Elena A. Buraya
Moscow State Linguistic University
Moscow, Russian Federation
helen812@list.ru

Prosodic rhythm in Australian English (Gender differentiation)

Abstract

This paper develops the concept of rhythmic classes through the approaches based on metrics designed to quantify speech rhythm. The emergence of a number of PC software packages, that considerably facilitate the analysis of a speech signal, inspired linguists to renew their interest in speech rhythm which waned when Rhythm Class Hypothesis did not prove to be supported with experimental data. New approaches to speech rhythm study are based on the so-called rhythmic metrics – the Deltas, the PVIs and those based on voice timing which are viewed as acoustic rhythmic correlates. Thus, the idea of the paper is to test within the scope of one study the validity of all the three methods using the material of Australian English. The paper also aims at analyzing rhythmical patterns of Australian English, comparing the obtained data with the results on other varieties of English and finding out whether the gender factor has any influence on Australian English speech rhythm. The evaluation of the research data testifies to the ability of each of the three methods to be quite reliable in quantifying speech-rhythm differences. The profiles of the metrics provide acoustic evidence for gender differentiation in Australian English speech rhythm with female speech being more stress-timed.

Keywords: rhythm, gender, Australian English, rhythm acoustic correlates, metrics, PVI, Deltas, voice timing

Бурая Елена Анисимовна
Московский государственный лингвистический университет
г. Москва, Российская Федерация
helen812@list.ru

Просодический ритм в австралийском варианте английского языка (гендерный аспект)

Аннотация

В работе развивается идея разделения языков на ритмические классы, в основе которой лежат подходы, основанные на методиках, предназначенных для измерения речевого ритма. Развитие компьютерных технологий позволило создать программы, которые открывают новые возможности в исследовании речи. Это возродило интерес лингвистов к изучению речевого ритма, угасшего, когда стало ясно, что гипотеза Пайка-Аберкромби о типологическом различии ритмических классов не была подкреплена экспериментальными данными. Новые подходы к изучению речевого ритма базируются на так называемых ритмических метриках – дельтах, PVI и метриках, вычисленных на основе периодов звонкости, которые рассматриваются как акустические корреляты речевого ритма. В связи с этим в работе на материале австралийского варианта английского языка предпринимается попытка проверить в рамках одного исследования надёжность каждой из трёх методик. В исследовании также проводится сравнение полученных результатов с данными по другим вариантам английского языка. Анализ и сопоставление эксперимен-

тальных данных позволяют говорить о надёжности каждой из трёх методик для измерения речевого ритма. Значения ритмических метрик также свидетельствуют о гендерной дифференциации ритма австралийского варианта английского языка: экспериментальные данные говорят о более выраженном тактосчитающем характере женской речи.

Ключевые слова: ритм, гендер, австралийский вариант английского языка, метрики, PVI, дельты, периоды звонкости

1. Introduction

The current research sets out to investigate Australian English speech rhythm. Australian English (AusE) is a major variety of the English language. It differs from other English speaking varieties in vocabulary, grammar, spelling and of course, pronunciation. While phoneticians have already described AusE segmental phonology, its suprasegmental level, including rhythm, has so far been poorly investigated.

With evolving computer technology there appeared new software designed for studying speech and thus new methods of analyzing speech rhythm, namely, the Deltas, the Pairwise Variability Index and the method based on voice timing, have been developed. The paper is an attempt to apply the mentioned approaches to AusE rhythm and assess their validity. The research also focuses upon gender differences in AusE speech rhythm.

2. Typological approach in the study of rhythm

A classical view on speech rhythm maintains that languages fall into two distinct rhythm classes, namely stress-timed and syllable-timed. The terms were evidently coined by Pike [Pike, 1945] to describe the rhythm of Romance languages on the one hand, and Germanic languages on the other hand. The -timed component here points to an isochronous repetition of an element. In syllable-timed languages such an element is a syllable, so successive syllables, whether stressed or unstressed, are supposed to be of equal duration. In stress-timed languages such an element is an interstress interval, so time periods between stressed syllables are said to be equal.

Pike's ideas were later developed by his British follower Abercrombie who claimed that the two types of timing are strictly categorical, that is, there are two and only two types of rhythmic classes. And "…every language in the world is spoken either with one kind of rhythm or with the other" [Abercrombie, 1967, p. 97].

The ideas put forth by Pike and Abercrombie marked the beginning of a typological study of speech rhythm.

The theory of the two types of speech rhythm was suggested by the authors on the basis of their authority impressions, but later experiments failed to provide any acoustic evidence for the isochrony of the mentioned intervals [Roach, 1982].

However, the idea has already become appealing to linguists and appeared to be backed by perceptual data. This has obliged some researchers to retreat from 'objective isochrony' in favour of 'subjective isochrony'. Isochrony became to be accepted as a concept that relates to speech perception. And it is possible to say that "… every language in the world is perceived with one kind of rhythm or with the other…" [Cauldwell, 2002, p. 18].

Some linguists [Dauer, 1983, among others] attributed the impression of stress- or syllable-timing to phonotactic properties of a language: the syllabic structure and the degree of vowel reduction. It is common knowledge that stress-timed languages allow for a greater complexity in syllable onsets and codas. Syllable-timed languages have a very limited number of syllable types with consonant clusters being rare. Vowel reduction is a property of stress-timed languages. In syllable-timed languages there is practically no vowel reduction.

Dauer also advocates a continuous model of rhythm. This suggests that there exists a variety of rhythmic types that could be viewed as a continuum on which languages may be placed depending on how syllable- or stress-timed their rhythm is. Thus, it is not the duration of certain units as Pike and Abercrombie supposed, but rather the syllable complexity and vowel reduction that are especially relevant to rhythmic differences.

3. Rhythmic acoustic correlates

At the turn of the XXI century the distinction between two rhythmic classes received renewed interest. Up-to-date approaches to speech rhythm are based on the so-called rhythmic metrics which, as linguists believe, are able to measure speech rhythm and compare it in different languages. The metrics are derived from the segmentation of speech into vocalic and consonantal intervals.

3.1. The Deltas

One of the approaches was initiated by Ramus and his co-workers [Ramus et al., 1999]. They measured the duration of consonantal and vocalic segments and calculated three parameters (or metrics): the standard deviation of consonantal intervals (marked ΔC), the standard deviation of vocalic intervals (marked ΔV) and the proportion of vocalic intervals in a sentence (marked %V). These parameters were later called 'the Deltas'.

The standard deviation of consonantal intervals (ΔC) is expected to point to syllable complexity: the more complex the syllabic structure the higher the value of the metric. So stress-timed languages are expected to have higher values of this parameter.

The standard deviation of vocalic intervals (ΔV) is expected to point to the degree of vowel reduction in unstressed syllables, so stress-timed languages are supposed to have a higher value of this parameter too.

The vocalic percentage (%V) points to both the presence/absence of vowel reduction and the complexity of the syllabic structure. This parameter is thought to be higher in syllable-timed languages.

The researches computed these metrics for some languages and in order to better visualize the results put the data on three planes: ΔV-ΔC, %V-ΔC and %V-ΔV.

The correlation %V-ΔC in figure 1[1] appeared to be most illustrative for language discrimination.

Figure 1 shows that the languages are not scattered randomly in the plane, they form groups that resemble rhythmic classes. British English and Dutch (stress-timed) cluster together in the left upper part of the plane. French, Italian and Spanish (syllable-timed) cluster together too, but away from stress-timed languages. Though the languages form two groups, the values of their metrics within each group are slightly different which means that there is a weak rhythmical distinction between them and this distinction is gradual.

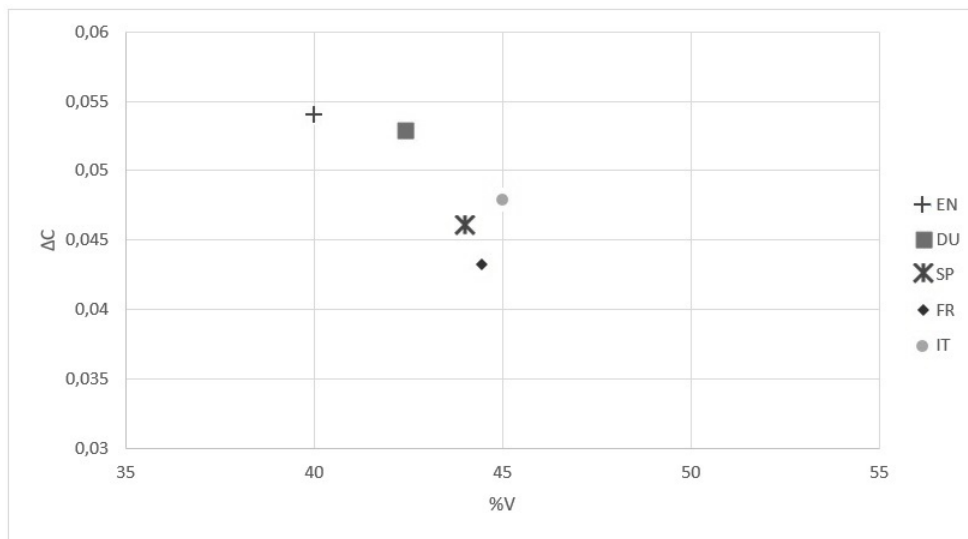[1] Figure 1 is a modified version of Figure 1 from Ramus et al. (1999).

F i g u r e 1. **Distribution of languages over the %V-ΔC plane**

3.2. Pairwise Variability Index

Another research designed to quantify speech rhythm was carried out by Grabe and Law [Grabe, Law, 2002]. They evolved an approach to speech rhythm analysis based on a different metric which they called "The Pairwise variability index – PVI". The PVI is also based on the segmentation of speech into vocalic and consonantal intervals.

To calculate the index they worked out the following equation:

$$rPVI = \sum_{k=1}^{m-1} |d_k - d_{k-1}| / (m-1)$$

where m is the number of items in an utterance and d is the duration of the kth item.

To eliminate the effect of speaking rate the linguists used a normalised version of the Index which they applied to vocalic intervals. The equation for this version, nPVI, is

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k-1}}{(d_k - d_{k+1})/2} \right| / (m-1) \right]$$

The researchers suppose that there is no need to normalize the index for speech rate while calculating it for consonantal intervals (rPVI), as their duration mostly depends on the character of articulatory gestures, necessary for consonant production.

PVI shows the level of variability in neighbouring syllables. In the so-called stress-timed languages where unstressed vowels are heavily reduced and the syllable structure is varied, the value of the index is supposed to be high. And in the so-called syllable-timed languages, which do not allow for vowel reduction, and where the syllable structure is rather limited, it is expected to be low. Then the researchers computed the values of the indexes for several languages and put them on a plane (fig. 2[2]).

In figure 2, three stress-timed languages – British English, Dutch and German – are grouped close to each other in the upper part of the plane, whereas two syllable-timed languages – French and Spanish – are located in its lower part. The position of the data projections in the planes (fig. 1 and fig. 2) clearly illustrate that rhythmic distinction between languages is not strictly categorical, as Pike and Abercrombie assumed, but rather gradual as Dauer supposed.

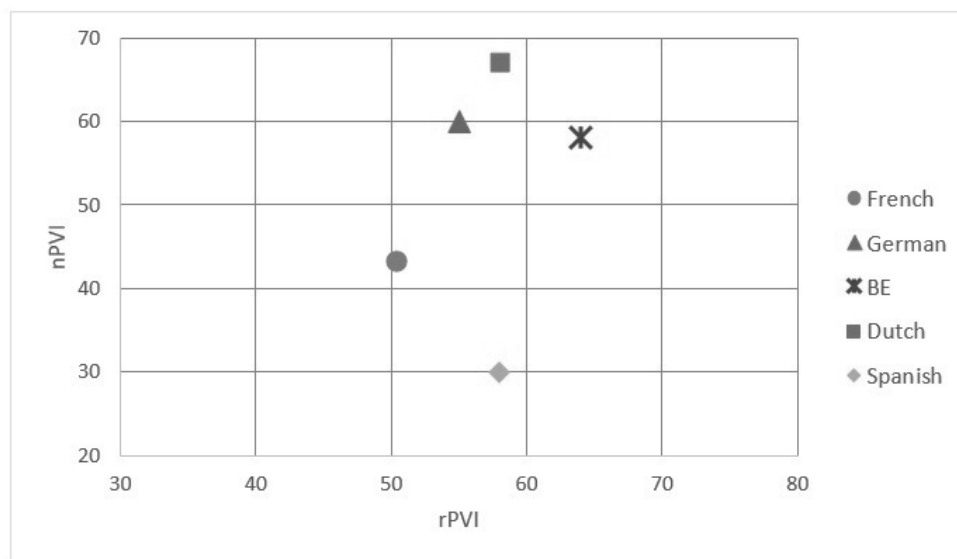[2] Figure 2 is a modified version of Figure 1 from Grabe, Low (2002).

F i g u r e 2. **Distribution of languages over the rPVI-nPVI plane**

The new approaches to speech rhythm mean a conceptual shift in its analysis. Pike and Abercrombie believed that the type of rhythm is a categorical concept with the members of the same category being equal. The new approaches assume that there exists a diversity of rhythms. The distinction between them has a gradient character, i.e. one language may be more or less stress-timed than another.

These rhythmic metrics (the Deltas and the PVIs [Grabe, Low, 2002]) began to be treated by linguists as acoustic correlates of speech rhythm, as units that measure rhythm. The metrics have provided rhythmologists with a precise method for calculating, analyzing and comparing rhythmic differences.

3.3. Indexes based on voice timing

The third line of research was initiated by Fourcin and Dellwo [Fourcin, Dellwo, 2009]. Their work was inspired by experiments with French infants' behaviour [Nazzi et al., 1998]. These experiments have proved that infants seemed to distinguish speech in French, Italian and Spanish from phrases spoken in Dutch and English. Fourcin and Dellwo supposed that the ability of newborns to differentiate between rhythmic classes might mean that there are firm signals in speech that allow drawing such kind of distinction. The researchers assumed that the cues for language differentiation might be influenced by whether they are physically voiced or voiceless. For this reason, they concluded that infants may prefer voiced-voiceless variability cues to distinguish between rhythmic classes.

Using this prediction as a benchmark, the linguists changed the unit of segmentation. They used the durations of voiced (VO-) and unvoiced (UV-) periods in the speech signal.

Fourcin and Dellwo applied the metrics developed by Ramus et al. and those developed by Grabe and Low to voiced and unvoiced intervals. Thus, the newly derived metrics are called:

• %VO: the percentage of total time over which speech is voiced
• ΔUV: the standard deviation of voiceless intervals.
• nPVI-VO: the rate normalised difference between consecutive voiced intervals.
• rPVI-UV: the non rate normalised difference between consecutive unvoiced intervals.

The main experimental advantage of this method is that the segmentation does not depend on any linguistic knowledge about the language, and it is much less time-consuming

than the methods based on the segmentation into vocalic and consonantal units as it relies only on the F0 measurements.

4. Gender differences in speech rhythm

Gender speech differences have lately been the subject of many researches. However, very few works tackle the issue of gender differences in speech rhythm. Nokes and Hay [Nokes, Hay, 2003] found proof of speech rhythm gender differentiation in the New Zealand variant of the English language. However, Szakay [Szakay, 2006], who also considered this problem, failed to find significant gender distinction. Callier [Callier, 2011] investigated the problem in Chinese. The study appeared to be quite positive about rhythmic gender differences.

5. Material and tools

The research material includes 20 realizations of the same text read by ten Australian men and ten Australian women. The recordings were taken from accent.gmu.edu, the speech accent archive. Each sample lasts 50 seconds. The sounding material was downloaded to the computer. To calculate the metric values, the segmentation of the text was done in Praat (Version 6.1.01) [Boersma, Weenink, 2019].

The files annotated in PRAAT were further processed using Correlatore (2010) to calculate the rhythmic metrics. This program does it automatically on the basis of TextGrid file.

6. Australian English speech rhythm. Results and evaluation

6.1. Delta values

As mentioned earlier, the ratio %V-ΔC (fig. 1) demonstrably shows the difference between languages in speech rhythm. The larger the ΔC value and the smaller the %V value the more stress-timed the language is. First, the %V and ΔC values for each of the 20 AusE speakers were calculated, and then the data were averaged. The average %V value equals 42.2 points and the average ΔC value amounts to 5.4 points. Figure 3 presents the results.
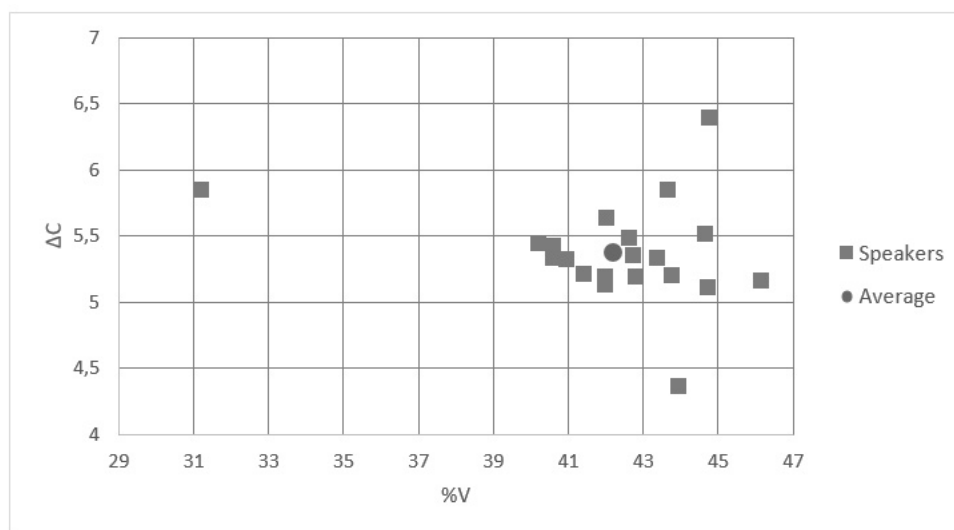

F i g u r e 3. **%V-ΔC values for 20 AusE speakers**

Then the average %V-ΔC values for AusE were put on another plane to compare them with those for other languages (See fig. 4). It is evident that AusE clusters with other stress-timed languages, British English and Dutch, being more syllable-timed than BE, though.
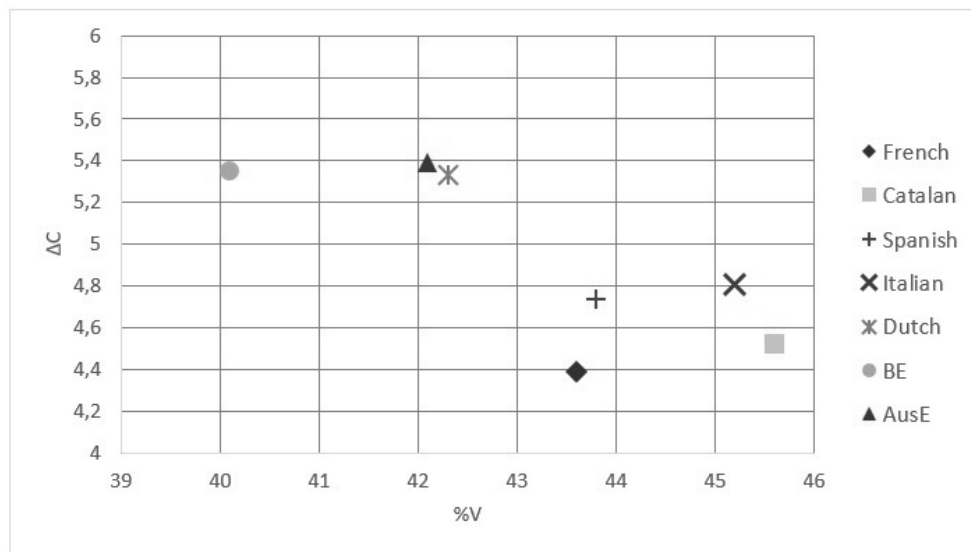
F i g u r e 4. **Distribution of AusE and other languages over the %V-ΔC plane**

Since the research also concentrates on gender differences, the average ΔC and %V values were computed for male and female speakers separately. The average ΔC value for male speech amounts to 5.3 points vs 5.4 points for female speech. The average %V value for male speech equals 42.5 points vs 41.8 points for female speech. Figure 5 shows the location of the metrics on the plane together with the average data.
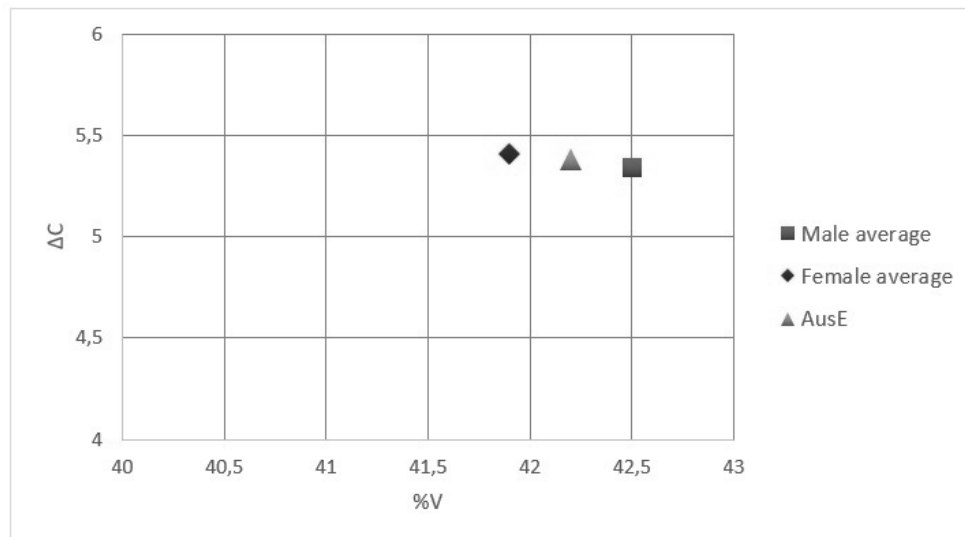

F i g u r e 5. **AusE gender differences (%V- ΔC ratio)**

The position of the values in the graph points to gender differences in AusE speech rhythm. The figure also shows that AusE female speech appears to be more stress-timed than male speech, as it displays slightly lower %V and higher ΔC values compared to male speech.

6.2. PVI values

Sections 3.1 and 3.2. made it clear that rhythmically languages form a continuum. Even typologically the same languages display a slight difference in their metric values.

The nPVI value (which is most informative than rPVI) appears to be larger for stress-timed languages. ꜰor the purpose of this study, the nPVI and rPVI data for each of the 20 AusE speakers were calculated and then the average values were obtained. Figure 6 demonstrates the results. It turned out that, the average nPVI value for AusE speech amounts to 57.9 points and the average rPVI value equals 85.8 points.
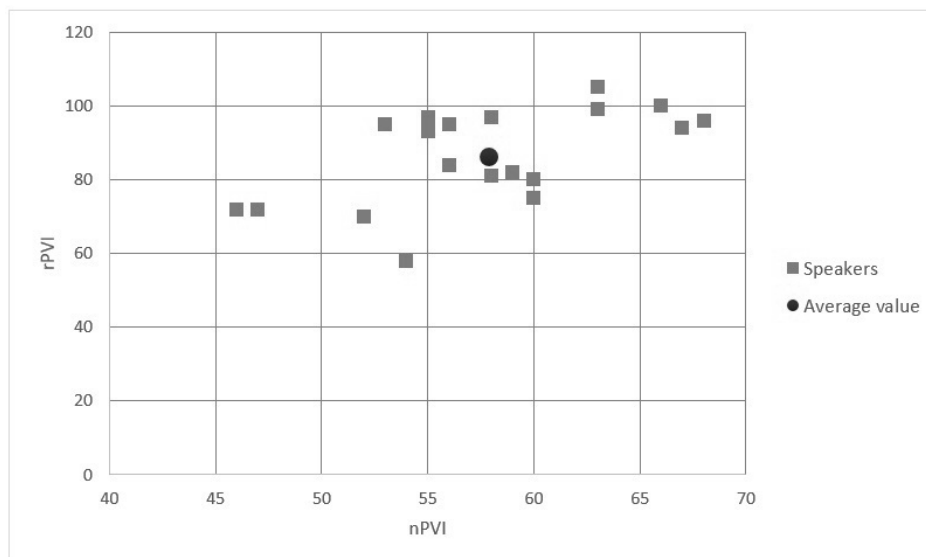

F i g u r e 6. **PVI values for 20 AusE speakers**

Now it is possible to compare the PVI data on AusE with those on other English varieties, namely, British English (BE), Singaporean English (SE) (the data are taken from Grabe, Low 2002), Pakeha English[3] and Maori[4] English (the data are borrowed from Szakay 2006).

Figure 7 shows that British, New Zealand and Australian varieties have very close but slightly different PVI values with AusE being somewhat more stress-timed than BE. Singaporean and Maori English show much lower values of the index. It is known that nativised varieties of English tend to be more syllable-timed, as they are strongly affected by their syllable-timed substrate languages.
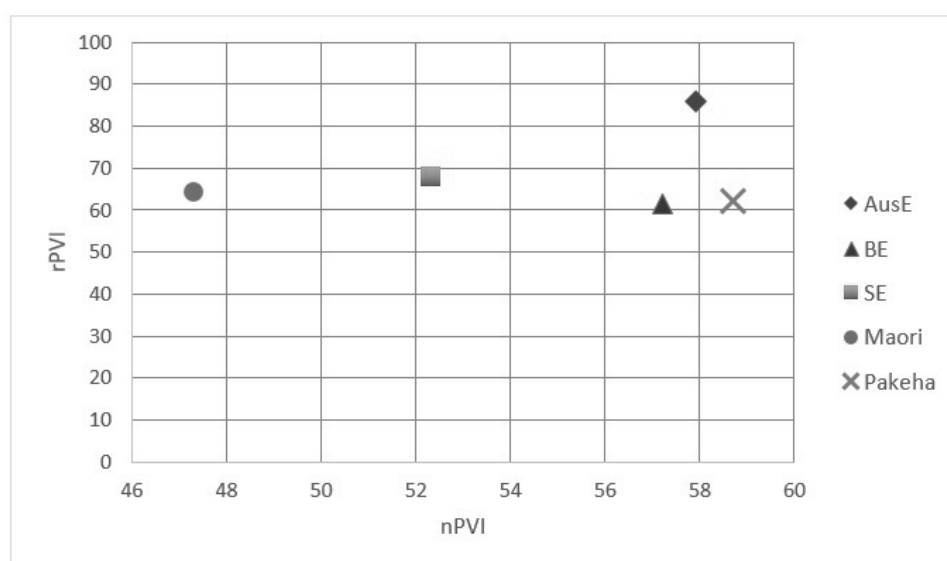

F i g u r e 7. **Distribution of varieties of English over the rPVI-nPVI plane**

---

[3] Pakeha English – New Zealand English spoken by the population of European descent.

[4] Maori – the aboriginal people of New Zealand.

Average nPVI and rPVI values for male and female speakers were then computed. The results show that the average nPVI value for AusE male speech equals 55.8 points vs 60.0 points for female speech while the average rPVI value for male speech reaches 82.1 points vs 80.0 points for female speech.
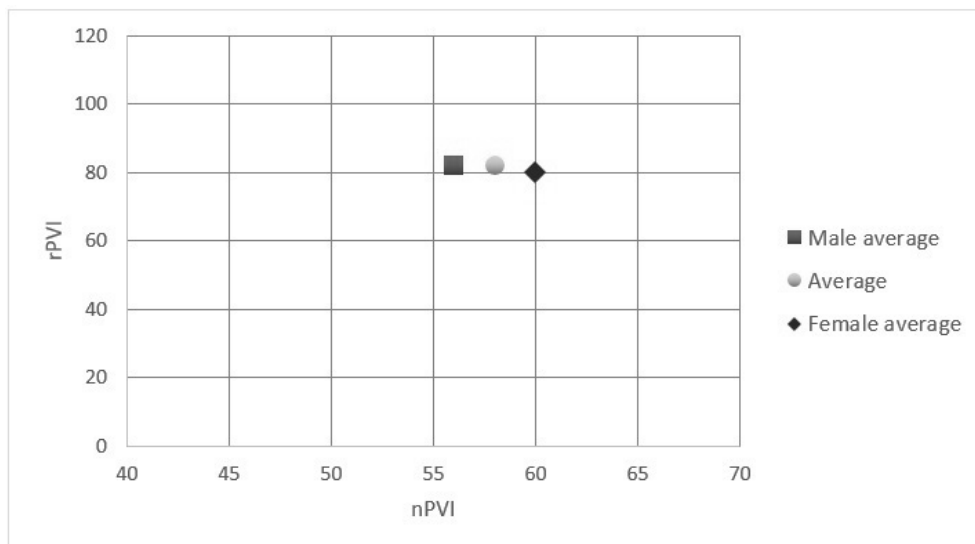

F i g u r e 8. **AusE gender differences (rPVI-nPVI ratio)**

The position of the metric values in figure 8 visualises gender differentiation in AusE speech rhythm. Both vocalic and consonantal PVI values vary for male and female speakers. The figure also shows that AusE female speech tends to be more stress-timed than male speech, as its average nPVI value is higher.

6.3. Delta and PVI values based on voice timing

The research is also an attempt to test the metrics based on the voice timing segmentation pattern [Dellwo, Fourcin, 2007]. As mentioned earlier, Fourcin and Dellwo promote the idea of voiced (VO-) and unvoiced (UV) intervals. They are defined on the basis of the F0 measurements: intervals between two F0 markers were labelled 'unvoiced' (UV), sequences of F0 markers were labelled 'voiced' (VO). The authors, however, have not provided any data obtained when using these metrics. Thus, it does not seem possible to compare the results on AusE with the results on other English varieties. However, we found it reasonable to apply this method for gender differentiation.

6.3.1. Delta values based on voice timing

The measurements of the average %VO- ΔUV ratio were computed first. Then the same metrics were calculated for male and female speech. The findings show, that the average ΔUV value for male speech amounts to 5.2 points vs 5.3 points for female speech and the average male %VO value equals 65.5 points vs 60.3 points for female speech.

The position of the metrics in figure 9 suggests gender differences in AusE speech-rhythm. The figure also convincingly demonstrates that AusE female speech appears to be more stress-timed, than male speech, as it displays lower %VO and higher ΔUV values compared to the data on male speech.
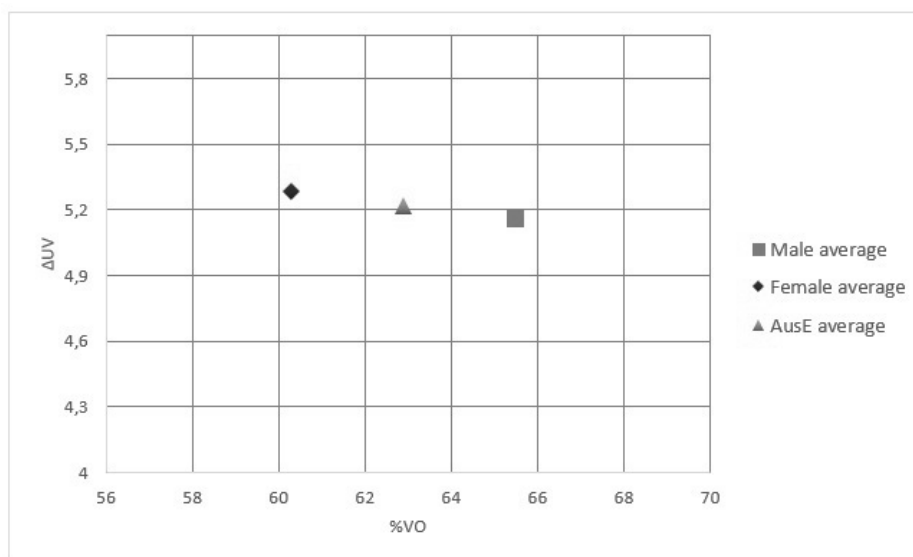
F i g u r e 9. **AusE gender differences based on voice timing (the Deltas)**

6.3.2. PVI values based on voice timing

The same procedure was applied to the PVIs. The average UV-PVI value for AusE male speech amounts to 74.4 points vs 67.2 points for female speech. The average male VO-PVI value equals 63.8 points vs 74.5 points for female speech. The results are presented in the UV-PVI – VO-PVI plane shown in figure 10.
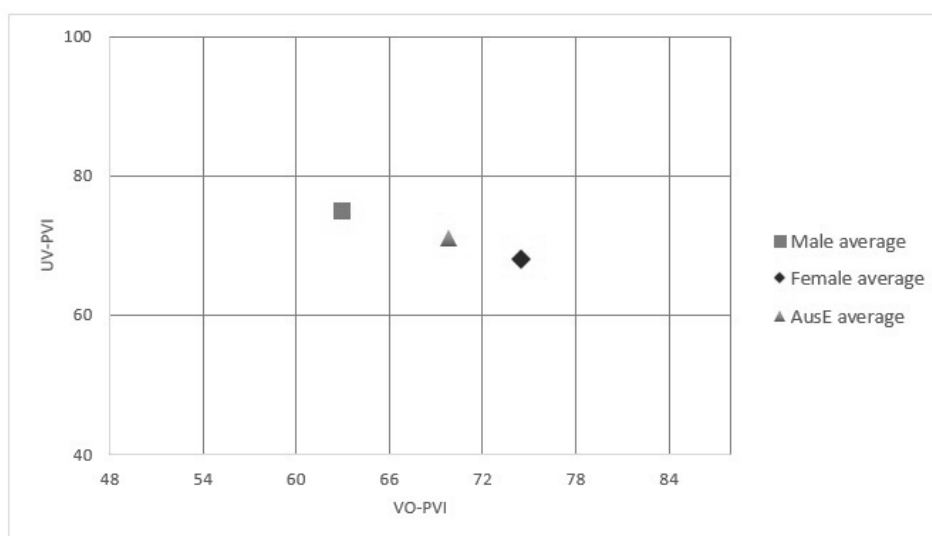

F i g u r e   1 0. **Gender differences based on voice timing (the PVIs)**

Figure 10 shows that the projections of both voiced and unvoiced PVI values are different for male and female speech with female speech being more stress-timed as its average VO-PVI value is somewhat higher.

7. Conclusion

The computation of different sets of rhythmic acoustic correlates – %V-ΔC, nPVI-rPVI, %VO-ΔUV and nPVIVO- rPVIUV – aimed at quantifying AusE speech rhythm have yielded the following results:

1. As expected, AusE belongs to the cluster of stress-timed languages. Its metrics – the Deltas and the PVIs – are very close to those of British English. Opposing tendencies were, however, recorded: the Delta method defined AusE as a more syllable-timed variant, whereas the PVI method characterized it as slightly more stress-timed than British English. The issue needs to be investigated further.

2. The application of the three tested methods to measure gender differences in speech rhythm led to comparable results, as at every step of the investigation the tendencies are all consistent with each other. This attests to the reliability of the methods, at least in gender differentiation.

3. AusE speech-rhythm has marked gender differences.

4. AusE female speech appears to be more stress-timed than AusE male speech.

References

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh : Edinburgh University Press, 1967. 203 p.

Boersma, P., Weenink, D. (2019). *Paat: Doing phonetics by computer*: Computer program (Version 6.1.01). Retrieved February 11, 2019 from <http://www.fon.hum.uva.nl>.

Callier, P. (2011). Social Meaning in Prosodic Variability. *University of Pennsylvania Working Papers in Linguistics*, 17 (1), 41–50.

Cauldwell, R. (2002). The Functional Irrhythmicality of Spontaneous Speech: A Discourse View of Speech Rhythms. *Apples – Journal of Applied Language Studies*, 2 (1), 1–24.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11 (1), 51–62.

Grabe, E, Low, T. L. (2002). Durational variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven, N. Waner (Eds.), *Papers in Laboratory Phonology 7*. (pp. 515–546). Cambridge University Press.

Fourcin, A., Dellwo, V. (2009). *Rhythmic classification of languages based on voice timing*. Retrieved September 25, 2016 from <discovery.ucl.ac.uk>id/eprint/15122/1/15122/>.

Mariano, P., Romano, A. (2010). Un confront tra diverse metriche ritmiche usando Correlatore. In S. Schmid, M. Schwarzenbach, D. Studer (Eds.), *La dimensione temporal del parlato : Proc. of the V National AISV Congress, 4–6 February* (pp. 79–100). University of Zurich, Collegiengebaude, Torriana (RN): EDK.

Nazzi, Th., Bertoncini, J., Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24 (3), 756–766.

Nokes, J., Hay, J. (2012). Acoustic correlates of Rhythm in New Zealand English: A Diachronic Study. *Language Variation and Change*, 24 (1), 1–31.

Pike, K. (1945). *The Intonation of American English*. Ann Arbor : University of Michigan Press.

Ramus, F., Nespor, M., Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73 (3), 265–292.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73–79). London : Edward Arnold.

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In P. Warren, C. Watson (Eds.), *Proc. of the 11th Australian International Conference on Speech Science & Technology* (pp. 421–426). Retrieved April 2016 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.551.4291&rep=rep1&type=pdf>.