

Laryngeal adjustment in voiceless consonant production: I. An experimental study of glottal abduction in loud versus normal speech

Susanne Fuchs¹, Phil Hoole², Xavier Pelorson³, Annemie Van Hirtum³, Pascal Perrier³, Klaus Dahlmeier⁴, Johanna Creutzburg¹

1 ZAS – Centre for General Linguistics, Berlin

2 IPSK – Institut for Phonetics and Speech Communication, Munich

3 ICP & CNRS – Institut de la Communication Parlée, Grenoble

4 Otorhinolaryngologist, Berlin

fuchs@zas.gwz-berlin.de

Abstract

Glottal abduction amplitude and duration during normal and loud speech were investigated by means of simultaneous recordings of transillumination, fiberoptic films and acoustics. Three German speakers were recorded producing several voiceless consonants and consonants clusters in word initial and word medial position. The aim of our study was to extend previous work on Danish /p/ (Andersen 1981) and test his model. Andersen proposed that loudness variations would coincide with a larger glottal opening amplitude for louder speech, but with no differences in the overall glottal opening duration. Our results concerning the production of voiceless consonants in word initial position are generally in agreement with Andersen's proposal. However, results for glottal opening in word medial position did not show similar differences. The amplitude of glottal opening in this position is considerably reduced for both normal and loud speech so that differences often diminish. Another intriguing question arises, although limited to our results in word initial position: Are the observed differences in glottal aperture controlled by the neural nervous system or are they due to physical factors (increased subglottal pressure in loud speech) or both? Further modelling work on this issue is presented in Van Hirtum (submitted to this conference).

1. Introduction

Variations of prosody such as speech rate, stress or loudness have been of particular interest in order to study the variability of articulatory kinematics or speech sounds. Changes of these modes may provide useful insights into the underlying control strategies or can at least allow certain hypothesis to be excluded.

Investigating supralaryngeal kinematics, Schulman (1989) found evidence that loud (shouted) speech can be described as a natural perturbation, where lip and jaw movements are amplified while the commonly produced distinctions are maintained. His findings for shouted speech showed a shorter acoustic consonant duration and a longer vowel duration in comparison to normal speech. He proposed that "loud is more" p.310. A comparable principle has been reported by de Jong (1995) varying stress patterns. Stress was associated with 'localized hyperarticulation' yielding more extreme (peripheral) articulatory positions and longer durations compared to normal speech, particularly in vowel production. McClean and Tasko (2003) correlated surface action potentials with simultaneously recorded kinematics (speed, distance and

duration) of lip and jaw movements. Speech rate and loudness in one test utterance were varied and compared to each speakers habitual rate and loudness. In the loud speech condition (approximately + 6 dB relative to normal speech) an increase of jaw and lip surface potentials correlated consistently with an increase of movement amplitude and velocity. Similarly consistent correlations between neural activity and kinematics could not be found for variations in speech rate, suggesting a different control for loudness and speaking rate.

Considering laryngeal movements and changes in intensity and rate, i.e. glottal opening in voiceless obstruents the only investigation we are aware of is Andersen (1981). Based on his experimental results, Andersen proposed a model for glottal abduction with 2 separate mechanisms: In louder speech the glottal opening amplitude increases with constant overall glottal opening duration (see figure 1 for a schematic view) whereas in fast speech both, amplitude and duration decrease. Andersen provided additional evidence for 1 subject that louder speech coincided with a higher posterior cricoarytenoid (PCA) activity whereas the interarytenoid (INT) did not change. (PCA activity and INT suppression are usually typical antagonistic patterns in single voiceless consonant production, see Hirose 1975). We will further focus on the loudness condition.

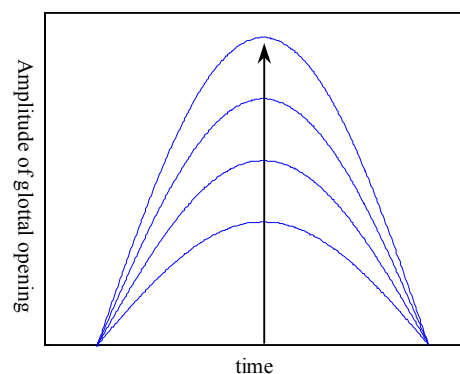


Figure 1: Schematic view of Andersen's model for loudness: with increasing intensity (marked with an array) amplitude of glottal opening becomes larger whereas the duration of glottal opening stays the same

If Andersen's model for loudness is correct, it could also be supposed that the larger glottal abduction amplitude in loud

speech is not only due to PCA activity, but also a result of an increased subglottal pressure. In order to verify this hypothesis, the aims of our study are:

- to carry out an experimental study and extend Andersen's work on /p/ to a series of voiceless consonants and consonant clusters in different positions in order to test his model, and
- to model loudness variations due to increasing subglottal pressure by means of numerical simulations (see Van Hirtum et al. submitted to this conference).

2. Method

2.1. Experimental set-up

By means of simultaneous recordings of transillumination, fiberoptic films, and acoustics we observed glottal opening amplitudes and overall glottal opening duration. Additional EPG data were obtained, but they are not in the scope of the current study. Three German native speakers were recorded, 2 males (CG, RW) and 1 female (SF). A standard endoscope of the type Olympus ENF (type P3) was inserted in the subject's pharynx and a photosensor was glued externally on the subject's neck below the cricoid cartilage. The endoscope was attached to a camera and connected to a video recorder with a monitor. The video images enabled the Otorhinolaryngologist to control the position of the tip of the endoscope and to identify whether saliva production influenced the signal. The video signals were taped to enable qualitative interpretation of the transillumination data.

Acoustic and transillumination data were sampled at 24 kHz, transillumination data were then downsampled to 200 Hz. The fiberoptic films have the standard video format of 25 images per second. The velocity of the transillumination signal was calculated as the first derivative. By analysing the velocity signal we defined the beginning and end of glottal opening and closing using a 5% threshold criterion in order to calculate overall glottal opening duration as the difference between the two landmarks. Fiberoptic stills have been taken from the images next to the acoustic burst for stops and around the middle of frication for fricatives. These images provide additional qualitative information of the transillumination data and are of further help for interpretations, since one of the disadvantages of the technique is that the distance between the tip of the fiberscope and the glottis is not known and can change during the whole recording (Hoole 1999). However, we will compare relative differences of glottal opening for normal and loud speech, but will not consider absolute values. In addition, small variations in glottal opening amplitude won't be taken as reliable differences.

2.2. Speech material

The speech material consisted of mainly bisyllabic real words with single consonants /p,t,f,ʃ,h/ and consonant clusters /pf,ts,tʃ,st,ʃp/ (see table 1). In some cases nonsense words were constructed which could be potential real words of the German lexicon. Two of the target words (X and Y) were put in the frame sentence: *Lese X wie Y bitte (Read X like Y please)*. Each sentence was produced in 3 different conditions in randomised but successive order: normal, loud and whispered (whispering is not taken into account in the current study). The level of intensity was self regulated by the subjects so that on the one hand a difference between normal and loud

could be perceived by the experimenter and on the other hand the placement of the endoscope was not disturbed by any additional head movements (to shout was not appropriate). Each sentence in each condition was repeated 5 times.

The consonants occurred either in word initial position (= stressed syllable) or word medial position (either in the post-stressed syllable when preceded by a tense vowel or ambisyllabic in the stressed/post-stressed syllable when preceded by a lax vowel). They were surrounded by the front vowels /i:,ɪ,y:,ʏ,ə/ in order to guarantee reliable transillumination data. Since the number of target words found for the relevant consonant/cluster and word position differed, the number of tokens varied. The surrounding vowel context did not influence the general results so that data were combined in order to increase the number of tokens and hence, strengthen further statistical analysis.

Table 1: *Speech material used in the corpus with relevant consonant/cluster, position in the word, graphematic representation of target words in different vowel contexts and number of tokens when target words for the relevant condition were repeated 5 times.*

Consonant/Cluster	Word position	/i:,ɪ/-context	/y:,ʏ/-context	Number of tokens
/p/	initial	piepe	Püterich	10
/p/	medial	piepe	Type Zyperm	15
/t/	initial	Tide, Tiefe, Tische	Type, Tüte	25
/t/	medial	biete	Tüte	10
/f/	initial	viele	fühle	10
/f/	medial	Tiefe	Hüfe	10
/ʃ/	initial	schiebe	Schübe	10
/ʃ/	medial	Tische	wüsche	10
/h/	initial	Hitsche	Hütsche, Hüfe, hüpe	20
/pf/	initial	Pfister	Pfütze	10
/pf/	medial	Zipfe	Hüpe	10
/ts/	initial	Ziele	Zyperm	10
/ts/	medial	Spitze	Pfütze	10
/tʃ/	initial	Chile	Tschüss	10
/tʃ/	medial	Hitsche	Hütsche	10
/st/	medial	Pfister	düste, büste	15
/ʃp/	initial	Spitze		5

3. Results

It should be mentioned that all consonant clusters were realised with only one glottal abduction peak.

3.1. Occurrence of glottal abduction gestures

In a previous study it has been found that /t/ in word medial post-stressed position is often considerably reduced or even disappears (Fuchs 2003). Table 2 lists the produced relative percentages of occurrence of glottal opening only for those cases where glottal opening did not occur in 100 percent of the cases.

Table 2: Relative percentages <100% of occurrence of glottal opening.

Consonant/ Cluster	Subject	Medial position normal vs. loud	
/p/	CG	66.6%	66.6%
	RW		86.6%
	SF		86.6%
/t/	RW	50%	50%

As can be seen for all the subjects only single stops in word medial position can disappear. Consonant clusters, single fricatives, and all tokens in word initial position were realised in all cases. A small trend concerning differences of the occurrence of glottal opening and intensity was found with louder speech showing a greater likelihood for the disappearance of glottal abduction in the word medial position.

3.2. Amplitude of glottal opening

The amplitude of glottal opening was calculated as the difference between the maximum of glottal opening and the baseline of the transillumination signal.

3.2.1. Word initial position

Since in word initial position louder speech seems to have a larger amount of glottal opening in comparison to normal speech a normalisation procedure was carried out. The amplitude of glottal opening for loud speech (GO_l) was set to 100 % and glottal opening for normal speech (GO_{nl}) was calculated in relation to this:

$$GO_{nl} = GO_n * 100 / GO_l \quad (1)$$

Analysis of variance using the General Linear Model in SPSS 11.51 with loudness as the independent factor and glottal opening amplitude as the dependent variable, split by consonant and subject, provided evidence that glottal opening for phonologically voiceless consonants in word initial position differs significantly between loud and normal speech. As can be seen in Figure 2, glottal opening is approximately 25-50% smaller in normal speech compared to loud speech. There were a few exceptions for subjects RW and SF: for both speakers /h/ could not be distinguished, additionally, /pf/ and /ʃ/ for RW and /f, sp/ for SF. The differences for /h/ can be explained with respect to vocal fold vibrations. /h/ is realised with vocal fold vibrations during the whole glottal opening period (see Sawashima et al. 1975), which might be explained by the missing intraoral air pressure increase, caused by an oral constriction or closure in obstruent production.

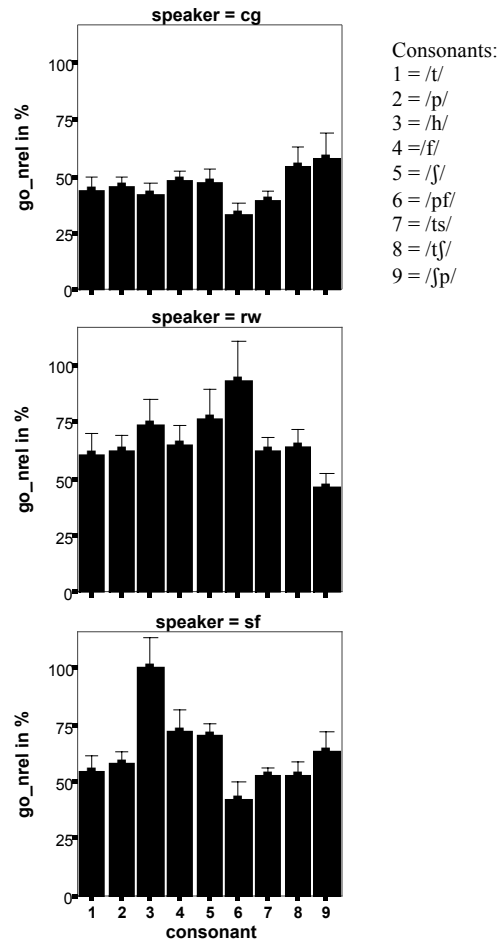


Figure 2: Averaged amount of glottal opening from normal speech in relation to loud speech (set to 100%), error bars correspond to +/-1std. error; subject:CG,RW,SF (from top to bottom);x-axis: different consonants

3.2.2. Qualitative evidence from fiberoptic stills for the word initial position

Qualitative images from fiberoptic stills support the findings from the transillumination signal. Figure 3 and 4 exemplify the differences for glottal opening amplitude, particularly in the posterior region between the arytenoid cartilages.

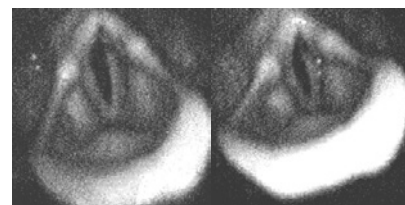


Figure 3: Fiberoptic stills from the image next to the burst in initial /p/ in piepe from CG; left loud, right normal

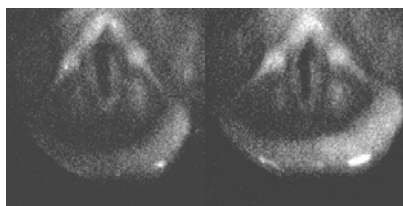


Figure 4: see figure 3, but for initial /t/ in Tüte

3.2.3. Word medial position

Glottal opening amplitudes in word medial position do not show similar differences regarding intensity as the ones in word initial position. Statistical analysis yielded no reliable distinction for all the consonants for speaker RW. For SF similar results are found with some exceptions in the clusters /pf, tʃ/. Subject CG's findings exhibit significant differences between loud and normal speech for all the consonant clusters, including /ʃ/. These results can be interpreted with respect to 'strong vs. weak' syllable positions (see e.g. Krakow) or to stressed vs. unstressed syllables (e.g. de Jong 1995, Löfqvist 1980). The word medial position in our corpus belongs most of the time to an unstressed syllable. Those syllables/positions are known to be realised with less extreme articulatory positions, shorter durations and much more variability. They are generally preferred places for reduction, coarticulation and assimilation processes. Our data on glottal opening in word medial position support this trend. However, consonant clusters seem to be less affected than single consonants.

3.2.4. Qualitative evidence from fiberoptic stills for the word medial position

The fiberoptic stills for the word medial position reveal the small amount of glottal opening (see figure 5). Based on these images it is rather difficult to decide whether loud speech is produced with a wider open glottis in comparison to normal. In the example in figure 5 the opposite could be assumed, but images vary from one to the other. We won't take these differences into account since first, transillumination cannot be calibrated and, second the framerate of standard video is too low for reliable statements, especially in word medial position where consonants are often shorter in comparison to the word initial position

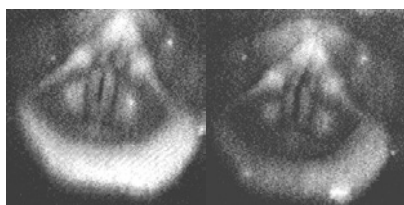


Figure 5: Fiberoptic stills from the image next to the burst in medial /p/ in piepe from CG; left loud, right normal

3.3. Overall glottal opening duration

In order to test Andersen's model for the loudness condition we obtained beside glottal opening amplitudes the overall glottal opening duration (GO_dur). Figure 6 shows averaged mean values of GO_dur for word initial position.

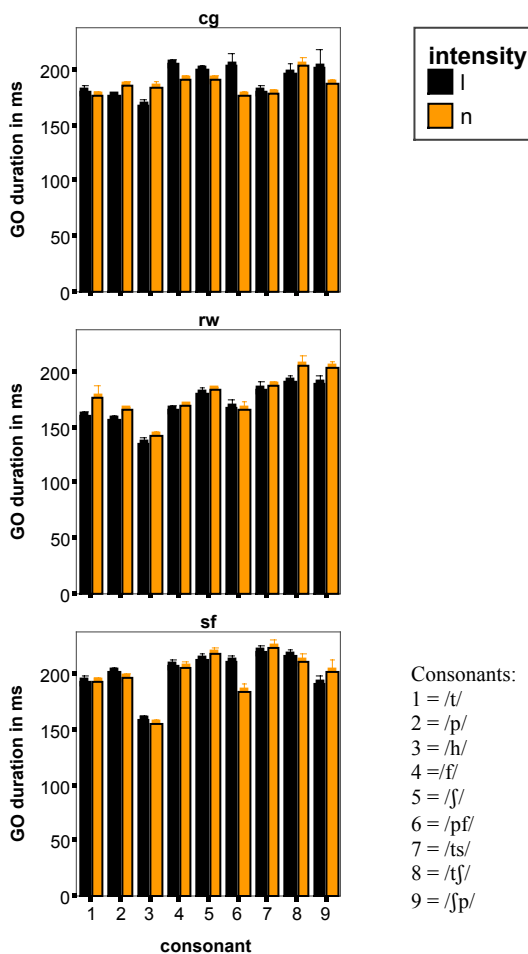


Figure 6: Averaged glottal opening duration in ms with error bars corresponding to +/-1std.error; subject:CG,RW,SF (from top to bottom) ;x-axis: different consonant

In general, our experimental data are in agreement with Andersen, i.e. although glottal opening amplitude is larger in louder speech, the duration of glottal opening stays the same. Most of the durations do not differ significantly. There are some weakly significant differences for CG in /h/, /f/ and /pf/, but they do not show consistent differences with respect to intensity: /h/ in normal speech is longer and /f,pf/ in louder speech. For RW none of the results differ significantly and for SF /pf/ exhibits a significantly longer glottal opening duration in loud speech, all others do not exhibit such patterns.

Averaged glottal opening duration for the word medial position show comparable findings with some exceptions for CG in /ʃ, tʃ/, RW in /st/ and SF in /f,p/ with consonants in louder speech being always shorter than in normal speech.

4. Discussion and Conclusion

We investigated glottal opening amplitude and duration in normal and loud speech by means of transillumination, fiberoptic films and acoustics in order to extend previous work on Danish /p/ from Andersen (1981) and show the validity of his model. Our results provide evidence that louder speech coincides with a larger glottal opening amplitude but overall glottal opening duration does not change with respect

to normal speech. Thus, Andersen's model can describe the behaviour of glottal abduction regarding changes in intensity. However, Andersen also mentioned that patterns may change when increasing loudness to a shouting character (approximately >10dB compared to normal). Based on our findings we additionally propose that Andersen's model holds only for the strong syllable positions, i.e. to the word initial position which was stressed. Distinctions in this position are more pronounced in comparison to weaker syllable positions (in our study the word medial position).

Amplitudes of glottal opening in word medial position are generally very small and shorter in duration, but amplitudes are more affected by a reduction process than timing. For the medial position differences between loud and normal speech are rather speaker and consonant dependently, i.e. some speakers produce some consonants with a larger glottal abduction in loud speech and others do not. On the one hand, such a result may be influenced by the fact that we did not control speakers' intensity change, i.e. one speaker could have produced a greater distinction in loudness than another. On the other hand, results could also be explained with respect to intensity variations in unstressed/medial position. Intensity is generally reduced in unstressed word medial position so that differences between loud and normal speech can diminish. In a next step we will investigate the dB differences.

Our results for the word initial position are of further relevance in discussing the control of loudness variation. Based on Andersen's EMG results it can be concluded that spatial variation in glottal opening is caused by muscular activity. However, Löfqvist, Baer and Yoshioka (1981) investigated to what extent the amount of glottal aperture can be controlled by the speaker under static and dynamic speech and nonspeech conditions with and without visual feedback. They suggested that the voluntary control of glottal opening is rather poor, i.e. not as accurate as lip or tongue movements, but similar to movements of the velum. They concluded: glottal opening movements in speech production are commonly related to supralaryngeal events, so that in addition to the degree of glottal opening the timing between laryngeal and supralaryngeal events is important.

The poor voluntary control of glottal opening may also be explained by aerodynamic factors acting at a glottal level, i.e. voluntary control may be poor (in opposition to phonation) since there is no need for it, because sub- and intraoral pressure variations highly influence glottal abduction. In one of the prime examples of vocal fold research a similar issue on physical and centrally controlled factors was raised: Vocal fold vibrations in human speech could be generated either as the result of aerodynamical phenomena interacting with the elastic properties of the vocal folds (the myoelastic theory), or as intended movements that would be precisely controlled via the neuro-muscular system (the neurochronaxic theory). Thanks to the modelling studies from Van den Berg (1958) it has been shown that the neurochronaxic theory was wrong.

In order to verify possible physical factors which could account for the differences in glottal opening amplitude in different intensity conditions we will use physical modelling (see Van Hirtum et al.).

Acknowledgements

This work was supported by a grant from the German Research Council GWZ 4/8-1, P.1. We thank Jörg Dreyer for technical support, Christine Mooshammer for corpus

construction and Anke Busler for segmentation of fiberoptic stills.

5. References

- [1] Andersen, P., 1981. The effect of increased speaking rate and (intended) loudness on the glottal behaviour in stop consonant production as exemplified by Danish p. *ARIPUC* 15, 103-146.
- [2] De Jong, K.J., 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of Acoustical Society of America* 97, 491-504.
- [3] Fuchs, S., 2003. *Articulatory correlates of the voicing contrast in alveolar obstruent production in German*. Unpublished PhD thesis: QMUC Edinburgh.
- [4] Hirose, H., 1975. The posterior cricoarytenoid as a speech muscle. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, University of Tokyo 9, 47-66.
- [5] Hoole, P., 1999. Techniques for investigating laryngeal articulation. Section A: Investigation of the devoicing gesture. In *Coarticulation: Theory, data and techniques*. W. Hardcastle; N. Hewlett (eds.). Cambridge: University Press: 294-300.
- [6] Krakow, R. A., 1999. Physiological organization of syllables: a review. *Journal of Phonetics* 27, 23-54.
- [7] Löfqvist, A., 1980. Interarticulator programming in stop production. *Journal of Phonetics* 8, 475-490.
- [8] Löfqvist, A.; Baer, T. and Yoshioka, H., 1981. Scaling of glottal opening. *Phonetica* 38, 265-276.
- [9] Mc Clean, M.; Tasko, S.M., 2003. Association of orofacial muscle activity and movement during changes in speech rate and intensity. *Journal of Speech Language and Hearing Research* 46, 1387-1400.
- [10] Sawashima M.; Hirose, H.; Ushijima, T. and Niimi, S., 1975. Laryngeal control in Japanese consonants with special reference to those in utterance initial position. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, University of Tokyo 9, 21-26.
- [11] Schulman, R., 1989. Articulatory dynamics of loud and normal speech. *Journal of Acoustical Society of America* 85(1), 295-312.
- [12] Van den Berg, J., 1958. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1(3), 227-244.
- [13] Van Hirtum, A.; Rutu, N.; Pelorson, X.; Fuchs, S. & Perrier, P., 2004. Laryngeal adjustment in voiceless consonant production: II. Physical modelling. *Submitted to this conference*.