

VOWEL GESTURES AND SPECTRA: FROM RAW DATA TO SIMULATION AND APPLICATIONS

Sidney A. J. Wood

Universities of Umeå and Lund, Sweden

ABSTRACT

This paper discusses the relationship between vowel gestures and the spectral output. Some sample model experiments on gestures analysed from x-ray films are presented, and examples are given of phonological applications (the tense vs lax contrast, Bulgarian vowel reduction, assimilations in Circassian).

1. GESTURE TO SPECTRUM

The single-cavity theory related timbre to the size of the front cavity. One interpretation (Roudet) claimed that compensations, such as trading lip rounding and tongue retraction, were unforeseeable and that articulation was therefore irrelevant [12]. On the other hand, Bell [1,73-74] postulated that defined settings of tongue height and retraction would yield predictable configurations and timbres. This theory was only discarded once phoneticians had accepted at least two formants. Paget [11, ch. 3-5] assigned them to the throat and mouth cavities respectively, and he also assumed there were other cavities for additional resonances.

Joos [6,57-59] pointed out that all resonances are modes of oscillation of the entire vocal tract and cannot be ascribed to minor side chambers. A gesture contributes to them all. Thus, there is no simple causal relationship between height and F1, or retraction and F2. The spectral consequences of a gesture depend on how it widens or narrows the vocal tract locally with respect to the standing wave of each resonance mode [3,4,5,9,15; see also 2,10]. Seen from this vantage point, Roudet's compensations and Bell's vowel configurations do not work [13,19]. Following [3,4,5,9,15], we can get a rough idea of the spectral consequences

of a gesture from how it shapes the vocal tract locally in relation to the nodes and antinodes of each resonance mode. My own calculations on configurations from several languages show that the nodes and antinodes never wander very far from the theoretical locations given in [3]. The relative sensitivity of each mode to the gesture can be judged from the local distributions of kinetic and potential energies. Published energy data based on Russian [4], French [8] and Arabic [16] are very similar for similar vowels, and thus mainly reflect universal acoustical properties of the vocal tract rather than language specific habits.

Published nomograms give the magnitude of a formant shift that can be attributed to a gesture, although the three-parameter models [4,14] are difficult to understand in gestural terms, as they are really models of the area function, and not of the manoeuvres that created it. Lindblom and Sundberg [7] have done a similar systematic mapping of the spectral consequences of manipulating a simulated vocal tract.

The alternative to reading nomograms is to simulate the gesture on a vocal tract profile and calculate the vocal tract resonances for successive configurations. This is the approach used for the model experiments reported here, that were designed to elucidate the spectral contributions of vowel gestures analysed from x-ray motion films of speech. The method is outlined in [17].

2. VOWEL GESTURES

The gestures were analysed from x-ray motion films by tracking articulator displacement between successive picture frames [17,22]. The actual gestures were

as follows. (1) Four tongue body manoeuvres (with respect to the mandible) directed towards the hard palate, velum, upper pharynx or lower pharynx [16], (2) mandibular depression, (3) lip rounding with respect to the jaws, (4) tongue blade elevation or depression with respect to the tongue body, and (5) larynx depression.

3. MODEL EXPERIMENTS

3.1 Palatal vowels

The analysed x-ray profiles revealed a set of [i-ɪ]-like vowels with closer jaw openings, typically 6-9 mm, and a more open set of [e-ɛ]-like vowels, with jaw

openings 9-14 mm [18,21]. They are further divided into *tense* [i-e]-like (with a bunched tongue posture) and *lax* [ɪ-ɛ]-like (with a flatter tongue). An example of a model experiment is given here in Fig. 1. The tongue was modified in 6 steps from a bunched to a flatter posture, at two jaw positions (8 and 14 mm). At 8 mm, the palatal passage ranged from 0.3 to 2.0 cm² cross section, and at 14 mm from 1 to 3.2 cm². This also reproduces the frequently observed phenomenon that the tongue is *higher* (the palatal passage is narrower) for [e] than [ɪ].

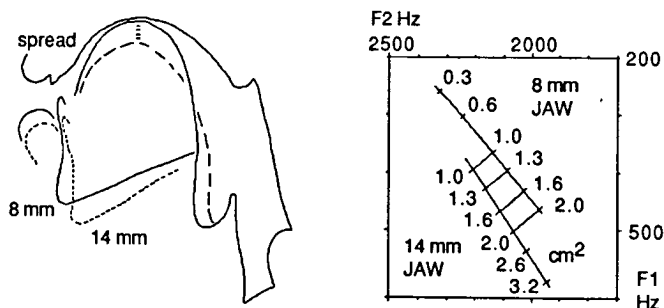


Figure 1. Left: the 6 degrees of palatal tongue bunching (relative to the mandible) at two typical jaw positions for [i-ɪ] and [e-ɛ]-like vowels; the profile illustrates the two extreme postures at the closer jaw position; the same six postures were then repeated at the more open position. Right: the resulting F1 and F2 frequencies at the 6 degrees of palatal constriction and two jaw positions.

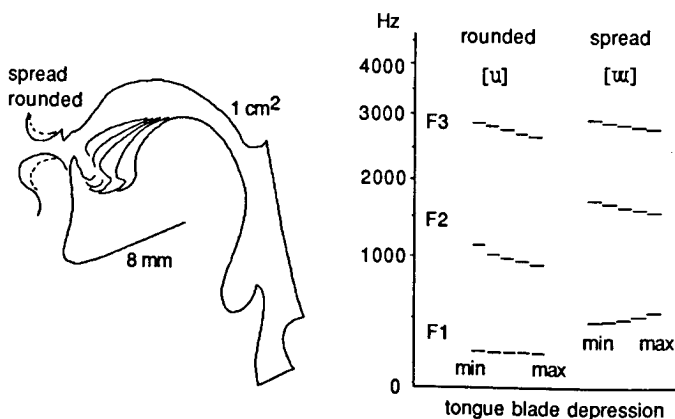


Figure 2. The effect of tongue blade position on palatovelar [u] and [ɯ]-like vowels (close jaw opening).

Other experiments [19] demonstrated the role of larynx depression in rounded palatal vowels and revealed some quantal relationships between component gestures and their spectral correlates. The experiments reported in [18] successfully reproduced the variation observed in natural speech, provided a precision of at least ± 1 mm was allowed for compensatory interplay between the tongue and mandible.

3.2 Palatovelar vowels

The [u-ʊ] and [ɯ]-like vowels also have a close jaw opening ranging 6-9 mm. *Lax* [ʊ] has a less bunched tongue body (a less constricted faucial passage and less ATR), while the lips are less rounded and

the larynx less depressed [21].

The example in Fig. 2 illustrates the effect of tongue blade depression on palatovelar vowels. X-ray pictures usually show how languages contrasting [u] and [y] have the tongue blade tightly depressed in [u], which increases the gravity of this vowel (lower F2). The less grave English [u] (higher F2) has traditionally been ascribed to an advanced tongue body whereas it is in fact due to tongue blade elevation.

Other experiments confirm that very efficient restoration of lip rounding is needed to compensate for mandibular variation in [u]. A similar experiment for [o] is described in the next section.

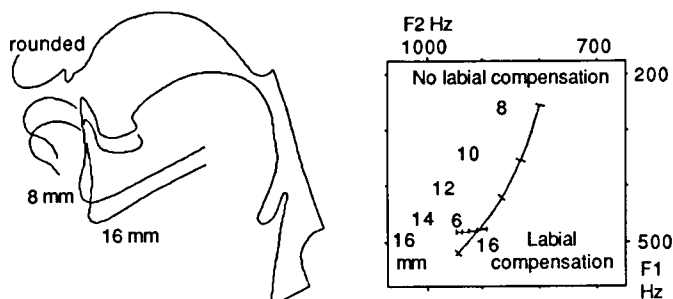


Figure 3. The effect of varying the jaw opening from 8 to 16 mm in an [o]-like vowel, both with labial compensation (the lip opening at 14 mm jaw opening was preserved, while the jaw was stepped from 16 to 6 mm), and without labial compensation.

3.3. Pharyngovelar vowels

Pharyngovelar (uvular) [o-ɔ] and [ɯ]-like vowels have an open jaw opening (9-14 mm). *Lax* [ɔ] has less lip rounding, less tongue blade depression, less ATR and less larynx depression. The example in Fig. 3 shows the success of perfect labial compensation for mandibular variation. Figure 3 also demonstrates that uncompensated mandibular variation in [o] can shift the spectrum towards an [u]-like timbre (i.e. the close, rounded palatovelar and close, rounded pharyngovelar configurations are spectrally ambiguous).

3.4. Low pharyngeal vowels

The low pharyngeal [ɑ-æ]-like vowels are usually produced with a jaw opening of

9-14 mm. The narrowed lower pharynx is widened from about 0.5 cm² for [ɑ], through 1.5 cm² for [a], to 2.5 cm² for [æ]. Grave [ɑ] is darkened to [ɔ] by adding lip rounding. The example at Fig. 4 illustrates the effect of varying the degree of low pharyngeal constriction, together with spread and rounded lips, all at one 14 mm jaw opening.

An interesting finding in [23] was that the "indeterminate" Bulgarian vowel /ã/ has an [a]-like low pharyngeal constriction but a close jaw position. This is a configuration that is precluded by the Bell model (where the mandible is explicitly disregarded and where vowels implicitly cannot be both *high* and *low*).

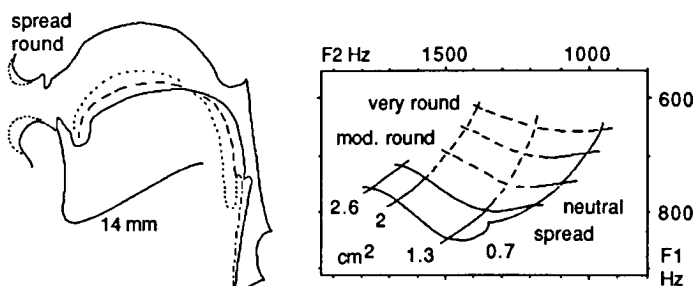


Figure 4. The effect of varying the degree of low pharyngeal constriction from 0.7 cm^2 to 2.6 cm^2 cross section, and lip rounding, at one mandible position (14 mm). The dashed areas are not generally relevant in speech.

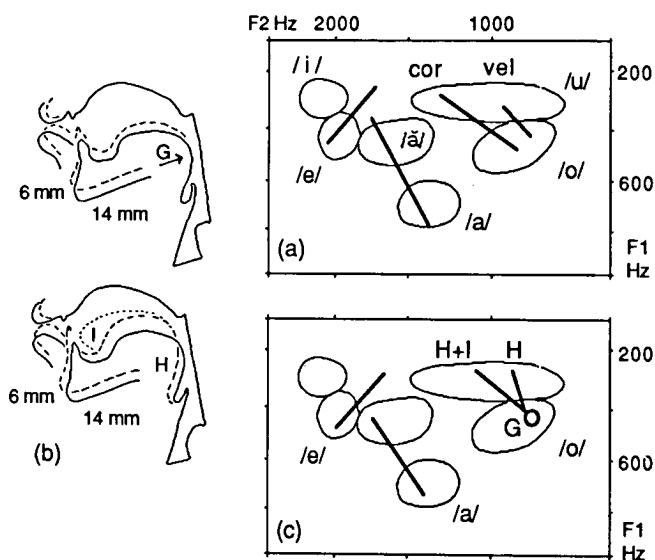


Figure 5. Simulation of vowel reduction in Bulgarian: (a) F1/F2 regressions for all /a,e,o/ (stressed and unstressed), also showing coronal and velar allophones of /o/, (b) simulated mandible variation 16-6 mm for /o/ (G perfect lingual and labial restoration, H no lingual or labial restoration, I tongue blade elevation), (c) simulated spectral reductions.

4. APPLICATIONS

The simulation of vowel gestures in model experiments is a useful tool for solving phonological problems. An example is vowel reduction in Bulgarian [23], where unstressed open /e,a,o/ merge with the reflexes of close /i,ä,u/ (Fig. 5a). The hypothesis was that this results from the

speaker omitting gestures that otherwise ensure the spectral contrasts. Fig. 5b illustrates the case of /o/. If lingual and labial compensation are turned off in unstressed syllables, and the jaw opening weakened, the spectrum becomes ambiguously similar to that of [u] without the tongue articulation having to be reorga-

nised to palatovelar (cf. Fig. 3). The simulations reproduced the observed regressions (Fig. 5c). A further point is that /o/ (stressed and unstressed) was brighter (higher F2) after coronal consonants. This was simulated by having the tongue blade elevated (I in Fig. 5c).

The spectral ambiguity arising from some configurations may also offer an explanation for some anomalous vowel assimilations in the Circassian language Kabardian [20]. The close phoneme is [u] or [u]-like in pharyngovelar contexts, i.e. they may well be pharyngovelar rather than palatovelar. This mirrors the reduction of Bulgarian /o/. Close pharyngovelar is one more configuration that is alien to the Bell model. Finally, there is a similar alternation in Kabardian between [o] and [a] in pharyngovelar environments that ostensibly seems to involve a tongue shift in only the one case. An alternative explanation is that rounded pharyngovelars and rounded low pharyngogals can be spectrally ambiguous. This is currently being investigated.

5. REFERENCES

- [1] BELL, A. M. (1867), *"Visible Speech"*, London: Simpkin, Marshall.
- [2] BOE, L. J. & PERRIER, P. (1990), "Comments on 'Distinctive regions' by M. Mryati et al", *Sp. Com.* 9,217-230.
- [3] CHIBA, T. & KAJIYAMA, M. (1941), *"The Vowel, its Nature and Structure"*, Tokyo: Phon. Soc. Jap.
- [4] FANT, C. G. M. (1960), *"The Acoustic Theory of Speech Production"*, The Hague: Mouton.
- [5] FANT, C. G. M. (1980), "The relation between area functions and the acoustic signal", *Phonetica* 37,55-86.
- [6] JOOS, M. (1948), *"Acoustic Phonetics"*, suppl. *Lang.* 24.
- [7] LINDBLOM, B. E. F. & SUNDBERG, J. (1971), "Acoustical consequences of lip, tongue, jaw and larynx movements", *J.A.S.A.* 50,1166-1179.
- [8] MRYATI, M. & CARRE, R. (1976), "Relations entre la forme du conduit vocal et les traits caractéristiques acoustiques des voyelles françaises", *Phonetica* 33,285-306.
- [9] MRAYATI, M., CARRE, R. & GUERIN, B. (1988), "Distinctive regions and modes; a new theory of speech production", *Sp. Com.* 7,257-286.
- [10] MRAYATI, M., CARRE, R. & GUERIN, B. (1990), "Distinctive regions and modes; a reply to Boë and Perrier", *Sp. Com.* 9, 231-238.
- [11] PAGET, R. (1930), *"Human Speech"*, London: Kegan Paul, Trench, Trübner.
- [12] ROUDET, L. (1911), "La classification des voyelles de M. Sweet", *Rev. de Phon.* 1,347-356.
- [13] STEVENS, K. N. (1972), "The quantal nature of speech: evidence from articulatory-acoustic data", in *Human Communication, a Unified View*, ed. by E. E. David & P. B. Denes, 51-66, New York: McGraw Hill.
- [14] STEVENS, K. N. & HOUSE, A. S. (1955), "Development of a quantitative description of vowel articulation", *J.A.S.A.* 27,484-495.
- [15] STEVENS, K. N. & HOUSE, A. S. (1961), "An acoustical theory of vowel production and some of its implications", *J. Sp. H. R.* 4,303-320.
- [16] WOOD, S. A. J. (1979), "A radiographic analysis of constriction locations for vowels", *J. Phon.* 7,25-43.
- [17] WOOD, S. A. J. (1982), "X-ray and model studies of vowel articulation", *Working Papers* 23,1-41, Dept. of Linguistics, Lund University.
- [18] WOOD, S. A. J. (1982), "Radiographic and model studies of the palatal vowels", *Working Papers* 23,119-155, Dept. of Linguistics, Lund University.
- [19] WOOD, S. A. J. (1986), "The acoustical significance of tongue, lip and larynx maneuvers in rounded palatal vowels", *J.A.S.A.* 80,391-401.
- [20] WOOD, S. A. J. (1990), "Vertical, monovocalic and other 'impossible' vowel systems; a review of the Kabardian vowels", *Working Papers* 36,191-212, Dept. of Linguistics, Lund University.
- [21] WOOD, S. A. J. (1991), "A radiographic and model study of the tense vs. lax contrast in vowels", in *Phonologica 1988*, ed. by W. Dressler et al, Cambridge University Press.
- [22] WOOD, S. A. J. (in press), "X-ray data on the temporal coordination of speech gestures", *J.Phon.*
- [23] WOOD, S. A. J. & PETERSSON, T. (1988), "Vowel Reduction in Bulgarian; the phonetic data and model experiments", *FoL Ling.* 22,239-262.