TEMPORAL-BASED SPEAKER SEX DIFFERENCES IN READ SPEECH: A SOCIOPHONETIC APPROACH

Sandra P. Whiteside
Speech Science, University of Sheffield, S10 2TA, England, UK.

ABSTRACT

This paper summarises the findings of a study investigating temporal-based speaker sex differences. Measures taken include sentence durations, syllable rates, consonant elisions, vowel reductions, VOTs of plosives and durations of fricatives. The findings are discussed within a sociophonetic framework.

INTRODUCTION

Speaker sex differences have been shown to exist in the acoustic signals of read speech. Some studies have looked at fundamental frequency differences and formant frequency differences [1, 2, 3]. Other studies have investigated differences in the glottal source [4] and first formant bandwidths and amplitudes [5]. However, there is also some evidence to suggest that differences also exist in the temporal domain [6]. For example in the TIMIT database women tend to speak more slowly than men, men tend to reduce their vowels to [ə] more often than women and women tend to release sentence-final plosives more often than men. These findings were statistically significant [6].

This paper presents the results of a preliminary acoustic-phonetic investigation into speaker sex differences in the temporal domain. It focuses on the read speech data of three men and three women speakers with a British General Northern accent.

METHOD

Subjects

Three male and three female adult native speakers of English served as speakers. All speakers came from North of England and represented a British General Northern accent which can be defined as a non-rhotic accent of Standard English characterised in the vowel system by COULD/CUD and GAS/GLASS rhyming and a tendency to retain strong vowels where RP shows weakening e.g. computer /kɒmpjuːtə/ [7].

Speech material

Ten repetitions of five sentences were read by three men and three women speakers. This made a total of three hundred sentences (150 for the men and 150 for the women). Both the sentences and the speakers formed part of the APLAWD [8] speech corpus of British General Northern (GN) accent speakers. The sentences are as follows:

Sentence One.: *George made the girl measure a good blue vase.*
Sentence Two. : *Why are you early you owl ?*
Sentence Three .: *Cathy hears a voice amongst Spar's data.*
Sentence Four. : *Six plus three equals nine.*
Sentence Five. : *Be sure to fetch a file and send theirs off to Hove.*

Recording procedures

High quality recordings were made in a sound proof studio at the University of Leeds, using recording procedures described in [8]. High quality audio cassette copies were used to digitise the speech samples onto a Macintosh LCII computer using a Farallon™ Macrecorder and Signalyze™. A sampling rate of 11kHz was used.

Analysis

Using Signalyze™ [9] the durations of each of all the three hundred sentences were measured. It must be noted that these durations did not include any of the pauses observed in the sentences as the observed pauses were subtracted from overall sentence duration. However, the incidence of pauses was noted and some of these observations are discussed below. A summary of the sentence duration results can be found in table 1. Syllable rates (syllables/second) for each of the sentences were also calculated. A summary of these results can be found in table 2. In addition each group of sentences (60 data items for each sentence) was examined for specific linguistic and acoustic phonetic phenomena as outlined below. Observations were tested for statistical significance using a statistical package (Statview ™). The results of these tests are summarised in tables 3 to 7 where an asterisk (*) indicates statistical significance.

Sentence 1 (/ ˈʤɔːʤ ˈ meɪd ðə ˈ gɜːl ˌmeʒə(ɹ)ə ˈgʊd ˈbluː ˈvɑːz/)

i) The occurrence of schwa elisions in 'measure a' was examined. Whether the speakers realized the utterance as [ˈmeʒə ə], [ˈmeʒɹə] or [ˈmeʒəɹə] was noted using auditory and acoustic analysis; ii) In addition the occurrence of pauses after 'girl' was investigated. It was predicted that the occurrence of pauses would coincide with a lengthening in the duration of 'girl'; iii) VOT values were taken for /g/ in 'girl' and 'good'; iv) the duration of [ʒ] in 'measure' was measured and whether it was voiced or devoiced was noted and v) dB differences between the amplitude peaks of the vowels /ɔ/ in 'George' and /ɑ/ in 'vase' were noted. See table 3 for statistical analyses.

Sentence 2 (/ˈwaɪ ˈɑ/ə juː ˈɜːlɪ juː ˈaʊl/)

i) Whether 'are' was fully represented as the vowel [ɑ], reduced as the schwa [ə] or elided altogether was noted. Auditory and acoustic analysis was used to make these decisions. The presence of a schwa was noted if there was a separate intensity peak in the speech pressure waveform; ii) The incidence of pauses after 'early' was noted together with iii) dB differences between the peak amplitude of /aɪ/ in 'Why' and the peak amplitude of /aʊ/ in 'owl'. See table 4 for statistical analyses.

Sentence 3 (/ˈkæθɪ ˈhɪəz ə ˈvɔɪs əˈmʊŋst ˈspɑːz ˈdeɪtə/)

i) Pausing after 'voice' was noted; ii) In addition the duration of 'voice' was measured for each speaker; iii) The duration of the /sts/ cluster in 'amongst Spar's' was measured; d) Whether speakers realized this cluster as a reduced form ([s s], [s]) or as a full representation([sts]) was also noted. The criteria used for these judgements included auditory and acoustic analysis. Speech pressure waveforms were used for the acoustic criteria where: [sts] was realized as fricative followed by a closure phase and a subsequent pulse/transient which was followed by a fricative; [s s] was realized as two fricatives separated by a reduction in amplitude in the speech pressure waveform and [s] was realized as a single fricative; iv) VOTs were measured for /k/ in 'Cathy' and /d/ in 'data'; v) the duration of /z/ in 'hears' was measured, it was also noted whether this segment was devoiced or not; vi) the duration of /s/ in 'voice' was measured and vii) dB differences between /æ/ in 'Cathy' and /eɪ/ and /ə/ were also measured. See table 5 for statistical analyses.

Sentence 4 (/ˈsɪks plʊs ˈθriː iːkwəlz ˈnaɪn/)

i) The vowel pair in /ˈθriː ˈiːkwəlz/ was examined to see whether speakers had fully realized the vowels acoustically or reduced them to a single vowel. Auditory analyses were also used in this procedure; ii) the duration of /ˈθriː ˈiːz/ was also measured; iii) the occurrence of a pause after 'three' was examined for each of the 60 sentences; iv) duration measurements were made for: initial and final /s/ in 'six', the word 'nine', /s/ in 'plus' and /z/ in 'equals'; v) whether /z/ in 'equals' was devoiced was noted; vi) dB differences between the vowel peaks in 'six' and 'nine' were measured and vii) whether speakers showed pre-plosive glottalization in 'six' was also noted. Statistical analyses for this sentence are given in table 6.

Sentence 5 (/bɪ ˈʃɔː tə ˈfeʧ ə ˈfaɪl ən(d) ˈsend ˈðeəz ɒf tə ˈhəʊv/)

i) Whether speakers paused after 'file' was noted; ii) the duration of the word 'file' was measured for each of the 60 sentences using the speech pressure waveform and auditory analysis; iii) duration measurements were taken for /ʃ/ in 'sure', /ʧ/ in 'fetch', /s/ in 'send' and /z/ in 'theirs'; iv) it was also noted whether /z/ was devoiced or not; v) dB differences between the vowel peaks of /ɔ/ in 'sure' and /əʊ/ in 'Hove' were measured and vi) the occurrence of glottalization in 'fetch' was noted. See table 7 for statistical analyses.

RESULTS AND DISCUSSION

Table 1 shows that sentence durations are longer for the women versus the men, with the women showing larger standard deviations. These results agree with those of Byrd [6]. The women also show lower syllable rates than the men as shown in table 2.

Table 1. Mean sentence durations and standard deviations of sentences 1 to 5 by speaker sex

	Men		Women	
Sent.	Mean (ms)	s.d. (ms)	Mean (ms)	s.d. (ms)
1	2479.1	113.9	3045.2	483.2
2	1420.6	94.1	1872.2	259.1
3	2325.9	131.2	2978.4	493.0
4	1706.6	177.2	2124.0	236.7
5	2705.1	224.0	3442.0	541.3

Table 2. Mean syllable rates (syllables per second) and standard deviations (s.d.) of sentences 1 to 5 by speaker sex

	Men		Women	
Sent.	Mean	s.d.	Mean	s.d.
1	3.853	0.264	3.234	0.393
2	4.735	0.434	3.806	0.51
3	4.313	0.242	3.44	0.592
4	3.144	0.14	2.844	0.341
5	4.466	0.378	3.584	0.644

Table 3. Statistical analyses for Sentence 1

Single factor ANOVA & X2 test results
speaker sex (SS) & sentence durations F=38.997, p =0.0001*
schwa elisions & sentence duration F=28.17, p=0.0001*
SS & schwa elisions F=0.62, p=0.4343
pausing after 'girl' & SS χ^2 =13.469, p=0.0002*
pausing after 'girl' & longer durations of 'girl F=136.017, p=0.0001*
VOT /g/'girl' & SS F=0.0004, p=0.957
VOT /g/ 'good' & SS F=8.341, p=0.0055*
Duration (ms) [ʒ] 'measure' & SS F=65.354, p=0.0001*
Duration [ʒ] 'measure' & voiced/ devoiced F=29.54, p= 0.0001*
Voiced/devoiced [ʒ] 'measure' & SS χ^2=23.72, p=0.0001*
dB difference /ɔ/ in 'George' & /ɑ/ in 'vase' & SS F=4.867, p=0.0313*

Table 4. Statistical analyses for Sentence 2

Single factor ANOVA results
speaker sex (SS) & sentence durations F=80.529, p= 0.0001*
SS & schwa elisions F=16.789, p = 0.0001*
SS & full representation of 'are' F=12.069, p = 0.001*
vowel reductions & sentence duration F=39.576, p = 0.0001*
schwa elisions & sentence duration F=8.511, p = 0.005*
pauses after 'early' & SS F=29.696, p = 0.0001*
dB differences 'Why' & 'owl' F=1.361, p=0.2481

For Sentences 1 to 4 (tables 3 to 6), this small group of speakers shows a link between speaker sex and pausing. In addition, there is also a link between the occurrence of pauses and longer duration values when words precede a pause. The findings here differ from those of Byrd [6] who found that there was no link between speaker sex and the occurrence of pauses. However what is interesting to note is that the findings here mirror some of the evidence of previous research which has shown that men tend to pause less frequently than women during a conversational speech setting. By not pausing men tend to dominate a conversation as this reduces both turn taking and any interruptions. Conversely women tend to pause more thus allowing themselves to be interrupted more frequently [10 &11].

The differences between Byrd's [6] findings and those here could be due to cultural differences between British English and American English speakers. That cultural differences exist in conversational style has been reported elsewhere [11]. However it is also possible that these differences are purely a result of individual speaker variation in the speakers investigated this study.

Table 5. Statistical analyses for Sentence 3

Single factor ANOVA & X2 test results
speaker sex (SS) & sentence durations F=49.087, p = 0.0001*
SS & pausing after 'voice' F=11.505, p= 0.0013*
SS & duration of 'voice' F=14.132, p=0.0004*
SS & /sts/ cluster durations F=25.67, p=0.0001*
SS & cluster reductions F=34.208, p = 0.0001*
/sts/ reductions & /sts/ durations F=47.823, p = 0.0001*
VOT /k/ 'Cathy & SS F=3.451, p=0.0683
Voiced/devoiced /z/ 'hears' & SS χ2= 30.0, p=0.0001*
Duration (ms) /z/ 'hears' & SS F=10.529, p=0.002*
Duration (ms) /s/ 'voice' & SS F=22.348, p=0.0001*
VOT /d/ 'data' & SS F=3.924, p=0.0529
Voiced/devoiced /z/ 'hears' & duration (ms) /z/ F=32.492, p=0.0001*
dB differences between /æ/ in 'Cathy' & /eɪ/ in 'data' & SS F=14.007, p=0.0004*
dB differences between /æ/ in 'Cathy' & /ɔ/ in 'data' F=38.545, p=0.0001*

Tables 6 and 7 showed no speaker sex differences in glottalization and few differences in VOTs. Tables 3 to 7 however, do show that the men in this study tended to either elide or reduce both vowels and consonants which contributed to shorter sentence durations. Conversely the women showed a tendency to realise speech segments more fully. This therefore meant

Table 6 Statistical analyses for Sentence 4

Single factor ANOVA & X2 test results
speaker sex (SS) & sentence durations F=59.753, p = 0.0001*
SS & vowel pair reductions F=39.479, p = 0.0001*
pauses after 'three' & duration of 'three' F=34.952, p = 0.0001*
SS & pauses after 'three' F=13.576, p = 0.0005*
Duration (ms) initial /s/ in 'six' & SS F=9.371, p=0.0033*
Duration (ms) final /s/ in 'six' & SS F=82.119, p=0.0001*
Duration (ms) of 'nine' & SS F=17.413, p=0.0001*
Duration (ms) /z/ in 'equals' & SS F=116.391, p=0.0001*
Voiced/ devoiced /z/ in 'equals' χ^2=16.463, P=0.0001*
Voiced/ devoiced /z/ in 'equals' & its duration value F=23.572, p=0.0001*
Duration (ms) /s/ in 'plus & SS F=14.288, p=0.0004*
dB differences between 'six' & 'nine' & SS F=1.599E-17, p=1
Glottalization in 'six' & SS All speakers showed glottalization effects

Table 7 Statistical analyses for Sentence 5

Single factor ANOVA & X2 test results
speaker sex (SS) & sentence duration F=34.952, p = 0.0001*
speaker sex & pauses after 'file' F=2.61, p = 0.1116
pauses after 'file' & duration of 'file' F=16.859, p = 0.0001*
Duration (ms) /ʃ/ in 'sure' & SS F=40.972, p=0.0001*
Duration (ms) /ʧ/ in 'fetch' & SS F=15.064, p=0 0003*
Duration (ms) /s/ 'send' & SS F=51.992, p=0.0001*
Duration /z/ 'theirs' & SS F=36.727, p=0.0001*
Voiced/ devoiced /z/ 'theirs' & SS χ^2=23.254, p=0.0001*
Duration /z/ & voiced/devoiced F=9.998, p=0 0025*
dB differences between /ɔ/ in 'sure' & /əʊ/ in 'Hove' & SS F=27.608, p=0.0001*
Glottalization in 'fetch' & SS All speakers showed glottalization effects

that the women's speech segments were on average longer than those of the men. These results agree with the finding that women enunciate more clearly than men [10]. We can argue that enunciating more clearly requires greater articulatory effort. From this we can suggest that the significant differences (tables 3 to 7) in the dB ratios for sentence-initial and sentence-final syllable-nuclei for the men and women speakers reflect greater articulatory effort by the women speakers, who had lower dB ratios. The findings may also reflect the different strategies men and women adopt in a conversational setting. However this is pure conjecture at this stage and reflects the need for further research.

The results of this preliminary investigation provide some acoustic-phonetic evidence that the men and the women in this data sample realise sentences differently when they are read in a controlled laboratory situation. Further research is planned using another British English database.

REFERENCES

[1] Peterson, G. E. and Barney, H. L. (1952). Control methods use in the study of vowels, *Journal of the Acoustical Society of America, 24,* 175-184.

[2] Wu, K. and Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis, *Journal of the Acoustical Society of America, 90(4),* 1828-1840.

[3] Childers, D. G. and Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis, *Journal of the Acoustical Society of America, 90(4),* 1841-1856.

[4] Fant, G. (1979b). Temporal fine structure of formant damping and excitation, *Speech Communication Papers, Acoustical Society of America,* 161-165.

[5] Karlsson, I. (1988). Glottal Waveforms for normal female speakers, *Speech Transmission Laboratory-Quarterly Progress and Status Report, Royal Technical Institute, Stockholm,* 31-36.

[6] Byrd, D. (1992). Preliminary results on speaker-dependant variation in the TIMIT database, *Journal of the Acoustical Society of America, 92(1),* 593-596.

[7] Wells, J. C. (1982). *Accents of English 2.* Cambridge University Press.

[8] Lindsey, G., Breen, A & Nevard, S. (1987). Spar's Archivable Actual-Word Databases, *UCL Report on SPAR Project.*

[9] Keller, E. (1992). *Signal Analysis for Speech and Sound.* InfoSignal™ Inc.

[10] Elyan, 0. (1978). Sex differences in speech style, *Women Speaking, 4,* 4-8.

[11] Tannen, D. (1992). *You just don't understand,* Virago Press.