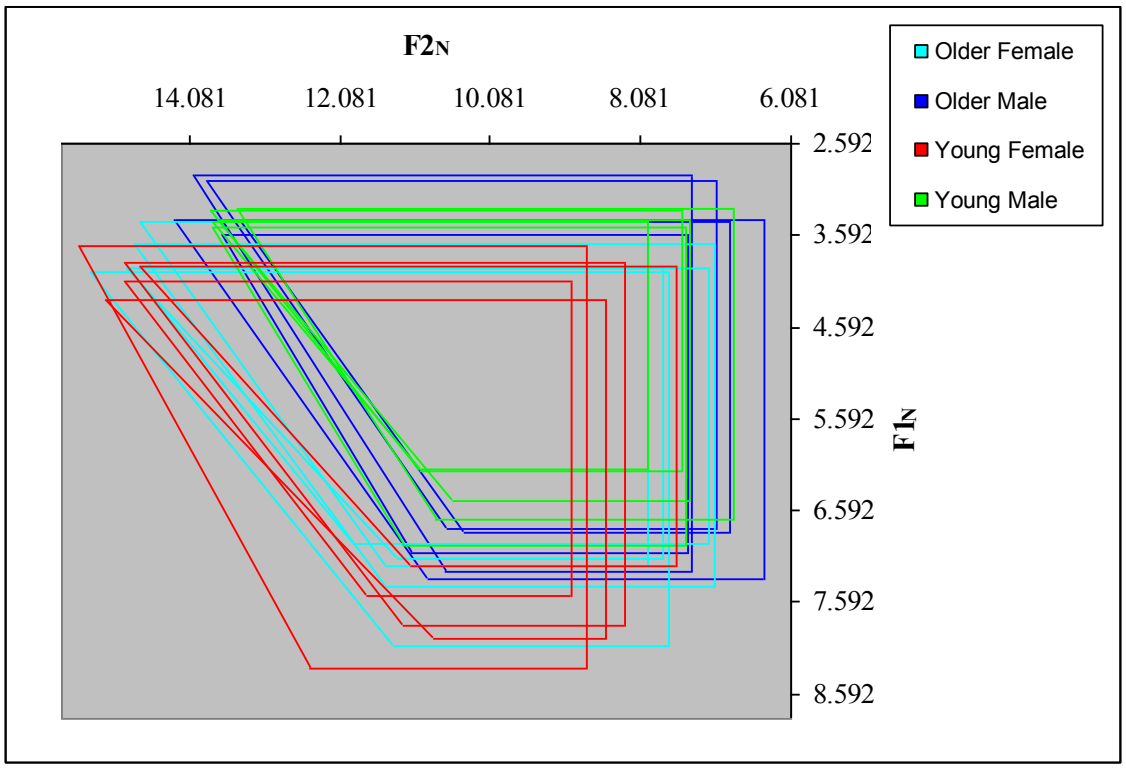


Figure 4: *Bark*-normalised vowel spaces of the 20 speakers

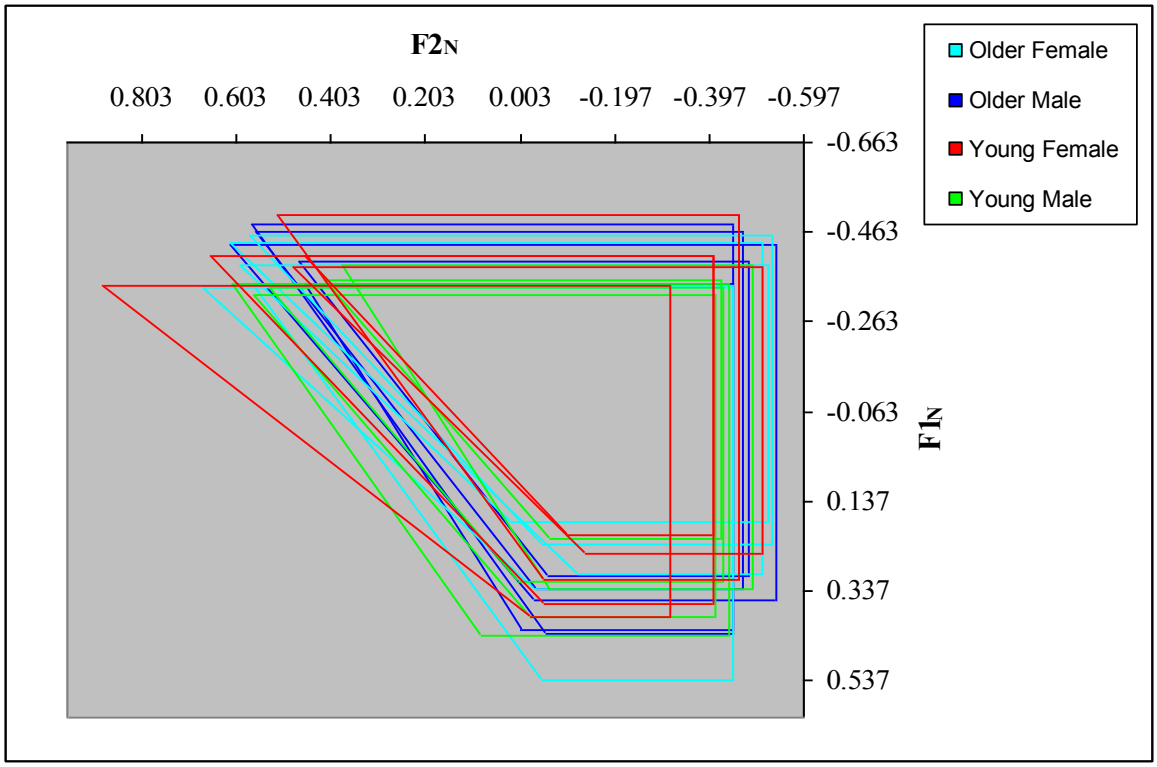


Figure 5: *Letter*-normalised vowel spaces of the 20 speakers

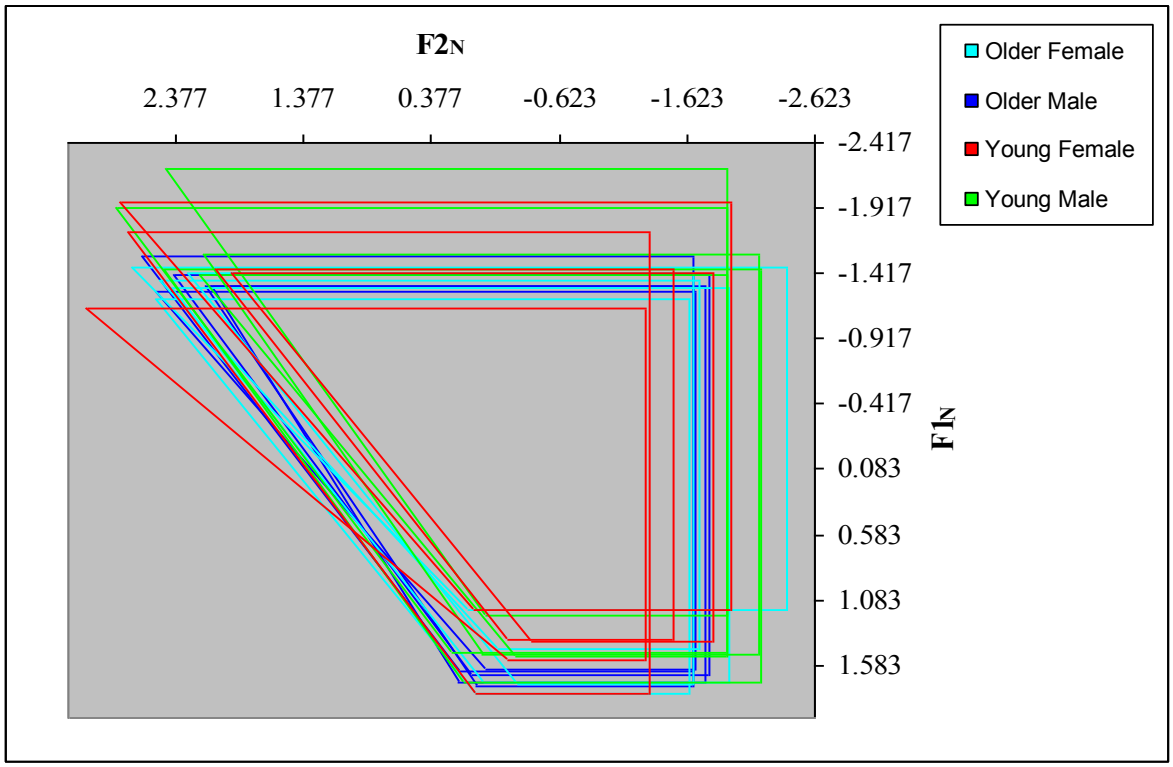
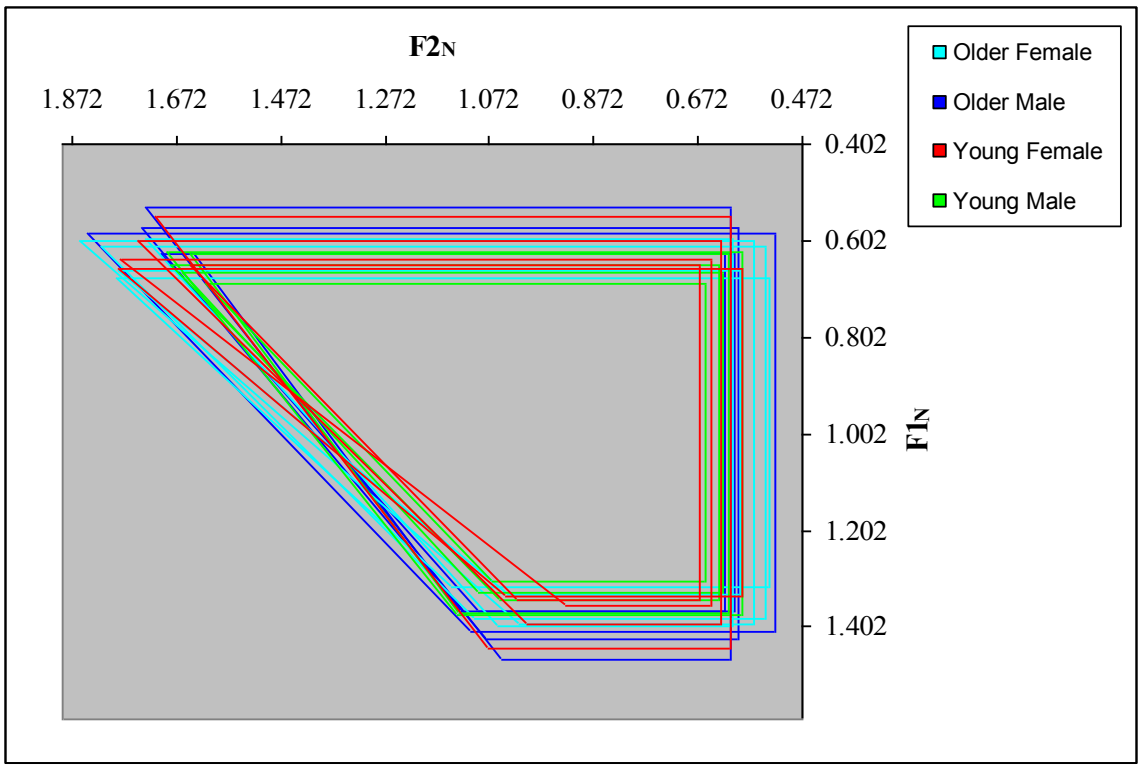


Figure 6: *Lobanov*-normalised vowel spaces of the 20 speakers

The overlap percentage for *Lobanov* was calculated to be 29.2%. This contrasts somewhat with the results of Fabricius et al. (2009: 427), who found the average overlap between a speaker's vowel space and the other speakers' vowel spaces collectively to be 56.4% for RP speakers, and as high as 68.8% for Aberdeen speakers for vowel spaces normalised using *Lobanov*. It is worth noting though, that the initial overlap percentages[5] for Fabricius et al.'s (2009) Hertz data are higher for both data sets than for this study, and the 16.6% overlap improvement by *Lobanov*-normalised data for this study is not that far removed from the 18.4% given by Fabricius et al. (2009: 427) for RP speakers. Inspection of the scale-drawn areas revealed that *Lobanov* failed to align vowel space areas of a number of the young speakers of both sexes, with mismatch at the top edge. However, looking beyond these speakers, the remaining vowel spaces appear to align fairly well as can be seen in Figure 6.

The three versions of *W&F* and *Bigham*, itself derived from *W&F*, were the most effective at aligning the vowel spaces. Indeed, there is a clear separation between these four procedures and the others with respect to overlap percentages, and all four showed vast improvement over raw Hertz measurements visually.

Bigham was ranked as being the procedure that performed the best at aligning the vowel space areas of the speakers, with an overlap percentage of 45.8%, more than triple that of the raw Hertz areas. Figure 7 clearly shows the high degree of alignment and overlap that normalising via *Bigham* facilitated, and comparison with Figure 3 (presented earlier) plainly shows this improvement visually. It is perhaps a little surprising that the overlap percentage isn't actually higher than 45.8% for *Bigham*. Certainly, the graphical results look to show greater overlap of the vowel space areas than this upon visual inspection. Nevertheless, it cannot be denied that increase to 45.8% from a baseline Hertz overlap percentage of 12.6% is a considerable improvement, albeit not to an idealised near-100% value.



[5] Fabricius et al. (2009) actually give their overlap calculations as decimals not percentages, but these have been converted to percentages here for direct comparison

Figure 7: *Bigham*-normalised vowel spaces of the 20 speakers

6.3. Overall Results

To complete the study, the results of the two tests were reviewed and aggregated to discover the overall relative performance of each procedure. A points system was developed with points awarded to procedures based on their rank position in the two comparisons. Points were then totalled and procedures were then ranked in order of points total with a lower score indicating a better result. Table 4 displays the points total and resulting overall ranking of each procedure, as well as summarising the previous results. The maximum possible points total was 40.

Normalisation Method	**Area SCV Reduction Rank**	**Area Aligning Rank**	**Total Points**	**Overall Rank**
Bigham	4	1	5	1
Gerstman	1	5	6	2
1mW&F	5	4	9	=3
Lobanov	3	6	9	=3
2mW&F	8	2	10	=5
origW&F	7	3	10	=5
LCE	2	14	16	7
Letter	6	13	19	=8
Nearey1	=10	9	19	=8
NeareyGM	=10	12	22	10
Bladon	=14	11	25	11
exp{Nearey1}	=18	8	26	12
ERB	9	18	27	=13
Nordström	20	7	27	=13
exp{NGM}	=18	10	28	15
Ln	=10	=19	29	=16
Log	=10	=19	29	=16
Bark	=14	16	30	18
Bark-diff	16	15	31	19
Mel	17	17	34	20

Table 4: Overall rankings for the 20 normalisation methods

7. Discussion

Consideration of the results brings to light some striking patterns and correlations, some in accordance with what other comparative studies of normalisation procedures have found, and some in apparent contrast.

The five vowel-intrinsic scaling transformations all performed poorly in the comparisons conducted, as did *Bark-diff*, itself a vowel-intrinsic method based on manipulations of *Bark*-transformed data. *Mel* was the worst-performing normalisation technique overall, followed by

Bark-diff and *Bark*. This result, coupled with similar findings by Adank (2003), Adank et al. (2004) and Clopper (2009) confirms that vowel-intrinsic methods are less than adequate for the purposes of vowel formant normalisation, for a sociophonetic study at least. The point should be made, however, that even the worst-performing procedure, *Mel*, was an improvement on the raw Hertz frequency measures, suggesting that any form of normalisation is better than not normalising at all. It is possible that vowel-intrinsic normalisation has a role to play in studies of human vowel cognitive perception, but with respect to sociophonetic research, their relative ineffectiveness to more superior methods makes their use redundant.

All formulations of Nearey's method performed relatively poorly in the comparative tests, and, thus, overall. This adds to recent literature that has reported outperformance of Nearey's method by other procedures. For example, Langstrof (2006) ranked *NeareyGM* less favourably than other procedures he tested, while Fabricius et al. (2009) showed *Nearey1* was statistically significantly outperformed by *Lobanov*, *1mW&F* and *origW&F*. In contrast, other researchers have ranked one or other of Nearey's methods as performing well (Clopper 2009; Adank et al. 2004; Adank 2003; Disner 1980; Hindle 1978). Indeed, Disner (1980) concluded that *Nearey1* was the most successful of the methods she tested at normalising effectively, while the same procedure was found by Adank et al. (2004) to be joint best at removing gender-related variation assumed to correspond to anatomical differences. Both of these results are in contrast to the findings of this study.

When considering the four different versions of Nearey's method that were compared, the results of these comparative tests showed that treatment of the formant values was not identical. The conclusion can be drawn, that use of an exponential Nearey formula rather than an original formulation does affect the outcome of the normalised data. Moreover, based on the results of these two tests, *exp{Nearey1}* and *exp{NGM}* both performed worse than their original counterpart. Use of a grand log-mean versus individual log-means was also observed to have an effect on the results of the normalisation. *Nearey1* was found to perform marginally better than *NeareyGM*. This finding corresponds with results of Adank (2003) and Adank et al. (2004).

Letter and *LCE* both performed well at equalising speaker vowel space areas, but comparatively poorly at aligning the vowel space areas, while *Nordström* and *exp{Nearey1}* were both far better vowel space aligners than equalisers. These results demonstrate the possibility of procedures performing to different levels of effectiveness depending on the method of comparison used, and suggest evaluation of procedures should ideally be based on a range of comparative tests.

There is a clear separation between the procedures ranked in the top six overall and the others. Of the six, *origW&F* and *2mW&F* performed least well overall. In correspondence with findings by Fabricius et al. (2009), *1mW&F*, which (following comments by Thomas & Kendall 2007) does not use the F_2 value of the point [a] when constructing the centroid *S* used to normalise data, outperformed *origW&F*. Furthermore, it performed strongly in both comparisons, and so ought definitely to be at least considered when choosing a normalisation procedure. In keeping with Fabricius et al.'s (2009) findings, *1mW&F* performed equally as well as *Lobanov*.

Lobanov, finishing in equal third place, is perhaps not ranked as highly here as in other comparative studies. The inclusion of additional methods not compared by other researchers may be related to this. For example, Adank (2003), who found *Lobanov* to be the best procedure overall, did not include in her study any formulation or derivation of the Watt & Fabricius method, two of which (*1mW&F* and *Bigham*) finished as high as, or higher than,

Lobanov in the overall rankings of this experiment. It should be noted that *Lobanov* still performed well, and considerably better than the majority of the procedures tested. Its existing widespread use in the sociolinguistic world is, therefore, warranted.

A method that performed strongly in both comparative tests but has been largely ignored as a viable procedure by studies thus far, was *Gerstman*. Langstrof (2006), Adank et al. (2004), Adank (2003) and Clopper (2009) all ranked *Gerstman* highly in terms of performance, and remarked on its effectiveness. This combined evidence points to *Gerstman* being an arguably adequate choice for a formant normalisation technique.

Based on the results of equalising and aligning the vowel space areas of different speakers, *Bigham* was evaluated as performing marginally the best overall, closely followed by *Gerstman*. The downside to both these procedures is that neither is available as part of NORM, therefore they are not as easily accessible or useable by researchers as other methods. Of the procedures available in NORM, *1mW&F* and *Lobanov* were the best-performing.

8. Conclusion

This chapter has provided an overview of the available and existing vowel formant normalisation procedures, and compared their effectiveness at normalising formant frequencies from multiple speakers of both sexes and of two age groups.

The research conducted extends the existing normalisation literature as it considers a large dataset consisting of large numbers of tokens from a wide range of vowels from a relatively large speaker sample. The data used were collected for a sociophonetic study, and therefore are neither laboratory-controlled, nor artificially created, and were recorded under conditions typical of sociolinguistic research, meaning the results are not solely applicable to speech data recorded under laboratory conditions.

20 different normalisation procedures were compared to give as full a picture as possible about their relative effectiveness. It is believed that the only established procedures not included are those of Miller (1989) and Labov et al. (2006). Miller's method was omitted as fundamental frequency measurements are needed to perform the formula, and these had not been made.

The decision was made to exclude Labov et al.'s method because the number of speakers used, 24, was deemed insufficient to give robust results. Thomas & Kendall (2007) note that Labov et al.'s procedure is a viable possibility only when the number of speakers is 'exceptionally high'. A figure of 345 speakers is posited (Thomas & Kendall 2007), far exceeding the 20 used for this study.

It is acknowledged that some of the methodology used as part of this study could be criticised. Jacewicz et al. (2007) found that a pentagon better estimated the complete vowel space area used by speakers, while Fox & Jacewicz (2008) hypothesised that a polygon defined by joining all perimeter vowels including diphthongs should be used to avoid underestimating the working space of the vowel system. Bearing these points in mind, the decision to construct and define vowel space areas as quadrilateral for the comparative tests involving equalising and aligning vowel spaces could be questioned. However, it should be noted that the method was used consistently, so the areas defined, though possibly not the exact total vowel space for a speaker, were still directly comparable as they were constructed in the same way for all speakers. Principally, the end results of the comparative tests produced robust results analogous to those of similar existing studies.

The nine best-performing methods all showed the typological classification of vowel-extrinsic, formant-intrinsic, speaker-intrinsic. It could be argued that this is abundant evidence in support of the hypothesis made by Adank (2003) and Adank et al. (2004), and recounted by Clopper (2009) and Fabricius et al. (2009), among others, that it is this type of normalisation method that performs best and is the most suitable to use for language variation research. In comparison, the poor performance by vowel-intrinsic methods, replicating findings by Adank (2003), Adank et al. (2004) and Clopper (2009), imply they are less than adequate for use in sociophonetic research.

Acknowledgements

Thank you to Paul Foulkes, Dom Watt, Bill Haddican, Dan Ezra Johnson, Greg Flynn and two anonymous reviewers for helpful comments and feedback. Thanks are also due to Jess Wardman and James Porter for help with Shapely and Python coding, and Huw Llewelyn-Jones for computer-related assistance. The Praat script used for formant extraction was devised by Phil Harrison of J.P.French Associates. Research was funded by the ESRC.

References

ADANK, PATTI. 2003. *Vowel Normalisation: A Perceptual-Acoustic Study of Dutch Vowels*. PhD Dissertation. Nijmegen: University of Nijmegen.

ADANK, PATTI, SMITS, ROEL & VAN HOUT, ROELAND. 2004. "A comparison of vowel normalisation procedures for language variation research". *Journal of the Acoustical Society of America* 116(5), 3099-3107.

BEAL, JOAN C. 2008. "English dialects in the North of England: Phonology". In: Bernd Kortmann & Clive Upton (eds) *Varieties of English. Vol. 1: The British Isles*. Berlin: Mouton de Gruyter. pp. 122-144.

BIGHAM, DOUGLAS S. 2008. *Dialect Contact and Accommodation among Emerging Adults in a University Setting*. PhD Dissertation. Austin: University of Texas at Austin.

BLADON, R.ANTHONY W., HENTON, CAROLINE G. & PICKERING, J.BRIAN. 1984. "Towards an auditory theory of speaker normalisation". *Language and Communication* 4(1), 59-69.

CLARK, URSZULA. 2008. "The English West Midlands: Phonology". In: Bernd Kortmann & Clive Upton (eds) *Varieties of English. Vol. 1: The British Isles*. Berlin: Mouton de Gruyter. pp. 145-177.

CLOPPER, CYNTHIA G. 2009. "Computational methods for normalising acoustic vowel data for talker differences". *Language and Linguistic Compass* 3(6), 1430-1442.

DISNER, SANDRA F. 1980. "Evaluation of vowel normalisation procedures". *Journal of the Acoustical Society of America* 67(1), 253-261.

FABRICIUS, ANNE H. 2008. *Vowel Normalisation in Sociophonetics: When, Why, How?* Paper presented at Sociolinguistics Circle, Copenhagen University, 16th September 2008.

FABRICIUS, ANNE H., WATT, DOMINIC J.L. & JOHNSON, DANIEL E. 2009. "A comparison of three speaker-intrinsic vowel formant frequency normalisation algorithms for sociophonetics". *Language Variation and Change* 21(3), 413-435.

FLYNN, NICHOLAS E.J. 2007. *A Sociophonetic Comparison of Adolescent Speakers in two Areas of Nottingham.* MA Dissertation. Colchester: University of Essex.

FLYNN, NICHOLAS E.J. fc. *Levelling and Diffusion at the North/South Border: A Sociophonetic Study of Nottingham Speakers* [working title]. PhD Dissertation. York: University of York.

FOX, ROBERT A. & JACEWICZ, EWA. 2008. "Analysis of total vowel space areas in three regional dialects of American English". In: *Proceedings of Acoustics '08 Paris*. pp. 495-500.

GERSTMAN, LOUIS. 1968. "Classification of self-normalised vowels". *IEEE Transactions of Audio Electroacoustic*s AU-16. pp.78-80.

GLASBERG, BRIAN R. & MOORE, BRIAN C.J. 1990. "Derivation of auditory filter shapes from notched noise data". *Hearing Research* 47(1-2), 103-138.

HINDLE, DONALD. 1978. 'Approaches to vowel normalisation in the study of natural speech". In: David Sankoff (ed) *Linguistic Variation: Models and Methods*. New York: Academic Press. pp. 161-171.

INTERNATIONAL PHONETIC ASSOCIATION. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

JACEWICZ, EWA, FOX, ROBERT A. & SALMONS, JOSEPH. 2007. "Vowel space areas across dialects and gender". In: J. Trouvain & W. Barry (eds) *Proceedings of the 16th International Congress of Phonetic Sciences*, 6th – 10th August 2007, Saarbrücken. Saarbrücken: Universität des Saarlandes. pp. 1465-1468.

JOHNSON, KEITH. 2003. *Acoustic and Auditory Phonetics*. (2nd edn) Oxford: Blackwell.

KAMATA, MIHO. 2008. *An Acoustic Sociophonetic Study of three London Vowels*. PhD Dissertation. Leeds: University of Leeds.

LABOV, WILLIAM, ASH, SHARON & BOBERG, CHARLES. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.

LANGSTROF, CHRISTIAN. 2006. *Vowel Change in New Zealand English – Patterns and Implications*. PhD Dissertation. Christchurch, New Zealand: University of Canterbury.

LOBANOV, B.M. 1971. "Classification of Russian vowels spoken by different speakers". *Journal of the Acoustical Society of America* 49(2), 606-608.

MILLER, JAMES D. 1989. "Auditory-perceptual interpretation of the vowel". *Journal of the Acoustical Society of America* 85(5), 2114-2134.

NEAREY, TERRANCE M. 1978. *Phonetic Feature Systems for Vowels.* PhD Dissertation reproduced by the Indiana University Linguistics Club. Indiana: Indiana University Linguistics Club.

NORDSTRÖM, PER-ERIK. 1977. "Female and infant vocal tracts simulated from male area functions". *Journal of Phonetics* 5(1), 81-92.

ROSNER, BURTON S. & PICKERING, J.BRIAN. 1994. *Vowel Production and Perception*. Oxford: Oxford University Press.

STEVENS, STANLEY S. & VOLKMANN, JOHN. 1940. "The relation of pitch to frequency: A revised scale". *American Journal of Psychology* 53(3), 329-353.

SYRDAL, ANN K. & GOPAL, H.S. 1986. "A perceptual model of vowel recognition based on the auditory representation of American English vowels". *Journal of the Acoustical Society of America* 79(4), 1086-1100.

THOMAS, ERIK R. 2002. "Instrumental phonetics". In: J.K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds) *The Handbook of Language Variation and Change*. Oxford: Blackwell. pp. 168-200.

THOMAS, ERIK R. & KENDALL, TYLER. 2007. *NORM: The Vowel Normalisation and Plotting Suite*. Online Resource. URL: <http://ncslaap.lib.ncsu.edu/tools/norm> Accessed: 17/11/2008.

TOLLFREE, LAURA F. 1999. "South East London English: Discrete versus continuous modelling of consonantal reduction". In: Paul Foulkes & Gerard J. Docherty (eds) *Urban Voices: Accent Studies in the British Isles*. London: Arnold. pp. 163-184.

TRAUNMÜLLER, HARTMUT. 1990. "Analytical expressions for the tonotopic sensory scale". *Journal of the Acoustical Society of America* 88(1), 97-100.
TRUDGILL, PETER. 1986. *Dialects in Contact*. Oxford: Blackwell.
WATT, DOMINIC J.L. & ALLEN, WILLIAM H.A. 2003. "Illustrations of the IPA: Tyneside English'. *Journal of the International Phonetic Association* 33(2), 267-271.
WATT, DOMINIC J.L. & FABRICIUS, ANNE H. 2002. "Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F_1~F_2 plane". *Leeds Working Papers in Linguistics and Phonetics* 9, 159-173.
WATT, DOMINIC J.L., FABRICIUS, ANNE H. & KENDALL, TYLER. 2010. "More on vowels: Plotting and normalising". In: Marianna di Paolo & Malcah Yaeger-Dror (eds) *Sociophonetics: A Student's Guide*. London: Routledge. pp. 107-118.
WATT, DOMINIC J.L. & MILROY, LESLEY. 1999. "Patterns of variation and change in three Newcastle vowels: Is this dialect levelling?". In: Paul Foulkes & Gerard J. Docherty (eds) *Urban Voices: Accent Studies in the British Isles*. London: Arnold. pp. 25-46.
WELLS, J.C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Appendix

Normalisation Method	**Equation(s) Used**
Log	$F_i^N = \log_{10}(F_i)$
Ln	$F_i^N = \ln.(F_i)$
ERB	$F_i^N = 21.4 \ln.(0.00437F_i + 1)$
Mel	$F_i^N = 1127 \ln.\left(1 + \frac{F_i}{700}\right)$
Bark	$F_i^N = 26.81 \left(\frac{F_i}{1960 + F_i}\right) - 0.53$
Bladon	$F_i^N = \begin{cases} 26.81\left(\frac{F_i}{1960 + F_i}\right) - 0.53 & \text{, speaker is male} \\ \left(26.81\left(\frac{F_i}{1960 + F_i}\right) - 0.53\right) - 1 & \text{, speaker is female} \end{cases}$
Bark-diff	$F_i^N = B_3 - B_i \quad , \quad i < 3 \quad (B_i = \text{Bark-transformed } F_i)$
LCE	$F_i^N = \frac{F_i}{F_i^{\max}}$

Gerstman	$F_i^N = 999\left(\frac{F_i - F_i^{min}}{F_i^{max} - F_i^{min}}\right)$
Lobanov	$F_i^N = \frac{(F_i - \mu_i)}{\sigma_i}$
Nordström	$F_i^N = \begin{cases} F_i & , \text{ speaker is male} \\ \left(\frac{\mu_{F_3}^{male}}{\mu_{F_3}^{female}}\right) F_i & , \text{ speaker is female} \end{cases}$ where $\frac{\mu_{F_3}^{male}}{\mu_{F_3}^{female}} = 0.876\ldots$ for this dataset.
origW&F	$F_i^N = \frac{F_i}{S(F_i)}$ $\quad S(F_i) = \frac{F_i[i] + F_i[a] + F_i[u']}{3}$ $F_i[i] = F_i[\text{FLEECE}]$ $F_i[a] = F_i[\text{TRAP}]$ $F_1[u'] = F_2[u'] = F_1[i]$
1mW&F	$F_i^N = \frac{F_i}{S(F_i)}$ $\quad S(F_i) = \begin{cases} \frac{F_i[i] + F_i[a] + F_i[u']}{3} & , i = 1 \\ \frac{F_i[i] + F_i[u']}{2} & , i = 2 \end{cases}$ $F_1[i] = F_1^{min}$, $F_2[i] = F_2^{max}$, $F_1[a] = F_1^{max}$, $F_1[u'] = F_2[u'] = F_1[i]$
2mW&F	$F_i^N = \frac{F_i}{S(F_i)}$ $\quad S(F_i) = \begin{cases} \frac{F_i[i] + F_i[a] + F_i[u']}{3} & , i = 1 \\ \frac{F_i[i] + F_i[u']}{2} & , i = 2 \end{cases}$ $F_1[i] = F_1^{min}$, $F_2[i] = F_2^{max}$, $F_1[a] = F_1^{max}$, $F_1[u'] = F_1^{min}$, $F_2[u'] = F_2^{min}$
Bigham	$F_i^N = \frac{F_i}{S(F_i)}$ $\quad S(F_i) = \frac{F_i[i'] + F_i[a'] + F_i[o'] + F_i[u']}{4}$ $F_1[i'] = F_1^{min}$, $F_2[i'] = F_2^{max}$, $F_1[a'] = F_1^{max}$, $F_2[a'] = F_2[\text{TRAP}]$, $F_1[o'] = F_1^{max}$, $F_2[o'] = F_2^{min}$, $F_1[u'] = F_1^{min}$, $F_2[u'] = F_2^{min}$

Letter	$F_i^N = \frac{F_i}{F_i[lett_{ER}]} - 1$

Nearey1	$F_i^N = \ln.(F_i) - \mu_{\ln.(F_i)}$
NeareyGM	$F_i^N = \ln.(F_i) - \mu_{\ln.(F_j)} \quad , \forall j = 1, \ldots, n$
exp{Nearey1)	$F_i^N = \exp.\{\ln.(F_i) - \mu_{\ln.(F_i)}\}$
exp{NGM}	$F_i^N = \exp.\left\{\ln.(F_i) - \left(\frac{\mu_{\ln.(F_1)} + \mu_{\ln.(F_2)}}{2}\right)\right\} \quad , \forall j = 1, \ldots, n$

Nicholas Flynn
Department of Language and Linguistic Science
University of York
Heslington
York
YO10 5DD
email: nejf100@york.ac.uk