Proceedings of the 4th International Speech Motor Conference. Nijmegen: Uitgeverij Vantilt, 2001. p. 70-73.

Vocal intensity: acoustic and articulatory correlates

Anja GEUMANN
Institut für Phonetik und Sprachliche Kommunikation,
Ludwig-Maximilians-Universität Munich, Schellingstr. 3/II. VG, 80799 Munich, Germany
geumann@phonetik.uni-muenchen.de

Abstract. Analyses of jaw movement (obtained by Electromagnetic Articulography) and acoustics show that loud speech is an intricate phenomenon. Besides involving higher intensity and subglottal pressure it affects jaw movements as well as fundamental frequency and especially first formants. It is argued that all these effects serve the purpose of enhancing perceptual salience.

1. Introduction

In earlier work vocal intensity was primarily investigated in its relation to sound pressure or aerodynamic conditions below or around the glottis (e.g. Ladefoged/McKinney 1963, Holmberg et al. 1988). More recent experiments also indicate that changes in vocal intensity are coupled with changes in articulatory dynamics within the vocal tract (Schulman 1989, Dromey et al. 1995, Geumann et al. 1999).

Ladefoged/McKinney (p. 460) described the way the listener identifies loudness in the following way: "It is of course probably true, that what the listeners were doing when they were judging the loudnesses [...] was assessing the amount of effort they themselves would have to make in order to produce corresponding sounds."

Here we would like to present articulatory and acoustic data that show that speakers indeed modify their articulatory dynamics according to intended volume level. Thus, the assumption that loudness for the speaker is directly related to physiological effort will be further supported.

2. Experiment

Kinematic and acoustic recordings were made of read phrases, produced by 6 German speakers (one female, five male). Pseudo-word 'VCV sequences were embedded in carrier phrases of the type "Hab das Verb ___ mit dem Verb ___ verwechselt". Data for some of the speakers were presented in Geumann et al. (1999).

The consonants were the alveolar German phonemes differing in manner of articulation /s, ʃ, l, n, d, t/ (/ʃ/ is postalveolar). They were placed in differing symmetric vowel height contexts /i__i, e__e, a__a/; both vowels were long, with main stress on the first vowel. All phrases were produced in loud and normal speech, which was elicited by simple instruction of the speaker. The loud and normal phrases were presented in random order. For each target consonant with given loudness and vowel context 12 repetitions were produced, totalling 72 repetitions of each consonant over all context and loudness conditions.

The two-dimensional (midsagittal) kinematic signals were recorded with an electromagnetic transduction system (Articulograph AG100, Carstens Medizinelektronik, for more technical details see Hoole 1996). Four sensors were placed on the tongue, three sensors were used to track jaw movement. One was placed on each of inner and outer surface of the gums beneath the lower incisors, a third sensor was placed on the angle of the chin. Reference sensors were located on upper jaw and the nasion. Jaw movement was corrected for head movement and postprocessed data were rotated, so that the x-axis is parallel to the occlusal plane, with the origin at the position of the reference coil above the upper incisors. Only data of the outer jaw sensor will be considered below.

Simultaneously, acoustic data were recorded with a Sennheiser MK H20 P48 microphone (preamplifier higain 44) approx. 30cm distant to the speaker’s mouth. Data were stored on a Sony PC 208Ax DAT-recorder, sampling rate of the data was 24kHz/16bit. Additionally, speakers were recorded frontally on video.

Further analysis of the articulatory data was performed with Matlab (using macros written by Phil Hoole). Manually selected minimal jaw position within the acoustic vowel (V1) segment was considered as vowel target position, absolute maximal jaw position following the minimal peak was identified as consonantal target position, following absolute minimal jaw position was considered as V2 target position. Analysis of the acoustic data was performed using the praat program package developed by Paul Boersma and David Weenink. Fundamental frequency and formants were extracted automatically, with some post editing.

3. Results

Data in Figure 1 indicate that in loud voice condition all speakers effectively produced vowels (as well as consonants) with higher intensity. However, it has been noted by Eriksson/Traunmüller (1999) and Wilkens/Bartel (1977) that listeners can estimate or reconstruct the speaker's loudness independently of the perceived sound pressure level.

Data in Figure 2 and 3 demonstrate that all speakers show a more open jaw position in loud voice condition in vowels, whereas, during coronal consonant production loud voice condition has no effect on jaw position in most speakers. It is discussed elsewhere (e.g. Shadle 1985, 1990, Lee et al. 1994, Narayanan 1995 Geumann et al. 1999, Geumann in prep.) that the jaw, carrying the lower incisors, plays an important role in the production of strident fricatives /s/ and /ʃ/. Either way jaw movements are found to be larger in loud speech.

The acoustic data exhibit a modified durational pattern with loud voice, with vowels lengthened and consonants shortened, which was stated for French data as well (Bonnot/Chevrie-Muller 1991). Besides, a rather strong effect on fundamental frequency and first formant has been found (see as well Traunmüller/Eriksson 2000). In loud voice condition, fundamental frequency is considerably higher (Figure 4) as is the first formant (Figure 5). Effects on second formant value (Figure 6) were noted mainly for /a:/, which is a non-front vowel in German. The third formant values (Figure 7) are in loud voice condition centralized towards a focus around 2600Hz.

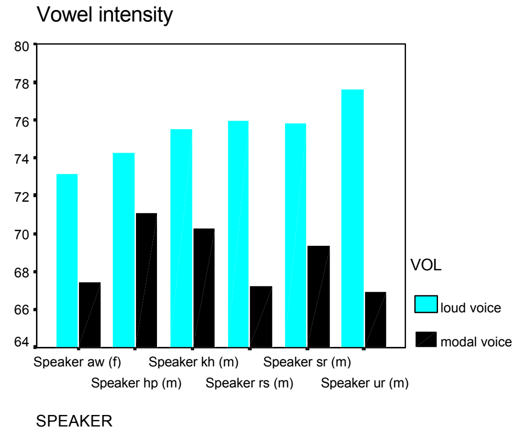

Figure 1. Intensity (in dB) of vowel segments (V1 and V2). T-test was performed and revealed a highly significant effect of volume level for all speakers. n = max. 432.

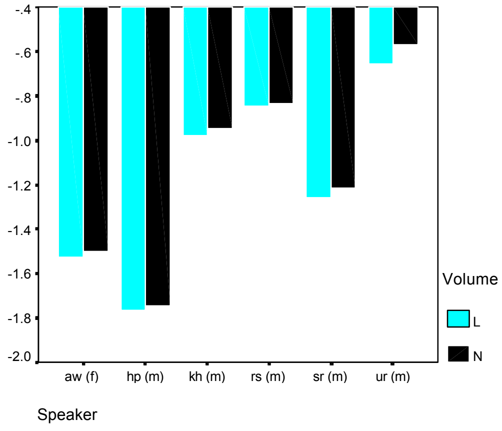

Figure 2. Jaw height (in cm) at consonantal target position for loud vs. modal voice. An independent samples t-test was performed and showed for speaker ur (p<0.001) and kh (p<0.01) a highly significant effect of volume level, for other four speakers no significant effect. n = max. 216.

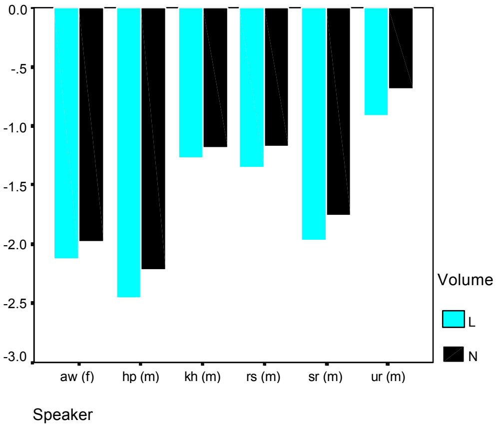

Figure 3. Jaw height (in cm) at stressed vowel (V1) target position for loud vs. modal voice. An independent samples t-test was performed and showed a highly significant effect (p<0.001) of volume level for all speakers. n = max. 216.

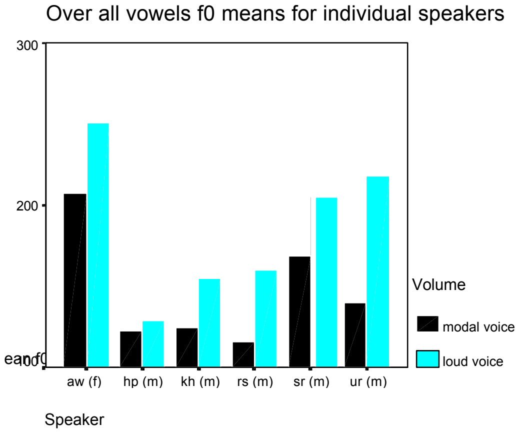

Figure 4. Effect of volume level on mean fundamental frequency (in Hz) in long vowels for individual speakers. Independent t-test comparison of means showed that effect of volume level is highly significant (p<0.001) for all speakers. n = max. 432.

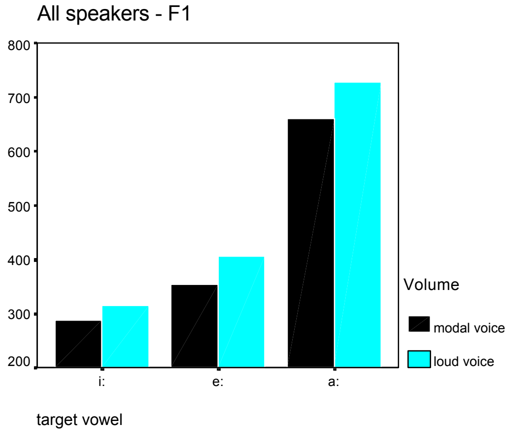

Figure 5. Effect of volume level on first formant. Data in Hz pooled for all speakers and for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F1 ($p<0.001$) for all vowels. n = max. 880.

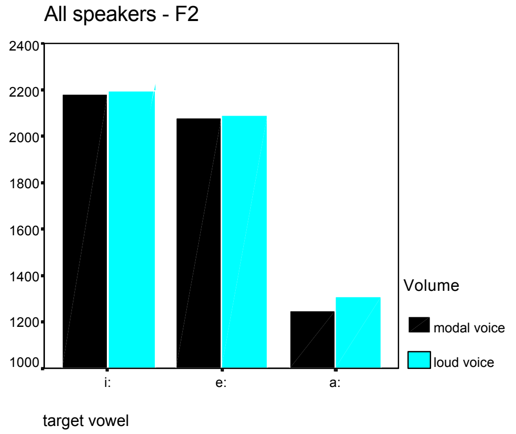

Figure 6. Effect of volume level on second formant. Data in Hz pooled for all speakers and for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F2 ($p<0.001$) for /a:/, significant effect ($p<0.05$) on /i:/, no significant effect on /e:/. n = max. 880.

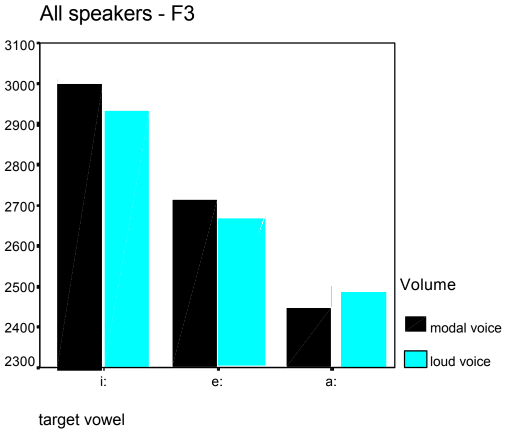

Figure 7. Effect of volume level on third formant. Data in Hz pooled for all speakers and for V1 and V2. An independent samples t-test was performed and showed a highly significant effect of volume level on F3. For /a:/ and /i:/ $p<0.001$, for /e:/ $p<0.01$. n = max. 880.

4. Conclusion

Articulatory and acoustic modification found with loud voice can be interpreted as an amplification of jaw movement, which might additionally help to shift acoustic durational patterns towards longer (more intense) vowels and shorter (in general less intense) consonants. Changes in fundamental frequency and first formant are easily explained by perceptive demands. A higher fundamental frequency is perceived more intense, and perception of vowel height refers to the difference between first formant and fundamental frequency (Syrdal/Gopal 1986). Raising of fundamental frequency therefore demands a higher first formant if vowel quality is to be preserved. On the other hand a higher first formant can be interpreted as resulting from a lower jaw position.

In addition, it was found that overall jaw activity can be quite different across speakers (see as well Hertrich/Ackermann 2000). Comparison of amplitudes of jaw and intrinsic tongue tip movements in consonant production revealed generally a higher jaw activity for speaker sr. Conversely, speaker ur only showed a comparatively small amount of overall jaw activity. In short, patterns of jaw activity can differ considerably across speakers.

To sum up, the phenomenon of loud speech can be said to be constituted by a varied collection of interdependent articulatory and acoustic parameters.

Acknowledgements

This work was partially supported by the German Research Council (DFG) (Ti69/31).

References

Bonnot, J.-F. P., Chevrie-Muller, C. (1991) Some effects of shouted and whispered conditions on temporal organization. *J. Phonetics 19, 473-483.*

Dromey, C., Ramig L.O., Johnson, A.B. (1995) Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: a case study. *Journal of Speech and Hearing Research*, 38, 751-764.

Eriksson, A., Traunmüller, H. (1999) Perception of vocal effort and speaker distance on the basis of vowel utterances. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 3, 2469-2472.

Geumann, A. (in prep.) Invariance and variability in articulation and acoustics of natural perturbed speech. Ph. Diss. University of Munich.

Geumann, A., Kroos, C., Tillmann, H.G. (1999) Are there compensatory effects in natural speech*? Proc. 14th International Congress of Phonetic Sciences* 1999. San Francisco, August, 399-402.

Hertrich, I., Ackermann, H. (2000) Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions. *Journal of the Acoustical Society of America* 107(4), 2236-2247.

Holmberg, E.B., Hillman, R.E., Perkell, J.S. (1988) Glottal airflow and transglottal air pressure measurements for male and female seakers in soft, normal, and loud voice. *Journal of the Acoustical Society of America* 84(2), 511-529.

Hoole, P. (1996) Issues in the acquisition, processing, reduction and parameterization of articulographic data, *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München (FIPKM)* 34, 158-173.

Ladefoged, P. (1990) On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein. In: J. Kingston & M. Beckman (Eds.) *Papers in Laboratory Phonology 1*, Cambridge: Cambridge University Press, 398-405.

Ladefoged, P., McKinney, N.P. (1963) Loudness, sound pressure, and subglottal pressure in speech. *Journal of the Acoustical Society of America* 35, 454-460.

Lee, S., Beckman, M., Jackson, M. (1994) Jaw targets for strident fricatives. *Proc. of the International Conference on Spoken Language Processing (ICSLP)* 1994 Yokohama, 37-40.

Narayanan, S.S. (1995) *Fricative consonants: An articulatory, acoustic, and systems study*. PhDiss. UCLA.

Praat [On-line]: Information at: http://www.fon.hum.uva.nl/praat/

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America* 85, 295-312.

Shadle, C.H. (1985) The acoustics of fricative consonants. Ph.D. thesis, MIT.

Shadle, C.H. (1990) Articulatory-acoustic relationships in fricative consonants. In: W. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 187-209.

Syrdal, A.K., Gopal, H.S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79(4), 1086-1100.

Traunmüller, H., Eriksson, A. (2000) Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America* 107(6), 3438-3451.

Wilkens, H., Bartel, H.-H. (1977) Wiedererkennbarkeit der Originallautstärke eines Sprechers bei elektroakustischer Wiedergabe. *Acustica* 37, 45-49.