THE ARTICULATORY CORRELATES OF DESCRIPTIVE CATEGORIES FOR SUPRALARYNGEAL VOICE QUALITIES

Francis Nolan[1] *and Barbara Kühnert*[1,2]

[1]*Department of Linguistics, University of Cambridge*
[2]*Institut für Phonetik und sprachliche Kommunikation, University of Munich*

ABSTRACT

This paper reports a pilot experiment using EMA (electromagnetic articulography) to monitor the 'articulatory settings' hypothesised to underlie supralaryngeal components of voice quality as defined in Laver's [1] descriptive framework. The experiment also tests whether segmental articulation is modified to preserve acoustic targets under changes of articulatory setting.

1 INTRODUCTION

'Voice quality' is used here in the sense of Abercrombie [1]: '...a quasi-permanent quality running through all the sound that issues from [a speaker's] mouth' (p.91). Laver [2] proposes an analysis for voice quality based on auditorily identified components, each hypothesised to be associated with an 'articulatory setting', a long-term biasing of the vocal organs towards a given configuration. Auditory components such as *breathy*, and *palatalised*, would result from a general tendency to adduct the vocal cords less strongly, and to bias the configuration of the tongue towards [i], both relative to neutral baselines.

Laver's descriptive framework 'stands on an auditory foundation' (despite the superficially articulatory labels such as *palatalised*), but 'the auditorily-identified components all have correlates ... capable of instrumental verification – [at] the articulatory, physiological, and acoustic levels' (p.7). Although much research has been done on correlates for laryngeal components, there has been little on the correlates of supralaryngeal components. Nolan [3] carried out an acoustic analysis of suprasegmental components in Laver's framework, and found systematic shifts in formant frequencies. Esling [4] also looked at spectral correlates of components such as *velarised* and *laryngo-pharyngalised*. Neither study included articulatory measurement. The first aim of the present experiment is to test the articulatory settings claimed to underlie *palatalised* and *velarised* voice.

Although a voice quality component can be seen as resulting from a bias in articulatory activity which pervades the whole of a person's speech, segmental requirements may impose limitations. For instance, *nasalisation* may be reversed by fricatives and plosives. But a more intriguing segment-setting interaction would be if a segment's articulation adjusted to preserve an acoustic requirement despite an 'unhelpful' setting. English [ɹ] and [ʃ] provide a potential example. The low F3 and/or F4 of [ɹ] and the relatively low first major spectral prominence of [ʃ] depend on a sufficiently large cavity in front of the major constriction. This is achieved by a post-alveolar constriction, augmented for many speakers by a degree of lip rounding and protrusion. If the acoustically effective size of the cavity is reduced in size, as in smiling, the tongue constriction may compensate by moving back, as noted by Andrew Crompton (pers. comm.). The second aim of the experiment is to explore this kind of compensatory articulation under *lip-rounded* and *lip-spread* settings.

2 EXPERIMENT

The first author acted as subject. He has extensive familiarity with Laver's framework, and has attended a training workshop in its use. Two sentences were used. The first, 'When the sunlight strikes raindrops in the air they act like a prism and form a rainbow', is part of the 'Rainbow' passage used in Nolan [3]. The sentence was read three times implementing each of the components *palatalisation* and *velarisation*, and three times *neutrally* as a control. The second sentence was 'The red rooster shattered the rural quiet with three very short shrieks', designed to contain several examples of [ɹ] and [ʃ]. This was

produced three times implementing each of *lip-rounding*, *lip-spreading*, and *neutral*. Only by using controlled performances by a phonetician is it possible to study the effect of a single component. In real speakers the effect of a component would be conflated with other factors, including anatomical differences. The approach adopted here is comparable to using a phonetician's Cardinal Vowels to explore the acoustic-articulatory mapping in the vowel space.

Time-aligned acoustic, EMA (Carstens AG 100), and electro-palatographic recordings were made (the latter not discussed here) in collaboration with Phil Hoole at the Institut für Phonetik und sprachliche Kommunikation, Munich. Receiver coils were attached as follows: lower lip, lower front teeth (for vertical jaw movement), tongue blade, tongue front, tongue back, and (to allow compensation for head movement) the upper front teeth and bridge of the nose. Data processing was carried out as described in Hoole [5].

3 RESULTS

Compared with a neutral tongue-body setting, the 'centre of mass' of the tongue body would be expected to be raised and slightly fronted in *palatalisation*, and raised and retracted in *velarisation* (Laver [2] p.46).

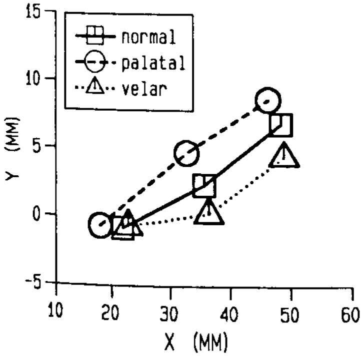

Fig 1. Mean position of EMA tongue-coils for three voice quality components.

Fig 1 shows the effect of these three settings (including neutral) on the position of the EMA coils. The x-axis shows front-back position in the occlusal plane, that is, parallel to the plane at which the teeth meet, and the y-axis shows position perpendicular to the occlusal plane. The front of the mouth is at the left. The points joined by lines are the position of the blade, front, and back coils for each setting. Each point is the mean of all analysis frames for utterances in a particular setting (approximately 2400: 4s utterance duration x 3 repetitions x 200 frames/s).

Palatalisation does, as the label suggests, involve raising and fronting of the body of the tongue. Less predictably, *velarisation* apparently brings a lowering of the tongue. What may be happening is that the tongue body is arching back and up behind the rearmost coil (placed below the join of the hard and soft palate), and the effect on the 'visible' part of the tongue is one of lowering, as the mass of the tongue retreats back. Nonetheless it does seem that the main secondary constriction is further back than the velar region, and formant data for subsets of high front and open vowels (Table 1) reveal that (particularly for the high front vowels) the values for *velarisation* in the present study are similar to those for *pharyngealisation* in the earlier recording of the same speaker.

Table 1. Formant frequencies by setting (data in italics: Nolan [3], Table 4.4).

	F1	**F2**	**F3**
	High Front Vowels		
Neutr.	396 *400*	1843 *1850*	2502 *2480*
Pal.	409 *390*	1927 *1950*	2650 *2650*
Vel.	448 *405*	1664 *1825*	2489 *2525*
Pharyn.	*465*	*1675*	2430
	Open Vowels		
Neutr.	669 *690*	1255 *1210*	2502 *2480*
Pal.	667 *670*	1562 *1430*	2650 *2650*
Vel.	649 *620*	1209 *1195*	2489 *2525*
Pharyn.	*685*	*1170*	*2460*

Laver (pp.55–6) notes there is difficulty in defining *velarisation*. Whether an ambiguity in the framework or inaccurate performance explains the present result might be determined by having listeners trained in the framework assess the present recordings.

Fig. 2 shows EMA data for the second sentence, which was realised with *neutral*, *lip-rounded*, and *lip-spread* components. For these components the lower lip coil (not shown) varies as expected, while the contour of the tongue is essentially the same. However it is noticeable that for *lip-spread*, each of the tongue coils is slightly more retracted, and the coil on the tongue front is slightly lowered. This would not be predicted from the definition of the components, which imply purely labial settings.

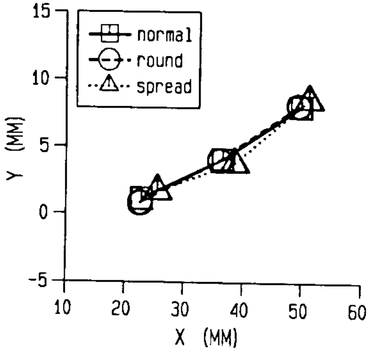

Fig 2. Mean position of EMA tongue-coils for labial voice quality components.

It is appropriate then to examine the segments for which compensatory articulation was predicted. Fig. 3 shows the average position of the tongue coils for the acoustic midpoint of [ɹ] in 'red', 'rooster', and 'very'. Fig 4 shows [ʃ] from 'shattered' and 'short'. For both sounds, the tongue is generally retracted (compare with the average coil positions for the whole utterance in Fig. 2); but it is relatively less retracted in lip-rounded, and more retracted in lip-spread; and in [ʃ] the blade is also lowered in lip-spread, suggesting retroflexion. The extra retraction (and retroflexion) would help to maintain the acoustic lowering effect otherwise reduced by the lip-spreading. In lip-rounded, the tongue may be taking advantage of the extra rounding, and retracting less than normal for these sounds. This supports the notion that there may be compensation in the articulation of a segment to maintain its target acoustic characteristics in the face of different settings.

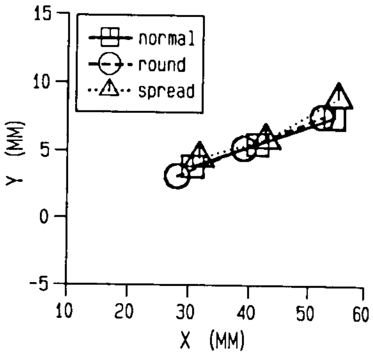

Fig 3. Mean position of EMA tongue-coils for [ɹ].

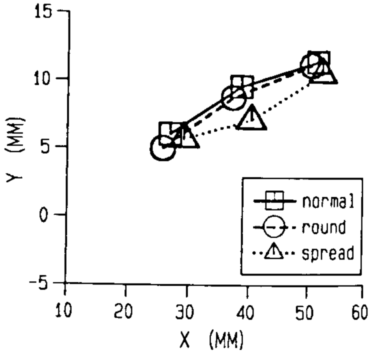

Fig 4. Mean position of EMA tongue-coils for [ʃ].

To test whether acoustic invariance is achieved, formant values were measured for [ɹ], and average spectra derived for [ʃ]. For [ɹ] (Table 2) it appears that F2 and F3 are relatively constant between *neutral* and *lip-rounded*; the articulatory retraction has not, however, fully compensated the effect of *lip-spreading*. The resultant spectra for [ʃ] in *neutral*, *lip-rounded*, and *lip-spread* are overlaid in Fig. 5. Although there are differences in the overall shape of the spectra, it is noticeable that the first major peak, which may contribute to the lower energy cutoff important in the perceptual distinction between [ʃ] and [s], is virtually constant at around 1600 Hz.

Table 2. Formant frequencies for [ɹ]

	F1	F2	F3
Neutr.	347	1358	1766
Rnd.	307	1357	1711
Spread	376	1711	1918

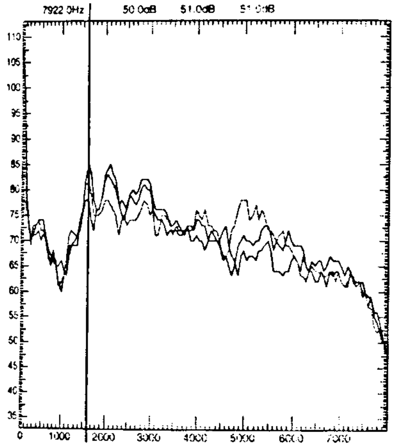

Fig 5. Overlaid average spectra for [ʃ] under three labial conditions.

4 DISCUSSION

This experiment is only a small beginning to the articulatory validation of the settings hypothesised to underlie perceptual components of voice quality. It has, however, shown the applicability of EMA as a techniques of analysis, and the results have borne out the assumptions made about *palatalisation* and raised questions about *velarisation*. Testing the mapping of terms in the framework onto physical dimensions is not a trivial exercise. Just as potentially the descriptive dimensions 'height' and 'frontness' for vowels may not closely correspond to articulatory fact for a set of vowels, so descriptive labels such as *palatalised* (voice) may not be accurate. Indeed it seems that the auditory quality labelled *velarised* may involve a rather more backward bias of the tongue body than the name implies.

The experiment also demonstrates the complex interaction between settings and segments: [ɹ] and [ʃ], whose criterial acoustic properties are sensitive to the pre-apical cavity, compensate for lip settings by fine adjustments in the coronal articulation. It is tempting to argue from this to a model of speech production which takes auditory goals as primary, and derives articulatory configurations by predictive modelling; but such an argument lies outside the scope of the present paper.

REFERENCES

[1] Abercrombie, D. (1967) *Elements of General Phonetics*. Edinburgh: EUP

[2] Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge: CUP.

[3] Nolan, F. (1983) *The Phonetic Bases of Speaker Recognition*. Cambridge: CUP.

[4] Esling, J.H. (1987) Vowel shift and long term average spectra in the Survey of Vancouver English. In *Proc. 11th ICPhS* Vol 4, 243-6. Tallinn: Academy of Sciences of the Estonian SSR.

[5] Hoole, P. (1993) Methodological considerations in the use of EMA in phonetic research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation* 31, 43-64.