seems that the current approach can challenge Frid's (2003) results (96.9% on a much larger (70kw) training corpus), while still retaining the advantage of the more interpretable rule representation. Frid goes on to predict lexical prosody; we hope to get back to this topic.

4 Future directions

Outside incorporating more sophisticated compound splitting, there are several interesting directions. The template set is currently small. Likewise, the feature set for each corpus position may be extended in other ways, for instance by providing classes of graphemes – C and V is a good place to start, but place or manner of articulation for C and frontness for vowels might also be considered. Such classes might help finding generalizing rules over, say, front vowels or nasals, and might help where data is sparse; the extracted rules are also likely to be more linguistically relevant. If so, segments should preferably be chosen such that they fall clear into classes.

Another, orthogonal approach is "multidimensional" TBL (Florian & Ngai 2001), i.e., TBL with more than one variable. For instance, the establishment of stress pattern may determine phoneme transcription, or the other way round. For most TBL systems, rules can change one, prespecified attribute only (although many attributes may provide context). This is true for μ-TBL as well; however, we are currently considering an extension.

Interesting is also the idea to try to predict quantity and stress reductively, with Constraint Grammar-style reduction rules (i.e., "if Y, remove tag X from the set of possible tags"). Each syllable is assigned an initial set of all possible stress levels, a set which is reduced by positive rules ('ending -<ör># has main stress; thus its predecessor does not') as well as negative ('ending -<lig># never takes stress'). μ-TBL conveniently supports reduction rules.

References

Bouma, G., 2000. A finite state and data oriented method for grapheme to phoneme conversion. *Proceedings of the first conference on North American chapter of the Association for Computational Linguistic*, Seattle, WA.

Brill, E., 1992. A simple rule-based part of speech tagger. *Third Conference on Applied Natural Language Processing,* ACL.

Brodda, B., 1979. Något om de svenska ordens fonotax och morfotax: Iakttagelse med utgångspunkt från automatisk morfologisk analys. *PILUS 38*. Institutionen för lingvistik, Stockholms universitet.

Florian, R. & G. Ngai, 2001. Multidimensional Transformation-Based Learning. *Proceedings of the Fifth Workshop on Computational Language Learning (CoNLL-2001)*, Toulouse.

Frid, J., 2003. *Lexical and Acoustic Modelling of Swedish Prosody.* PhD Thesis. Travaux de l'institut de linguistique de Lund 45. Dept. of Linguistics, Lund University.

Hedelin, P., A. Jonsson & P. Lindblad, 1987. *Svenskt uttalslexikon* (3rd ed.). Technical report, Chalmers University of Technology.

Kondrak, G., 2000. A new algorithm for the alignment of phonetic sequences. *Proceedings of the first conference on North American chapter of the ACL*, Morgan Kaufmann Publishers Inc, 288-295.

Lager, T., 1999. The μ-TBL System: Logic Programming Tools for Transformation-Based Learning. *Third International Workshop on Computational Natural Language Learning (CoNLL-1999)*, Bergen.

Olsson, L-J., 1998. Specification of phonemic repesentation, Swedish. *DEL 4.1.3 of EC project "SCARRIE Scandinavian proof-reading tools" (LE3-4239).*

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics 145
Working Papers 52 (2006), 145–148

The Articulation of Uvular Consonants: Swedish

Sidney Wood

Dept. of Linguistics and Phonetics, Centre for Languages and Literature, Lund University
sidney.wood@ling.lu.se

Abstract

The articulation of uvular consonants is studied with particular reference to quantal aspects of speech production. Data from X-ray motion films are presented. Two speakers of Southern Swedish give examples of [ʀ]. The traditional view, that uvular consonants are produced by articulating the tongue dorsum towards the uvula, is questioned, and theoretical considerations point instead to the same upper pharyngeal place of articulation as for [o]-like vowels. The X-ray films disclose that these subjects did indeed constrict the upper pharynx for [ʀ].

1 Introduction

1.1 The theory of uvular articulations

This study begins by questioning the classical account of uvular consonant production (e.g. Jones, 1964), that the tongue dorsum is raised towards the uvula, and that the uvula vibrates for a rolled [ʀ]. Firstly, it is not clear how a vibrating uvula would produce the acoustic energy of a typical rolled [ʀ]. A likely process exploits a Bernoulli force in the constricted passage to chop the voiced sound into pulses when air pressure and tissue elasticity are suitably balanced, which requires that intermittent occlusion is possible between pulses. Unfortunately, there are free air passages either side of the uvula that should prevent this from happening. Secondly, these same passages should likewise prevent complete occlusion for a uvular stop, and they should also prevent a Reynolds number becoming sufficiently small for the turbulence of uvular fricatives.

If the uvula is not a good place for producing consonants known as "uvular", how else might they be produced? Wood (1974) observed that the spectra of vowel-to-consonant transitions immediately adjacent to uvular consonants were very similar to the spectra of [o ɔ]-like vowels, or to their respective counterparts [ɤ ʌ], and concluded that they shared the same place of location, i.e. the upper pharynx, confirmed for [o ɔ]-like vowels by Wood (1979). Mrayati et al. (1988) studied the spectral consequences of systematic deformations along an acoustic tube, and also concluded that the upper pharynx was a suitable location for these same consonants and vowels. Observations like this are obviously relevant for discussions of the quantal nature of speech (Stevens 1972, 1989). Clarifying the production of uvular consonants is not just a matter of correcting a possible misconception about a place of articulation. It concerns fundamental issues of phonetic theory.

1.2 This investigation

The uvular articulations were analysed from cinefluorographic films, a method that enables simultaneous articulatory activity to be observed in the entire vocal tract, and is therefore suitable for studying the tongue manoeuvres associated with uvular consonants. Two undisputed sources of uvular consonants are [ʀ] in southern Swedish, and [ʀ q χ] in West Greenlandic Inuit. The subjects of the films are native speakers of these languages. Examples from the Greenlandic subject have been published in e.g. Wood (1996a-b, 1997). Examples from one Swedish subject are reported in this paper. Examples from a second Swedish subject will be presented at the conference.

2 Procedures

Wood (1979) gives details of how the films were made. One reel of 35mm film was exposed per subject at an image rate of 75 frames/second (i.e. 1 frame every 13.3ms), allowing about 40 seconds per subject. Each frame received a 3ms exposure.

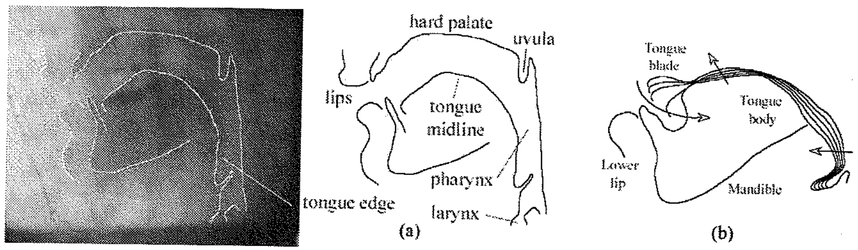

Figure 1. (a) Example of profile tracing and identification of prominent features. Note the difference between the tongue midline and the edge contours. (b) Examples of tongue body and tongue blade maneuvers (five successive film frames in this instance).

In the film by SweA, Swedish sibilants were commuted through different vowel environments. The uvular variant of Swedish /r/ occurs in the present indicative verb ending {/ar/} followed by the proposition {/i:/}, yielding several tokens of the sequence [aʀi].

In the film by SweB, the long vowels of Swedish (diphthongized in this dialect) were placed in a /bVd/ environment. The uvular variant of Swedish /r/ occurs where the subject recited the date and location of the film session. The word "four", *fyra* (/fy:ra/), is reported here, yielding the sequence [əyʀa].

3 Examples from the subject SweB

The frame by frame tongue body movement by SweB in [yʀa] is summarised in Figure 2. The sequence of profiles from [y] through [ʀ] to [a] is shown in Figures 3-5 (every other film frame, i.e. about 27ms between each illustration).

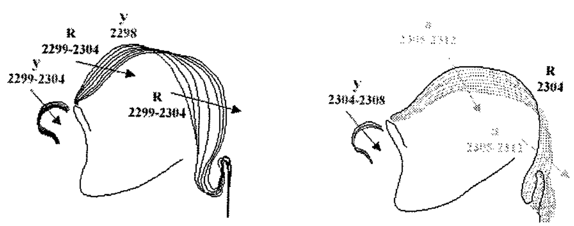

Figure 2. Subject SweB. Frame by frame tongue body movement (13.3ms for each step) in the transition from [y] to [ʀ] (left) and [ʀ] to [a] (right). The numbers refer to frames on the film.

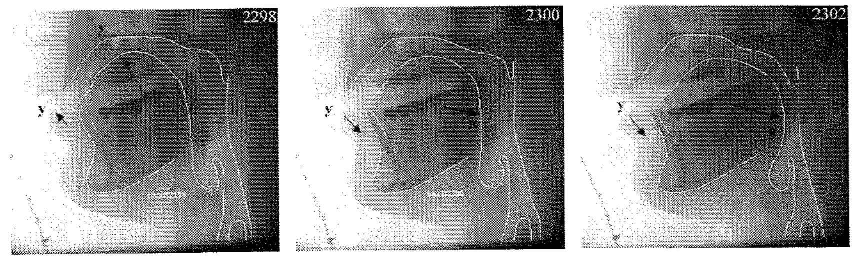

Figure 3. Every other profile from the sequence [yʀa], starting from the most complete /y:/ profile (left, frame 2298): tongue body raised towards the hard palate and lips rounded. The tongue body was then retracted for the transition to [ʀ], and the lip rounding withdrawn (2300 centre, 2302 right). Continued in Figure 4.

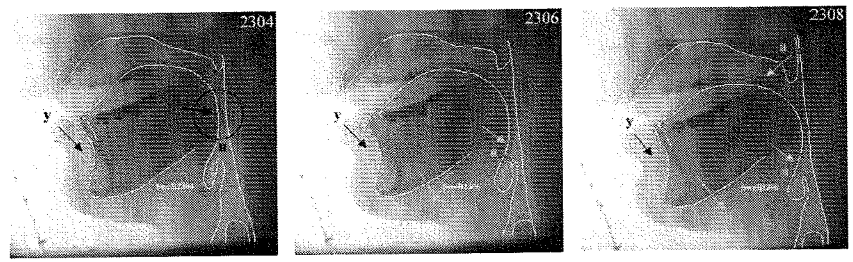

Figure 4. Profiles from the sequence [yʀa], continued from Figure 3. The transition to [ʀ] continued to frame 2304 (left), concluding with a narrow pharyngeal constriction (circled). This retraction was accompanied by slight depression, so that the tongue dorsum passed below the uvula and was directed into the upper pharynx. The lip rounding of /y:/ is still being withdrawn. Activity for /a/ was then commenced, continuing through frames 2306, centre, and 2308, right. The tongue body gesture of /a/ is directed towards the lower pharynx, accompanied by mandibular depression. The velar port opened slightly in frame 2308 (right) (this sequence is phrase final and was followed by a breathing pause). Continued in Figure 5.

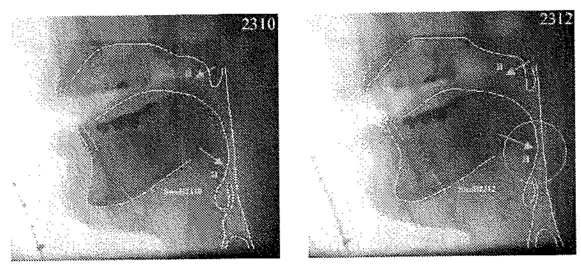

Figure 5. Profiles from the sequence [yʀa], continued from Figure 4. The transition from [ʀ] to [a] continued through frame 2310 (left) to frame 2312 (right), concluding with a narrow low pharyngeal constriction (circled), as expected from Wood (1979).

4 Discussion and conclusions

The retracting tongue body manoeuvre from [y] to [ʀ], seen in Figure 1 (left) and in profiles 2300 to 2304 in Figures 2 and 3, was depressed slightly. Consequently it passed below the uvula and continued into the pharynx. For this instance of [ʀ], the subject did not elevate the tongue dorsum towards the uvula. Similar behaviour was exhibited by the West Greenlandic subject for [ʀ q χ], and by the second Swedish subject whose results will be presented at the conference.

The target of the tongue body gesture of [ʀ] was the upper pharynx, as hypothesized. This was also the case in the other data to be reported at the conference. The upper pharynx is also the region that is constricted for [o] and [ɔ]-like vowels, which means that this one place of articulation is shared by all these consonants and vowels.

The upper pharynx is a more suitable place than the uvula for producing "uvular" stops, fricatives and trills. The soft smooth elastic surfaces of the posterior part of the tongue and the opposing posterior pharyngeal wall allow perfect occlusion, or the creation of apertures narrow enough for the generation of turbulence.

References

Jones, D., 1964. *An Outline of English Phonetics*. Cambridge: W. Heffer & Sons Ltd. (9th amended edition).

Mrayati, M., R. Carré & B. Guérin, 1988. Distinctive regions and modes: a new theory of speech production. *Speech Communication 7*, 257-286.

Stevens, K.N., 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In P.B. Denes & E.E. David, Jr. (eds.), *Human Communication: A Unified View*. New York: McGraw Hill, 243-255.

Stevens, K.N., 1989. On the quantal nature of speech. In J. Ohala (ed.), *On the Quantal Nature of Speech. Journal of Phonetics 17*, 3-45.

Wood, S.A.J., 1974. A spectrographic study of allophonic variation and vowel reduction in West Greenlandic Eskimo. *Working Papers 4*, Dept. of Linguistics and Phonetics, University of Lund, 58-94.

Wood, S.A.J., 1979. A radiographic analysis of constriction locations for vowels. *Journal of Phonetics 7*, 25-43.

Wood, S.A.J., 1996a. Temporal coordination of articulator gestures: an example from Greenlandic. *Journal of the Acoustical Society of America* (A). *Poster presented at 131st meeting of the Acoustical Society of America*, Indianapolis.

Wood, S.A.J., 1996b. The gestural organization of vowels: a cinefluorographic study of articulator gestures in Greenlandic. *Journal of the Acoustical Society of America 100*, 2689 (A). *Poster presented at the Third Joint Meeting of the Acoustical Societies of America and Japan*, Honolulu.

Wood, S.A.J., 1997. A cinefluorographic study of the temporal organization of articulator gestures: Examples from Greenlandic. In P. Perrier, R. Laboissière, C. Abry & S. Maeda (eds.), *Speech Production: Models and Data (Papers from the First ESCA Workshop on Speech Modeling and Fourth Speech Production Seminar, Grenoble 1996). Speech Communication* 22, 207-225.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 149–152

Acoustical Prerequisites for Visual Hearing

Niklas Öhrström and Hartmut Traunmüller
Department of Phonetics, Stockholm University
{niklas.ohrstrom|hartmut}@ling.su.se

Abstract
The McGurk effect shows in an obvious manner that visual information from a speaker's articulatory movements influences the auditory perception. The present study concerns the robustness of such speech specific audiovisual integration. What are the acoustical prerequisites for audiovisual integration to occur in speech perception? Auditory, visual and audiovisual syllables (phonated and whispered) were presented to 23 perceivers. In some of the stimuli, the auditory signal was exchanged for a schwa syllable, a dynamic source signal and a constant source signal. The results show that dynamic spectral information from a source signal suffice as auditory input for speech specific audiovisual integration to occur. The results also confirm that type (and absence) of lip rounding are strong visual cues.

1 Introduction

Visual contribution to speech comprehension was for a long time ignored by theorists and only accounted for when the auditory speech signal was degraded (Sumby & Pollack, 1954). However, McGurk and MacDonald (1976) showed that auditory speech perception could be altered by vision even when the auditory stimulus lacked ambiguity. They used [baba] and [gaga] syllables dubbed on visual stimuli with different consonants. A visual [gaga] dubbed on an auditory [baba] evoked the percept of /dada/. A visual [baba] dubbed on an auditory [gaga] was often perceived as /gabga/ or /bagba/. This demonstrated ordinary speech perception to be a bimodal process in which optic information about a speaker's articulatory movements is integrated into auditory perception. Traunmüller and Öhrström (in press) have demonstrated that this also holds for vowels. It has been shown experimentally that perception of features such as labiality and lip rounding is dominated by the visual signal. In addition, it is worth mentioning that crossmodal illusions are not necessarily restricted to speech perception: Shams et al. (2000) demonstrated that the visual perception of the numerosity of flashes can be altered by simultaneous auditory presentation of clicks.

Bimodal speech perception normally involves synchrony between the auditory and the visual information from a speaker's articulatory movements. Visual information can, therefore, be expected to have a substantial influence on auditory speech perception but visual hearing might require presence of a more or less authentic acoustic speech signal. This study aims at exploring acoustical prerequisites for visually influenced auditory perception to occur. How much information from the original acoustic signal can we remove and still evoke visual hearing? In this study the four long Swedish vowels /i/, /u/, /ɛ/ and /ɒ/ will be tested (appearing both phonated and whispered in a [b_d] frame). In the first condition the formant frequencies of the vowel will be changed (in this case a [ə] will be used). In the second condition the formant peaks will be flattened out, whereby an approximate source signal will be obtained. In the third condition, the formant peaks will be flattened out and the acoustic signal will be kept in a steady state. It can be expected that at least the visible type of lip rounding will have an influence on auditory perception.