The influence of body posture on the acoustic speech signal

Yvonne Flory[1], Francis Nolan[1]

[1]Department of Theoretical and Applied Linguistics, University of Cambridge
yvflory@gmail.com

ABSTRACT

Few studies address the influence of postural changes on the vocal mechanism, and where they do so, they most often neglect the acoustic implications. This study tested participants in upright, prone and supine posture and assessed the resulting changes in formant frequencies and F0. Results show that while most formants appear to be subject to compensation strategies by the speaker, F3 differs consistently between body postures. F0 as well rises in non-neutral postures.

This work is of relevance to the forensic-phonetic framework, which is reliant on higher formant frequencies in particular to carry out forensic speaker comparison. Further research is needed in order to assess articulatory movement with respect to its acoustic correlates.

Keywords: Forensic speaker comparison, acoustic phonetics, articulatory phonetics, posture variation

1. INTRODUCTION

Several studies have been carried out to assess the impact of posture on the vocal mechanism. Most of these studies have focused on articulator displacement of the tongue [5, 6, 11, 12, 14], jaw [10, 12] and lips [5, 14]. The tongue has been found to counteract the effect of gravity in supine posture [5, 8, 11]. The jaw as well showed backwards displacement in supine posture [10, 12 (for one of their two subjects)] and rotated towards or away from occlusion in prone and supine posture, respectively [10]. Other studies found that compensatory behaviour differed between speakers, and that some even overcompensated for gravitational pull [12]. Similarly, individual sounds exhibit differing behaviour: back vowels were more susceptible to gravitational pull in Kitamura et al. [5] than front vowels, which they explain with articulatory anchoring of the latter to the hard palate. Shiller et al. [10] also found that differences between upright and supine posture were not as pronounced as between upright/supine and prone posture.

Only a few studies also assessed acoustic parameters with respect to body posture changes, and those which did found contrasting results [10, 11, 12]. Tiede et al. [12] did not find differences in the first three formants, and suggest the overall configuration cancels out any anatomical differences. Stone et al. [11] did not find differences in formant frequencies either, but Shiller et al. [10], on the other hand, measured differences in F1 and F2 for the vowels /e/ and /æ/.

Unfortunately, the other studies did not address the potential correlation between anatomical and acoustic variation. Modelling this kind of variation is especially important for the forensic-phonetic framework with respect to forensic speaker identification or comparison. The greater the amount of within-speaker variation in two or more samples that are compared, the more difficult it is to make a statement about the probability of them having been spoken by one or two speakers. However, the more of this variation that can be explained by external factors, the better any estimation will be. F0 is probably the most variable acoustic parameter due to its linguistic and paralinguistic signalling function.

Formant frequencies also vary within speakers according to a number of factors, such as coarticulatory variation, channel restrictions [6], or perturbations of the vocal tract. Higher formant frequencies in particular are indicative of vocal tract shape, therefore a change to the shape can be expected to yield a change in frequency of F3 and F4. Even higher formants can be disregarded, as they are often not visible even in studio recordings. More often, recording conditions are worse than this in the forensic context.

The present study will investigate the acoustic differences induced by changes in body posture. Formant frequencies of the first four formants as well as fundamental frequency will be measured. The results will be discussed with respect to their implications for forensic speaker comparison.

2. EXPERIMENT

2.1. Methodology

2.1.1. Subjects

24 native speakers of British English were recorded, 12 male and 12 female. They were all between 18 and 30 years old (mean 22.13), grew up in the South-East of England and spoke with an SSBE pronunciation. None had any known neck or back

problems, or had been diagnosed with speech or hearing problems in the past.

2.1.2. Stimuli

The stimuli were embedded in highly controlled carrier sentences. The aim was to measure F0 and formant frequencies in suitable sounds. The sounds used were vowels, liquids, and glides (Table 1). Consonants adjacent to the vowels were controlled for in the statistical analysis with respect to place of articulation and voicing (/b, p, d, t, g, k/). Liquids were controlled for vowel context. [l] and [ɹ] were followed by the following vowels: /a, iː, uː/. [ɫ] was preceded by /iː, ɪ, ɛ, ɔː, uː, ʊ/.

Table 1: Stimulus sounds and the carrier sentences in which they were embedded.

Sound	Carrier sentence
/ɑː/	Say **C** + **/ɑ**ː**/** + **C** + **/ɑ**ː**/** please.
/iː/	Say **C** + **/i**ː**/** + **C** + **/i**ː**/** please.
/uː/	Say **C** + **/u**ː**/** + **C** + **/u**ː**/** please.
/j/	Say **/ɑ**ː **j ɑ**ː**/** please.
/w/	Say **/ɑ**ː **w ɑ**ː**/** please.
[l]	This is a **/f/** + **[l]** + **V** + **/t/.**
[ɫ]	It's **/f/** + **V** + **[ɫ]** time.
[ɹ]	This is a **/f/** + **[ɹ]** + **V** + **/t/.**

2.1.3. Procedure

Recordings were made in a sound insulated booth, using a Sennheiser MKH 40 P48 condenser microphone with a muffler and a Marantz PMD670 Portable Solid State Recorder, which saved each file in .wav format straight onto a CompactFlash memory card. The files were recorded mono through an XLR cable and had an audio sample resolution of 16 bit and a sampling rate of 44 kHz.

Each subject was asked to read stimuli from a PowerPoint presentation, at a self-controlled speed, in all three postures: upright, prone, and supine. The slides were randomised and contained one sentence each. The whole presentation contained three repetitions of each stimulus, except for the glides, which were presented five times each.

In both supine and prone condition, subjects lay on a table, cushioned with blankets. In the supine condition, they had a pillow to support and stabilise their neck, for additional comfort and to discourage unnecessary movement. In the prone condition their forehead rested on a flat foam cushion for comfort, while a gap was left between the two tables their forehead and body rested on, respectively. This gap was large enough to allow their jaw to move freely, but small enough to be reasonably comfortable.

The microphone was placed at a target distance of 5 cm to the side of their mouths in each condition, to aid the muffler in reducing any unwanted noise to a minimum and at the same time preventing high energy peaks in the signal from bilabial sounds. For the same reason subjects were asked to move as little as possible, and if it was absolutely necessary, to do so between sentences. Subjects were generally successful in remaining still, thus the distance to the microphone remained largely constant.

2.1.4. Acoustic analysis

Formant frequency measurements were taken in the centre of the vowels, glides and [l]. For [ɫ] and [ɹ], segmentation from the adjacent vowels was not always possible. Therefore, 11 markers were set throughout the sequence at 10%-intervals. Values were measured at each of these intervals, and the curves analysed with respect to the average location of the liquid; the mean of these intervals was used in the analysis. F0 was measured in the same manner.

2.1.5 Statistical analysis

All analyses were done in SPSS [4] using a repeated-measures design with Body Orientation as the main factor, and sex as a between-subjects factor. A repeated-measures ANOVA was also performed for each sex separately, because even if there were no interactions between Body Orientation and Sex, there could still be differences between sexes regarding effect size. Similarly, every analysis was split by Vowel or Vowel Context, except F0 analysis.

Sphericity violations of the homogeneity of variances were corrected using ε as calculated by the conservative Greenhouse-Geisser correction. F-values that violate sphericity are marked with asterisks in the following.

Differences between individual postures were calculated using the Bonferroni post-hoc test. This test is claimed to be the most robust where sphericity is violated (page 472, [1]). Brackets within figures denote significant results of these post-hoc tests. Where there are no brackets, the overall ANOVA was significant (unless otherwise mentioned), but none of the individual contrasts were.

2.2 Hypotheses

F0 was predicted to rise in prone posture, due to increased tension in the neck. Conversely, supine posture might induce higher values, but because less tension is expected, values were expected to remain similar to those in upright posture.

For the lower formants F1 and F2, compensation along the lines of a task dynamic model [9, 13] might be expected. They should therefore remain more stable between postures in order to maintain intelligibility. The higher formants F3 and F4 should be less susceptible to compensation strategies, and therefore differ more. However, the movement of higher formant frequencies is harder to predict.

With respect to differences across the sounds investigated, sounds within close proximity to neighbouring sounds in the phonetic space were expected to differ less than those that are further apart from their nearest neighbours. Therefore, /ɑː/ was predicted to exhibit more differences between postures than /iː/ or /uː/. /iː/ may also be subject to articulatory anchoring to the hard palate and molars, which is suspected of stabilising its articulation, resulting in some resistance to postural impact.

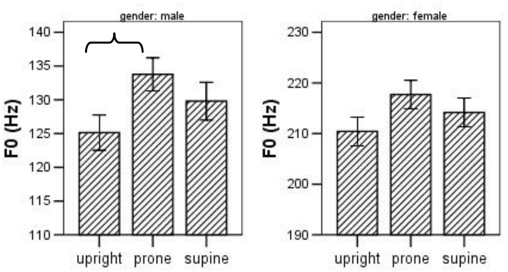
Liquids may be similarly affected by anchoring, to the extent they are in contact with the palate, which is mostly the case for the laterals. Therefore, on this basis we expect differences mostly in [ɹ] and the two glides. However, although the standard view that rhoticity is signalled by F3 has been challenged [2, 3], the likelihood that F3 is important perceptually makes it a candidate for compensation.

3. RESULTS

3.1 F0

F0 values were significantly higher in prone than in neutral body orientation for male speakers (F(2,22) = 5.66, ε = .93 , *p* < .01). Despite not being significant, the same pattern was visible for female speakers as well. Both non-standard body orientations triggered higher F0 values (Fig. 1).

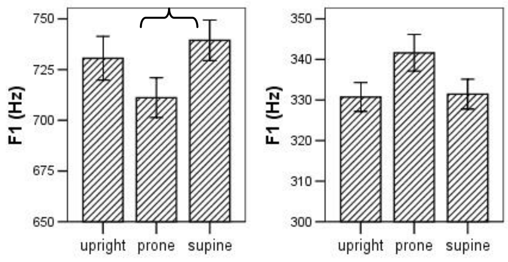
Figure 1: F0 differences for male (left) and female speakers (right).



3.2 Lower formants

Lower formants exhibited few differences between postures. F1 patterns of variation also differed across vowels.
F1 in /ɑː/ was lowest in prone posture (F(1.52, 33.4) = 5.97*, ε = .76, *p* < .05), while F1 in /uː/ was highest in prone posture, but only approaching significance (F(2,44) = 2.72, ε = .92, p = .77). See Fig. 2 for both /ɑː/ and /uː/. Both upright and supine values were similar for both vowels. /iː/ did not exhibit any differences.

Figure 2: F1 differences for /ɑː/ (left) and /uː/ (right).



F2 only showed a few erratic differences, which formed no pattern, and F2 is therefore not presented in detail here.

As F1 and F2 are vital for vowel quality, the lack of changes was expected. The focus of this paper will therefore lie on the higher formants.

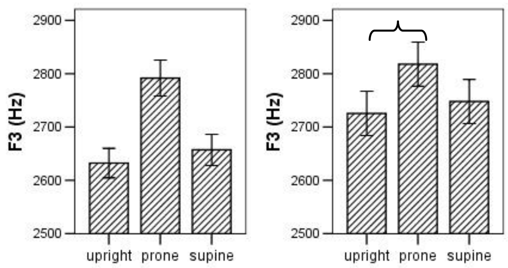
3.3 Higher formants

F3 showed the most consistent results across liquids and glides, and for some of the vowels.

Table 2: Results overview of F3 in vowels, liquids and glides, not broken down by adjacent sound. (f = only female speakers)

Sound	df1	df2	F	Sig.	ε
/ɑː/	1.27	28.04	5.37*	*p* < .05	.64
/iː/ (f)	2	22	3.19	*p* = .061	.89
/uː/	1.39	30.49	2.61*	n.s.	.69
[l]	2	44	4.55	*p* < .05	.95
[ɫ]	1.90	541.96	63.26*	*p* < .001	.95
[ɹ]	2	44	4.42	*p* < .05	.84
/j/ (f)	2	22	5.22	*p* < .05	.83
/w/	1.49	32.71	4.52*	*p* < .05	.74

Figure 3: F3 differences for /ɑː/ (left) and [l] (right)



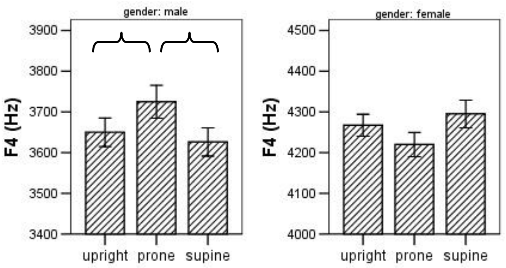
Values for F3 were highest in prone posture for /ɑː/, /iː/ (approaching significance), [l], [ɫ], [ɹ], /w/, and female speakers of /j/ (Table 2 presents a general overview of the ANOVA results; Fig. 3 shows the patterns for /ɑː/ and [l] as an exemplar). Values in upright and supine posture remained largely similar, with few exceptions.

F4 results were less systematic. Differences were only significant for [ɫ], and here only for *full* ($F(2,92) = 6.23$, $\varepsilon = .99$, $p < .01$), while values approached significance for *feel* ($F(1.58,72.80) = 2.68$, $\varepsilon = .79$, $p = .87$). However, there were significant interactions between Body Orientation and Sex for the words *feel*, *fill*, *fell* and *fall*. This was the result of female speakers' values being significantly lower in prone than in supine orientation, while male speakers exhibited the opposite pattern (Table 3; Fig. 4). These opposing trends will have cancelled out an overall effect with both sexes analysed together.

Table 3: Results overview of the opposing patterns between sexes for F4 in [ɫ].

Word	sex	df1	df2	F	Sig.	ε
feel	m	1.31	30.00	1.44*	n.s.	.65
	f	2	46	5.34	$p < .01$	.93
fill	m	2	46	1.50	n.s.	.98
	f	1.55	35.75	3.77*	$p < .05$	.78
fell	m	2	46	8.93	$p < .001$	.91
	f	1.37	31.55	5.27*	$p < .05$	.69
fall	m	2	46	7.10	$p < .001$	.98
	f	2	46	.12	n.s.	.96

Figure 4: F4 differences in *fell* for male (left) and female speakers (right)



4. DISCUSSION

F0 values were indeed higher in prone posture, and remained largely similar in supine posture, which was expected. It remains open whether this was the result of increased tension, or subglottal pressure, or both.

Formant values also behaved largely as expected. F1 and F2 are important for intelligibility of the utterance, and were therefore, we presume, subject to compensatory strategies. F1 did differ between postures in a few instances, and the patterns were different across vowels. This could be explained by the different impact of constrictions on the standing waves in the vocal tract. For /ɑː/, the constriction is made far back in the pharynx close to an F1 velocity node, which is responsible for its high F1. If this constriction is weakened by forward displacement of the tongue, F1 is expected to lower, which was the case in this study. The opposite may be postulated in the case of /uː/, whose low F1 will rise when the constriction is weakened.

The consistency of F3 differences across body postures indicates that speakers do not fully compensate for changes to the configuration of their vocal tract. The effect was small, but persisted across the majority of sounds investigated. Even liquids, which were expected to be less affected due to anchoring to the palate, showed strong effects.

F4 showed fewer and more erratic results with respect to body posture. The opposite patterns of F4 changes across speaker sex may be the result of the non-isomorphic vocal tracts in male and female speakers – male speakers for instance having proportionately longer pharynxes. With resonances of shorter wavelength, as is the case in F4, lack of exact equivalence in constriction location relative to the distribution of acoustic nodes can result in major differences in formant frequency.

For all formants, prone posture yielded the largest differences, in contrast to supine versus upright posture. Speakers have been shown to become familiar with perturbations to their vocal tract [7], and to be able to develop compensatory strategies. Less 'normal' postures such as lying prone might therefore be expected to provide more of a challenge for immediate compensation, with the consequence that formant frequencies differ more from those in an upright posture than do those in the more common supine posture.

5. CONCLUSION

The results presented in this paper contribute to our understanding of within-speaker variation, and are therefore of relevance to forensic speaker comparison. Overall, posture had relatively small and inconsistent effects on parameters often used in speaker comparison. However, in the case of F3 in particular, some clear effects emerged which forensic practitioners will need to be aware of.

In terms of phonetic theory, our data provide a new perspective on compensatory articulation. Future work will benefit from synchronous articulatory and acoustic recordings and analysis.

6. REFERENCES

[1] Field, A. 2011. *Discovering statistics using SPSS*. London: Sage.

[2] Heselwood, B. 2009. Rhoticity without F3: Lowpass filtering, F1-F2 relations and the perception of rhoticity in NORTH-FORCE, START and NURSE words. *Leeds Working Papers in Linguistics & Phonetics 14*, 49 – 64.

[3] Heselwood, B., Plug, L. 2011. The role of F2 and F3 in the perception of rhoticity: Evidence from listening experiments. *Proc. of the 17th ICPhS Hong Kong*, 867 – 870.

[4] IBM Corp. 2012. SPSS for Windows (Version 21.0). [Software]. Armonk, NY: IBM Corp.

[5] Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., ... Senda, M. 2005. Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner. *Acoustical Science and Technology, 26*(5), 465 – 468.

[6] Künzel, H. J. 2001. Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics*, *8(*1), 80 – 99.

[7] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: W. J. Hardcastle, A. Marchal (eds.): *Speech production and speech modelling* (pp. 403 – 439). Dordrecht & London: Kluwer Academic Publishers.

[8] Niimi, S., Kumada, M., Niitsu, M. 1994. Functions of tongue-related muscles during production of the five Japanese vowels. *Annual Bulletin of the Research Institue of Logopedics and Phoniatrics, 28*, 33 – 40.

[9] Saltzman, E. L., Munhall, K. G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333 – 382.

[10] Shiller, D. M., Ostry, D. J., Gribble, P. L. 1999. Effects of gravitational load on Jaw movements in speech. *Journal of Neuroscience, 19*(20), 9073 – 9080.

[11] Stone, M., Stock, G., Kumar, K., Epstein, M., Kamhamettu, C., Li, M., … Prince, J. 2007. Comparison of speech production in upright and supine position. *The Journal of the Acoustical Society of America, 122(*1), 532 – 541.

[12] Tiede, M. K., Masaki, S., Vatikiotis-Bateson, E. 2000. Contrasts in speech articulation observed in sitting and supine conditions. *Proc. 5th Seminar on Speech Production*, 25 – 28.

[13] Turvey, M., Saltzman, E. L., Schmidt, R. C. 1991. Dynamics and task-specific coordinations. In: N. I. Badler, B.A. Barsky, D. Zeltzer (eds.): *Making them move: Mechanics, control, and animation of articulated* figure (pp. 157 – 170). San Mateo, CA: Morgan Kaufmann.

[14] Wrench, A., Cleland, J., Scobbie, J.M. 2011. An ultrasound protocol for comparing tongue contours: Upright vs supine. *Proc. 14th ICPhS Hong Kong*, 2161 – 2164.