

Laryngeal correlates of prosodic variation

Christine Mooshammer¹

¹IPDS, Institute for Phonetics and Digital Speech Processing, Christian-Albrechts
University Kiel
24098 Kiel, Germany

timo@ipds.uni-kiel.de

***Abstract.** The aim of this study is to compare the effects of global vocal effort changes on voice source parameters with effects of focus and word stress. Seven speakers were recorded by means of a laryngograph processor. Glottal pulses were preprocessed by time and amplitude normalization. Apart from the more common shape indicators, the open and the speed quotients, we also carried out the functional version of a Principal Component Analysis. Results on the calculated voice source parameters indicate clear differences for vocal effort changes with a shorter open phase for loud speech as compared to normal and soft speech, but only very inconsistent evidence was found for the involvement of laryngeal mechanisms for signaling word stress or focus.*

1. Introduction

In Germanic languages, linguistic prominence is associated with an increase in duration, fundamental frequency, articulatory precision and intensity (see e.g. Beckman 1986, Fry 1955). Different types of prominence, such as word stress and sentence accent, differ with respect to the contribution of individual parameters to the production of prominence: whereas intensity or vocal effort has been shown to be more closely associated with word stress, sentence accent is signaled by rapid f₀ changes (see e.g. Sluijter & van Heuven 1996) because of the association with a pitch-accent. Both prominence types can be seen as a local enhancement of syllables or words relative to their contexts produced by different means. However, a largely unresolved and controversial issue is whether word and sentence stress changes are controlled physiologically by the same types of mechanisms.

In his pioneering work in 1967 Ladefoged found that subglottal pressure not only contributes to global paralinguistic vocal effort changes, but also to local variations of prominence, namely lexical word stress. In his view, short contractions of the expiration muscles, measured by EMG, cause an increase in subglottal pressure which would also decrease the spectral slope. Indeed, in an acoustical study Sluijter & van Heuven (1996) found that the spectral slope was flatter in stressed vowels independently on their accentual status. They attributed this difference in “spectral balance” to a quicker glottal closure or an increased vocal effort. Marasek (1997) suggested in a modeling study that greater subglottal pressure underlies only word stress whereas sentence accent is primarily controlled by vocal fold tension. However, there is also accumulating

counterevidence: Fant et al. (2000) found an increase in subglottal pressure, measured for one subject by tracheal puncture, comparable to vocal effort increases only for very high levels of emphatic stress but not for stressed vowels produced in more neutral environments. No clear-cut distinction in parameter changes for stress and accent were found by Campbell & Beckman (1997) and Hanson (1997). Furthermore, in a study by Heldner (2003) focal accent was also produced with a lower spectral tilt. All these studies suggest that there is no clear physiological distinction between stress and accent as was proposed by e.g. Sluijter & van Heuven (1996).

In order to explore this issue further, changes due to word stress, sentence accent and changes due to increased vocal effort (by asking a subject to vary loudness) were analysed by using laryngographic techniques. We reasoned that if word stress changes can be attributed to changes of vocal effort, then we should see a similar pattern of laryngographic changes in both word stress and raised loudness which is different from the types of laryngographic change that accompany sentence accent.

2. Method

The test words were two-syllable words in German that in traditional terms varied in primary lexical stress on the first syllable, i.e. the words 'Lena' with primary lexical stress on the first syllable /le:/ (henceforth *s* for strong) and 'Lenor' (name of a washing powder) with primary lexical stress on the second syllable and hence a lexically unstressed, but full vowel /e/ on the first syllable (henceforth *w* for weak). These words were embedded in dialogues which elicited either an accented production in a focused context [F] associated with providing 'new' information; and they were also embedded in sentences in which the word was unaccented because the information was given [U]. Questions were pre-recorded and presented via headphones. The answers were presented on a computer screen. Examples of questions and answers are given below:

a) Focussed and stressed: [F, s]

Q: Wolltest Du Dir Friedas Buch ausleihen? (*Did you want to borrow Frieda's book?*)

A: Nein, ich wollte LENAS Buch ausleihen. (*No, I wanted to borrow Lena's book.*)

b) Unfocussed and stressed: [U, s]

Q: Wie findest Du Lena? (*How do you like Lena?*)

A: Ich HASSE LEna und ihre Schusseligkeit. (*I hate Lena and her absent-mindedness.*)

c) Focussed and weak [F, w]:

Q: Kaufst Du Omo oder Lenor bei Schlecker? (*Do you buy Omo or Lenor at Schlecker?*)

A: Ich kaufe LEnor bei Schlecker. (*I buy Lenor at Schlecker*)

d) Unfocussed and weak [U, w]:

Q: Wäschst Du nicht gern mit Lenor und Omo? (*Don't you like washing with Lenor and Omo?*)

A: Ich HASSE LEnor und Omo. (*I hate Lenor and Omo*)

As a result, we had four possible accent x lexical-stress combinations: [F, s], [F, w], [U, s], [U, w] which were spoken in a comfortable vocal effort level (henceforth N). A further condition was that all the [F, w] combinations were produced by the speakers in either a loud (L) or a soft voice (S). For the loud conditions speakers were instructed to

speaking loudly without shouting; for the soft conditions the instruction was to speak softly without whispering. These 6 (lexical stress x accent in normal loudness + soft and loud levels) possible combinations were repeated nine times in randomized order. Acoustic and laryngographic signals (henceforth Lx) at a sampling rate of 16 kHz were obtained from 7 male subjects between 20 and 35, speaking a northern variety of Standard German. Some items had to be excluded because of mispronunciation or because the Lx signal was distorted.

From the speech signal the RMS energy was measured at the acoustically defined mid of the vowel /e:/ from the test words. From the Lx signal the two medial pitch periods during the vowel were extracted and further processed. Since we were not interested in shape differences depending on f_0 and since it is well-known that the Lx signal cannot be calibrated, the two pulses of all items were time and amplitude normalized to a uniform length of 1000 samples and an amplitude of 1 for the first glottal closure. In order to compensate for vertical larynx movements a line connecting the two minima was subtracted.

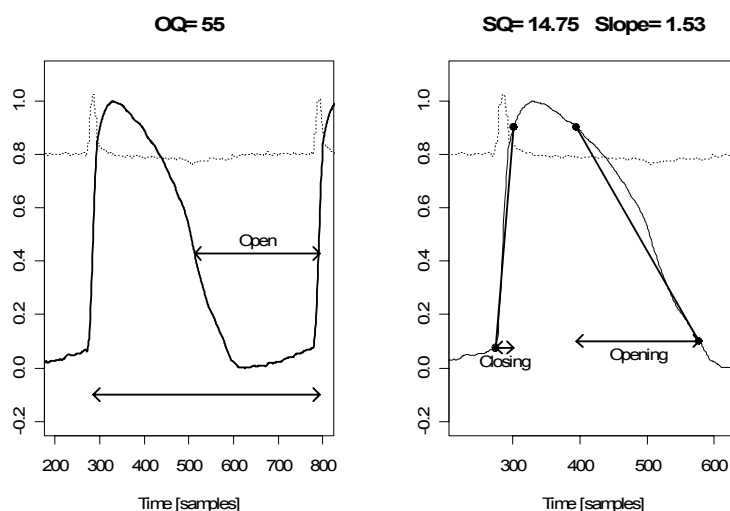


Figure 1. Left: Calculation of the open quotient (OQ); right: speed quotient (SQ) and the closing slope. Solid line: normalized glottal pulses, dotted line: first derivative in arbitrary units

The following EGG parameters were computed for the medial pitch period during the vowel /e/ in all test conditions using EMU/R: Open quotient (OQ, using the 3/7 threshold as instant of glottal opening as suggested by Howard 1990, see Fig. 1, right) was calculated as the percentage of the open glottis interval to the pitch period duration. The speed quotient (SQ, using a 10% threshold as suggested by Marasek 1997, see Fig. 1, left) was computed as the ratio between the closing duration and the opening duration. Furthermore the slope of glottal adduction was also computed. According to simulations by Marasek (1997) and results from the literature the parameters should change in the following way:

OQ: decrease for loud speech and in stressed items, increase for accented vowels

SQ: not change for loud speech and in stressed items, increase for accented vowels

Slope: increase for vocal effort and stress, no change for accent

An alternative more holistic approach to analyzing derived parameters such as OQ and SQ is the application of Functional Data Analysis (FDA, see e.g. Ramsay and Silverman 1997, Lucero and Koenig 2000). This method calculates functional versions of time-variant digital data by computing splines (in our case Fourier basis functions). One of the major advantages is that this method takes into account the continuity underlying the physiological system generating these data. After computing the spline functions a Principal Component Analysis was applied to the data. This kind of factor analysis extracts the relevant parameter for explaining the variance in the data and transforms the raw data to a new coordinate system. The cell means of the resulting factor scores and of SQ, OQ, Slope and RMS were further analysed by repeated measures ANOVA with Greenhouse-Geisser correction for violations of the sphericity assumption using R. For significant effects pairwise t tests with Bonferroni adjustments were calculated. Since the design was not full-factorial, i.e. vocal effort was only varied for the stressed and accented items, two separate analyses were carried out with subsets of data: for linguistic prominence only data spoken in normal volume were used with the independent variables stress and accent. For effects of vocal effort only data from accented and stressed vowels were taken into account.

3. Results

Table I gives the results of repeated measures ANOVAS for linguistic prominence (stress and accent) and vocal effort as independent variables and the dependent variables RMS, OQ, SQ and Slope. Figure 2 shows the mean values and standard deviations for the 6 conditions. Data used for vocal effort comparisons are plotted in dark grey and for linguistic prominence in light grey. The bar plotted in medium grey corresponds to data from stressed accented items in normal speech volume which are used in both subsets of the data.

Table I: Results of repeated measures ANOVA and pairwise t-tests for the dependent variables RMS, OQ, SQ and CSlope and the factors prominence and vocal effort.

Variable		Prominence	sig. t tests	vocal effort	sig. t tests
RMS	F df	21.16 *** 2.2, 13	[F, s] > [U]	108.6 *** 1.7, 10.1	L>N>S
OQ	F df	5.7 * 2.1, 12.4		13.1 * 1.3, 7.6	L<S
SQ	F df	2.13 1.6, 9.4		0.96 1.1, 6.8	
CSlope	F df	0.4 2.1, 12.7		0.63 1.6, 9.9	

Table I indicates that RMS was significantly affected by loudness and also by linguistic prominence with accented stressed items being louder than deaccented stressed and

unstressed items. As was mentioned before, we would expect glottal parameters of stressed items to pattern with loud items and unstressed with soft. However, none of the derived parameters OQ, SQ and CSlope showed significant differences for linguistic prominence. There was slight tendency for OQ to be lower for stressed items than for unstressed. Loudness affected only the OQ significantly with loud volume being produced with a significantly shorter open phase than soft volume.

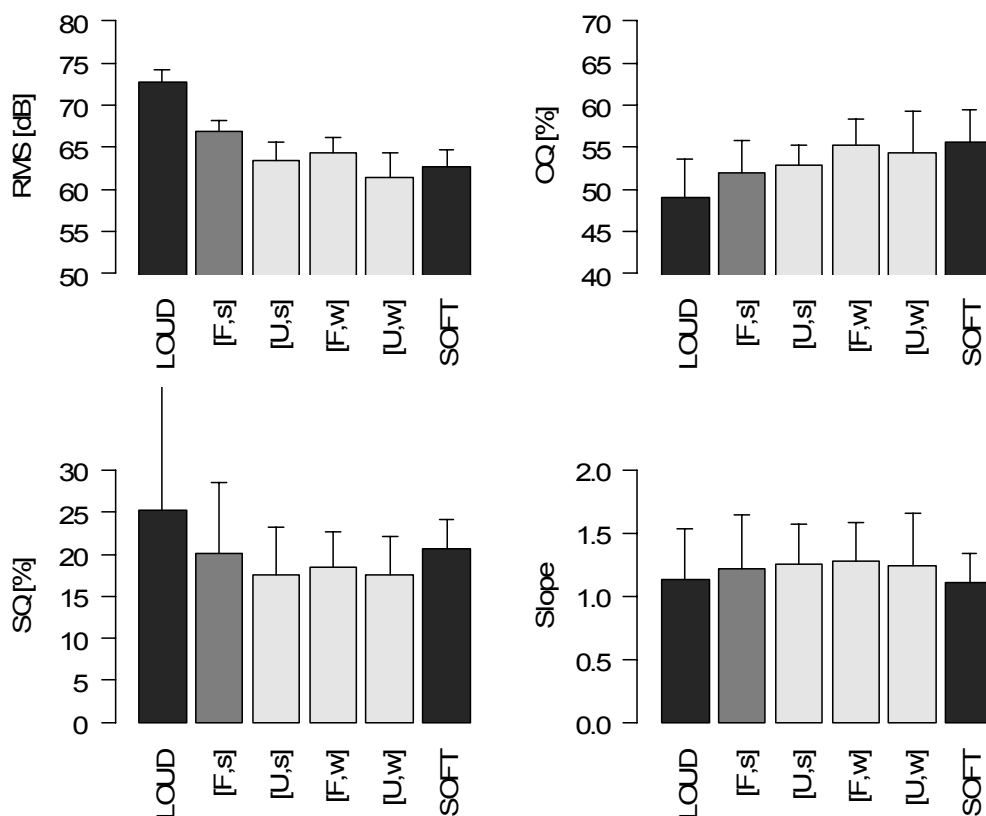


Figure 2. Means and standard deviations for the dependent variables RMS, OQ, SQ and Closing Slope for vocal effort and prominence variations.

Results of the PCA are shown in Figure 3. The upper two panels show the shapes of the glottal pulse for negative factor values indicated by minus signs, positive factor values by plus signs and the mean curve by a solid line. Shape differences as shown for the first (left) and second (right) factor explain about 64% and 14% of the variance in glottal pulse shapes. Positive values of the first factor are characterized by a longer open phase as can be seen from the longer flat and low plateau before the glottis is closing compared to the mean curve. Negative factor values can be characterized by a shorter open phase and by a steeper and higher rise. The second factor varies only very little. There is a tendency for a more symmetric shape for positive factor values.

In the two panels below mean and standard deviations of the factor scores of the first and second factor are displayed. The higher the values are the more similar is the shape to the “positive” glottal pulse shape. Hence, for the first factor glottal pulses from loud speech exhibit a “negative” pulse shape with a short open phase and a steep and short

closing phase. Soft speech is more similar to the positive glottal pulse and involves a longer open phase. Interestingly, the stressed items pattern with loud speech and the unstressed items with soft speech. For the second factor no clear picture emerges. Therefore, statistics are only calculated for the first factor and presented in Table II. In the first line denoted by “All” repeated measures ANOVA’s were calculated as described above for the other variables.

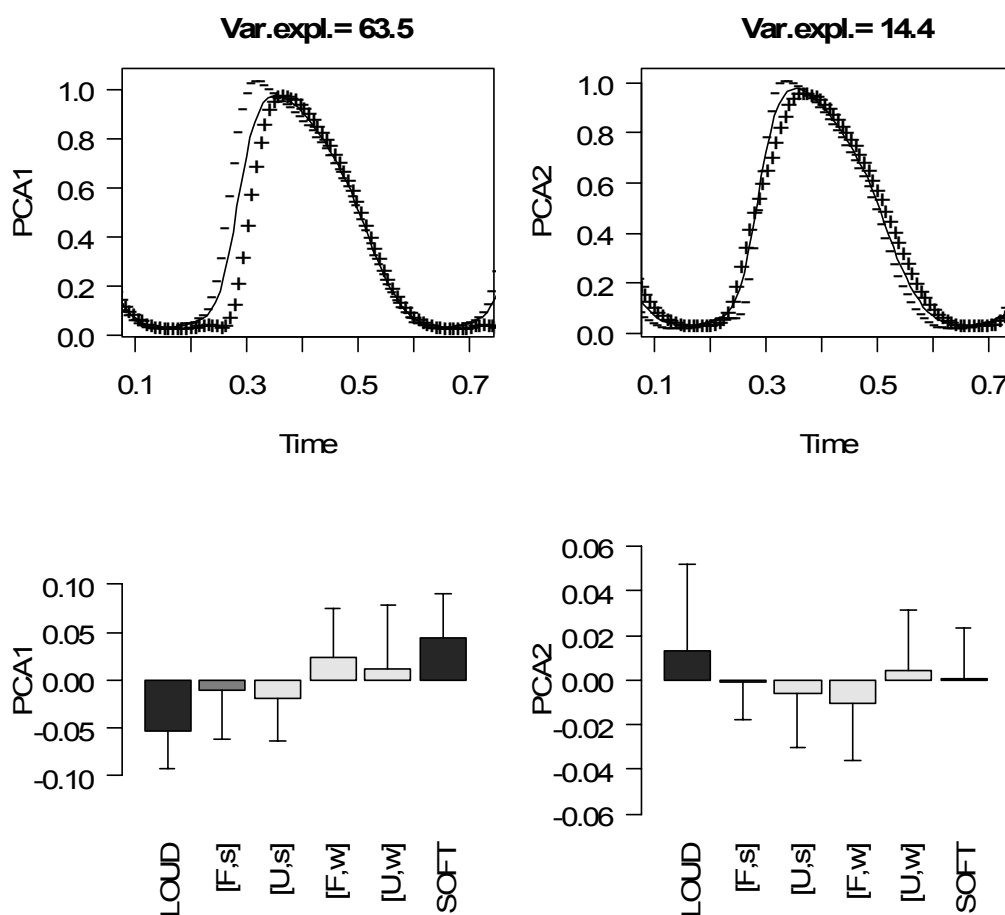


Figure 3. Upper panels: the shapes of the glottal pulse for negative factor values indicated by minus signs, positive factor values by plus signs and the mean curve by a solid line for the first factor (left) and the second factor (right). Lower panels: means and standard deviations of the factor scores for vocal effort and prominence variations.

According to our predictions glottal pulses for stressed items should pattern with loud speech and for unstressed items with soft speech. Scores of the first principal component shown in Figure 3, and results from the repeated measures ANOVA seem to confirm this assumption. However, when calculating PCA’s for individual speakers only one speaker, Sp7, distinguishes strong from weak syllables by glottal shape independently of accent. All other speakers either exhibit significant differences in the opposite direction (see Sp1, Sp2 and Sp.4) or none at all (Sp3, Sp5, Sp6). For the vocal effort condition, however, speakers behave much more consistently: 6 out of 7 speakers

produce loud speech with a different glottal shape as compared to normal and soft speech.

Table II: Results of repeated measures ANOVA and pairwise t-tests for the scores of the first principal component and the factors prominence and vocal effort in the first line and fixed effects ANOVAs for individual speakers below. Significant results contrary to our assumptions are printed in italics. The variance explained by the first factor is reported in the second column.

PCA1	Var. expl.	Prominence	sig. t tests	vocal effort	sig. t tests
All	63.5	5.1 * 2.7, 16.4		11.6 * 1.3, 7.9)	L < S
Sp1	78.6	4.5 ** 3, 32	[U, s] < rest, <i>[U, s] < [F, s]</i>	29.1 *** 2, 24	L < N, S
Sp2	76.9	10.5 *** 3, 32	[s] < [F, w] <i>[U, w] = L</i>	22.2 *** 2, 23	L, N < S
Sp3	78.6	0.2 3, 28		39.7 *** 2, 21	L < N < S
Sp4	78.7	4.9 ** 3, 32	[F.s] < [U.w] <i>[U.w] < [U.s]</i>	5.6 * 2, 24	L < N
Sp5	66.4	2.74 3, 30		1.3 2, 24	
Sp6	75.1	1.3 3, 26		20.5 *** 2, 23	L < N < S
Sp7	75.5	7.24 *** 3, 31	[s] < [w]	24.3 *** 2, 24	L < N < S

4. Discussion

In this study glottal pulse shapes, derived from laryngographic data, and their parameterizations were compared for 3 levels of vocal effort and 4 levels of linguistic prominence. Based on the literature we predicted that stressed items should be produced with more vocal effort than unstressed items.

For vocal effort changes the Open Quotient decreased going from soft to loud speech. This was also reflected in the factor scores of the first Principal Component which distinguished glottal pulse shapes mainly by the length of the glottal open phase. Vocal effort does not affect the steepness of glottal closure or the symmetry of the pulse, at least not for laryngographic data (see Marasek 1997 for a detailed discussion).

Linguistic prominence did not affect the Open Quotient as predicted by the results from vocal effort. Furthermore stressed items were not consistently produced by glottal pulses resembling loud speech, and unstressed items were not similar to soft speech. Two possible suggestions might explain the discrepancy between results from e.g. Sluijter and van Heuven (1996) and Ladefoged (1967): First, it could be that the laryngographic signal is not suitable for registering subtler differences in the glottal pulse shape. However, this can be partly refuted by the rather consistent and stable results for vocal effort changes. Secondly, the stressed-unstressed distinction is not produced by a higher subglottal pressure or different glottal settings. Similar

conclusions have also been reached by Campbell & Beckman (1997), Fant et al. (2000) and Hanson (1997). Whether upper vocal-tract adjustments for the stressed-unstressed distinction account for changes in spectral tilt, as presented by Sluijter & van Heuven (1996), will be investigated in the near future.

Acknowledgements

I would like to thank Christian Geng, Suse Fuchs and Phil Hoole for their valuable suggestions, Jonathan Harrington for help with the corpus design and with EMU/R, Jennifer Schneeberg for data labeling and Herbert Fuchs for assisting the recordings.

References

- Beckman, M. E. *Stress and Non-Stress Accent*. Dordrecht, The Netherlands: Foris, 1986.
- Campbell, N. and Beckman, M. E. Stress, prominence and spectral tilt. In *Proc. ISCA Workshop on Intonation: Theory, Models and Applications*, pages 67–70, 1997.
- Fant, G.; Kruckenberg A. and Liljencrantz. J. Acoustic-phonetic analysis of prominence in Swedish. In A. Botinis, editor, *Intonation: Analysis, modelling and technology*, pages 55–86. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2000.
- Fry, D. B. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27:765–768, 1955.
- Hanson, H. Vowel amplitude variation during sentence production. In *Proceedings ICASSP*, pages 1627–1630, 1997.
- Heldner, M. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, 31:39–62, 2003.
- Howard, D. M., Lindsay, G. A. and Allen, B. Toward the quantification of vocal efficiency. *Journal of Voice*, 4:205–212, 1990.
- Ladefoged, P. Stress and respiratory activity. In *Three Areas of Experimental Phonetics*, pages 1–49. London: Oxford University Press.
- Lucero, J. and Koenig, L. Time normalization of voice signals using functional data analysis. *Journal of the Acoustical Society of America*, 108:1408–1420, 2000.
- Marasek, K. Electroglottographic description of voice quality. In *AIMS*. Stuttgart, 1997.
- Ramsay, J. O. and Silverman, B. W. *Functional data analysis*. New York: Springer, 1997.
- Sluijter, A. and van Heuven, V. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100:2471–2485, 1996.