An EGG study on global and local vocal effort changes in German: preliminary results

Christine Mooshammer

Institut für Phonetik und digitale Sprachverarbeitung
Christian-Albrechts Universität zu Kiel, Germany
timo@ipds.uni-kiel.de

Abstract

The aims of this study are (1) to compare the effects of global vocal effort changes on voice source parameters with effects of focus and word stress, (2) to tease apart the influences of subglottal pressure and vocal fold tension by independently varying sentence accent and word stress. Therefore two EGG experiments were carried out. Results on the calculated voice source parameters indicate that in the absence of f0 differences word stress is indeed produced with a higher subglottal pressure. For focus, both seems to be involved: subglottal pressure and vocal fold tension.

1. Introduction

There is general agreement that subglottal pressure is the main contributor to variation of loudness. An increase in subglottal pressure causes a faster closing of the vocal folds and a longer closed phase which in turn decreases the steepness of the spectral slope (e.g. [3]). As was found by Ladefoged (1967) [9] in his pioneering work, subglottal pressure not only contributes to global paralinguistic vocal effort changes, but also to local variations of prominence, namely lexical word stress.

In Germanic languages prominence is produced by an increase in duration, fundamental frequency and intensity. At least two types of prominence can be distinguished: lexical word stress and sentence accentuation or focus. These two types differ with respect to the contribution of individual parameters to the production of prominence, whereas intensity or vocal effort is more relevant for increasing word stress, sentence accent is signaled by rapid f0 changes (see e.g. [12]) because of the association with a pitch-accent and, if the accented word is near the end of the prosodic phrase, due to the influence of boundary tones. Both prominence types can be seen as a local enhancement of syllables or words relative to their contexts.

Both, global and local intensity changes, have been studied extensively in the past (e.g. for global intensity [2, 6, 10, 13], for word stress [9], for focus [4]). Up to now the only systematic comparison between global paralinguistic and local linguistic changes in vocal effort was carried out by Pierrehumbert 1997 [10]. She analyzed the interaction between pitch accents and global intensity changes by adjusting the Liljencrants-Fant model parameters to the semi-automatically inverse filtered speech signal and concluded that (1) the increase of F0 for loud speech cannot simply be attributed to consequence of increased subglottal pressure but also involves some laryngeal adjustments and (2) that because parameters such as the Open Quotient and the skewness of the glottal pulse are affected by vocal effort changes and tonal adjustments in a different manner it is rather intricate to tease apart these two factors.

The aim of the current study is to investigate source characteristics for two different levels of stress (stressed vs. unstressed [±S]), two levels of focus (sentence accents due to new and given information [±F]) and three levels of global loudness (loud, normal and soft) in German. The hypothesis is that changes of voice source parameters due to linguistic prominence are in the same direction as global changes in vocal effort. Therefore two experiments were carried out: in the first laryngographic recording sentence accent was varied by simply underlining the words which should be accented by the subjects. Since this experimental setup didn't elicit deaccentuation a second more natural question-answer experiment was designed, where the deaccentuated word was in postfocal position.

2. Experiment 1

2.1. Recordings

Two male speakers of Standard German (both non-smokers between 20 and 30) have been recorded by means of a Laryngograph processor. The material consisted of real words containing [zVn] sequences with the tense vowels /i, e, a/. The test sequences were embedded in words with different stress patterns, e.g. *Sahne* (cream) stressed on the first syllable vs. *sanieren* (to redevelop) stressed on the second syllable. All test words were embedded in the carrier sentence "Ich habe ____ gesagt" (*I said ____.).* Sentence accent was elicited by instructing the subjects to focus on the underlined word which was either the test word (accented) or the sentence-final word *gesagt* (unaccented). For intensity variation, the subjects were instructed to speak at a comfortable intensity level (normal), as loudly as possible without shouting (loud), and softly without whispering (soft). All sentences were repeated at the three volume levels five times in randomized order.

All words containing /a/ were acoustically labeled using the PRAAT software. The derivative of the Lx signal (DEGG) was also computed by PRAAT. The following EGG parameters were computed for all periods during the vowels using EMU/R: Open quotient (OQ, using the 4/7 threshold as instant of glottal opening as suggested by Howard [7]), maximum of the derivative of Lx during closing and opening (Cpeak and Opeak), speed quotient (SQ, using a 10% threshold as suggested by Marasek [9]) and the slope of glottal adduction and abduction (Cslope and Oslope) (see Table 1). Intensity was measured by computing the RMS of the speech signal. For calculating fundamental frequency the peak of the Lx derivative closing phase was used. Besides abbreviations and description of the EGG parameters Table 1 also summarizes the predictions for changes of these

parameters due to effects of subglottal pressure (L) and active control of fundamental frequency (F0). The prediction given here are based on the computer simulations by Marasek [9] and some other sources. Because of the difficulty to tease apart the influences of subglottal pressure and laryngeal adjustments on the shape of the glottal pulse, we found predictions based on a model in which both dimensions can be varied independently more promising than experimental data.

Table 1: *Analysed EGG parameters, their description and predictions for changes due to increased subglottal pressure (L) and changes due to higher f0 actively controlled by laryngeal muscles (F). Predictions are mainly summarized from Marasek's model [9], (sometimes different predictions for airflow and EGG simulations are denoted by AF and EGG respectively). Other references are given in brackets.*

	Description	Prediction
OQ	Open Quotient: Percentage of open phase	L: decrease (AF), increase (EGG) F: increase
SQ	Speed Quotient: Symmetry of the glottal pulse	L: decrease (AF), not usable (EGG) F: decrease (AF)
Cpeak	Closing peak of the first derivative	L: increase [3] F: no prediction
Cslope	Slope of the glottal adduction	L: increase F: no effect
Opeak	Opening peak of the first derivative	L: increase [1] F: no prediction
Oslope	Slope of the glottal abduction	L: increase F: no effect

2.2. Results

2.2.1. Global vocal effort changes

For both speakers, differences in intensity levels between loud, normal and soft were highly significant and amounted to about 12 dB each for speaker GA and 7 dB for speaker BD. As was expected f0 also decreased significantly with decreasing vocal effort. Furthermore all EGG parameters showed significant differences in the expected direction: decrease of the open and speed quotient as well as an increase of the DEGG closing peak and the closing slope going from soft to normal (n.s. for both speaker DEGG closing peak, see 3.2.1 for a discussion) and from normal to loud speech (n.s. for speaker BD, speed quotient), which provides evidence for subglottal pressure differences. Contrary to the expectations the slope and DEGG peak for glottal opening increased going from normal to soft, which means that they approached the values for loud speech for both speakers. Parameters of the opening phase (Opeak and Oslope) have been investigated in a recent study: Alku et al. (2001) [1] found that for very loud voices the relative contribution of the opening phase can be seen as a secondary excitation of the vocal tract. In accordance with their study these parameters increased significantly going from normal to loud voice but not going from soft to normal voice. Further evidence for a lower subglottal pressure in soft speech is given by the fact that, whereas in loud and normal speech no devoicing of the syllable-initial [z] occurred, 54 % of [z] tokens for speaker GA and 42 % for speaker BD were fully or partially devoiced in soft speech (in syllable-initial position [s] and [z] do not contrast in German).

2.2.2. Local vocal effort changes

As discussed earlier, the final word *gesagt* was focused in order to elicit deaccentuation of the preceding test word. Unfortunately, and as could be seen by analyzing the f0 contours, the test words were produced with an f0 peak in most of the cases which was only somewhat lower than the f0 peak of the focused word. Therefore the effect of word stress will be discussed only in the accented condition. For studying the effect of focus, a second experiment was designed.

Table 2: *Significant difference for global (upper part) and local (lower part) effort changes calculated – using pairwise t-tests with Bonferroni adjustments. Parameters printed in italics contradict the predictions given in Table 1.*

GLOBAL	Speaker BD	Speaker GA
loud - normal	RMS, f0, OQ, Cpeak, Cslope, Opeak	RMS, f0, OQ, SQ, Cpeak, Cslope, Opeak, Oslope
Normal – soft	RMS, f0, OQ, SQ, Cslope, *Oslope*	RMS, f0 OQ, SQ, Cslope, *Opeak, Oslope*
LOCAL	Speaker BD	Speaker GA
Stressed - unstressed	RMS, f0, OQ, SQ, Cpeak, Cslope, *Opeak, Oslope*	f0 OQ *Cpeak, Cslope, Oslope*

As can be seen in Table 2 only speaker BD showed a significant decrease of intensity going from stressed to unstressed. This speaker also showed a significant increase of OQ and significant decreases of the speed quotient, the DEGG peak (Cpeak) and the slope of the glottal closing (Cslope). For the other speaker (GA), only f0 and OQ showed significant differences, and also significant changes that contradicted the prediction, i.e. Cpeak and Cslope were closer to loud speech for unstressed syllables.

Therefore the two speakers varied for the way in which they signalled word stress: speaker BD enhanced the stressed syllable by increasing the intensity and also showed significant differences of Lx parameters in the direction of loud speech whereas speaker GA marked the stressed syllable only by an F0 rise. For both speakers, the Open Quotient was significantly affected by the stress level. On the one hand this parameter is said to decrease with a tenser voice but on the other hand it increases with F0. As was found by Marasek [9], the Speed Quotient (airflow) decreases and the Closing Slope increases for an increase of subglottal pressure but not for an increase in F0 due to a higher muscular tension where the glottal pulse stays symmetrical. Therefore we assume that besides an increase in vocal fold tension speaker BD produced accented stressed syllables by an increase in subglottal pressure compared to accented unstressed syllables. This was not the case for speaker GA.

When inspecting the EGG parameters for the whole word we found that the initial voiced fricative not only showed a high amount of devoicing for soft voice but also affected the

source parameters at the beginning of the following vowel. Especially the Open quotient tended to be very high at the onset of the vowel which can be attributed to the fact that voiced sibilants are often produced with arytenoid separation [11]. Additionally the experimental design failed to elicit de-accentuation of the test word in most of the cases. Assuming that for deaccentuated words the stressed and the unstressed syllables are produced with the same low f0 it should be possible to tease apart the melodic and pressure effects.

3. Experiment 2

3.1. Recordings

In the second experiment, variation of accentuation was elicited by using a question - answer paradigm where the test word was given information in postfocal position (denoted by [-F]) or new information in focal position (denoted by [+F]). Questions were recorded beforehand and presented to the speakers by headphones. The test sentence was presented simultaneously on a monitor. They consisted of /le:n/ or /ze:n/ sequences with or without word stress (denoted by [+S] or [-S]). Sentences were constructed as follows with the test words underlined:

1a) Q: Wolltest Du Dir Friedas Buch ausleihen?
    A: Nein, ich wollte LENAS Buch ausleihen. [+F, +S]
1b) Q: Wie findest Du Lena?
    A: Ich HASSE Lena und ihre Schusseligkeit. [-F, +S]
2a) Q: Kaufst Du Omo oder Lenor bei Schlecker?
    A: Ich kaufe LENOR bei Schlecker. [+F, -S]
2b) Q: Wäschst Du nicht gern mit Lenor und Omo?
    A: Ich HASSE Lenor und Omo. [-F, -S]

In sentences 1a) and 1b) the test word "Lena" (a person's name) is stressed on the first syllable ([+S]) as opposed to "Lenor" (brand name of a washing powder) in sentences 2a) and 2b) in which the first syllable is unstressed ([-S]). In sentences 1a) and 2a), the test words are produced by focal accentuation ([+F]) whereas in sentences 1b) and 2b) they are deaccentuated ([-F]) due to given information and the postfocal position after the emphasized word "hasse" (hate). Similar sentences were constructed with the words "Sehnen" (sinews) and "Senat" (senate). Sentences where the test sequence was stressed and accented were also produced in loud and soft voices. All sentences were repeated eight times in randomized order. 4 subjects (all male non-smokers between 20 and 30, speaking a northern variety of Standard German) have been recorded by means of a Laryngograph processor, but only speaker GA has been analyzed so far (for details of the measurements see experiment 1).

3.2. Results

3.2.1. Global vocal effort changes

In the second experiment changes from loud to normal for initial /z/ affected the same parameters as in experiment 1. However, the change from normal to soft voice had neither an effect on the parameter SQ after /z/ nor after /l/. From both experiments, it can be summarized that changes due to a reduction of global vocal effort affect fewer parameters of the Lx signal compared to an increase of vocal effort. These effects are even weaker for initial /l/ compared to /z/.

As can be seen in Table 3 the parameter Cpeak, i.e. the closing peak of the derivative, does not play an important role in distinguishing between the three vocal effort levels. As has often been observed this parameter largely controls the overall spectral level and increases with higher voice levels. Therefore we assumed it to be consistently affected by volume increase. The reason that this was not the case can be found in the shape of the derivative: many cases showed multiple peaks. The distribution of the occurrence of single peaks is given in Table 4 (the criterion for a single peak was that the other peaks occurring during the closing phase do not exceed 50 % of the maximum) and suggests that it is positively related to vocal effort. Only 13 % of the vowels following /l/ produced in loud speech showed a single peaked closing phase compared to 87% for soft voice. Multiple peaks exhibited significantly lower values compared to single peaks. As to why multiple peaks in loud speech occur is again a matter of speculation. For glottal closing, Henrich et al. [5] found via simultaneous high speed and electro-glottographical recordings that the double closing peak was related to a time-lag of the closing of different parts of the glottis. They also found that the occurrence of these double peaks was highly speaker-dependent and could not attributed to a general pattern. Because of the high percentage of double peaks in our data, they could not be excluded from further analyses, but their occurrence can explain why this parameter quite frequently varies in the opposite direction. Although the opening peak was never well defined, i.e. never single-peaked, it showed a somewhat more consistent behavior especially with respect to local vocal effort changes.

Table 3: *Significant difference for global (upper part)and local (lower part) effort changes for speaker GA calculated - using pairwise t-tests with Bonferroni adjustments. Parameters printed in italics contradict the predictions given in Table 1.*

GLOBAL	/l/	/z/
loud - normal	RMS, f0, OQ, SQ Cslope, Opeak, Oslope	RMS, f0, OQ, SQ, Cpeak, Cslope, Opeak, Oslope
normal – soft	RMS, f0, OQ *Cpeak Opeak*	RMS, f0 OQ Cslope, Oslope
LOCAL	/l/	/z/
[+F]: stressed - unstressed	RMS, f0 OQ *Cpeak* Opeak, Oslope	RMS, f0 OQ, SQ *Cpeak* *Opeak, Oslope*
[-F]: stressed - unstressed	RMS Cpeak, Cslope Opeak, Oslope	OQ *Oslope*
[+S]: focus - non-focus	RMS, f0 OQ, SQ *Cpeak*, Cslope Opeak, Oslope	RMS, f0 OQ, SQ *Cpeak*, Cslope Opeak, Oslope
[-S]: focus - non-focus	RMS, f0 SQ Cpeak, Cslope Opeak, Oslope	RMS, f0 Cslope Opeak, Oslope

3.2.2. Local vocal effort changes

After the sonorant /l/, word stress significantly increased intensity and f0 for focused words and only intensity for non-focused words. The Open Quotient differs for word stress in the focused position only, which confirms the influence of f0 on this parameter. If the tonal differences are neutralized, then the slopes of the glottal pulse and the peaks of the derivative are significantly higher for stressed as compared to unstressed tokens. The symmetry of the pulse (SQ) was not affected by word stress in the unfocussed condition.

After the voiced fricative, stressed and unstressed vowels only differed in intensity and f0 when focussed. In the unfocussed condition only, the Open Quotient showed significant effects due to word stress. One reason for the different effects might be a rather peculiar pattern of devoicing for this speaker which is given in Table 4. Speaker GA devoices most of the initial /z/'s in unstressed but focussed syllables but none in unstressed and unfocussed syllables. This pattern wasn't observed in experiment 1 but devoicing influences the first few EGG periods of the fol-lowing vowel. Therefore the data for the accented unstressed as well as the unaccented stressed items (with 50 % of devoicing) show a very high amount of variability which is probably the reason why fewer parameters reach significance as compared to the items with an initial sonorant. As to why this pattern of devoicing occurs for this speaker we can only speculate. Obviously, Data on more subjects are needed.

Focus affects RMS and F0 in stressed and unstressed vowels, i.e. focussed unstressed vowels exhibit a higher fundamental frequency and intensity than non-focussed unstressed vowels. Therefore the domain of focus seems to be the word and not just the stressed syllable. The EGG parameters most consistently affected by deaccentuation of stressed syllables were the same as for the difference between loud and normal voice (except for Cpeak after /z/). Apart from the lack of effect on OQ after both consonants and SQ after /z/ this was also true for unstressed vowels.

Table 4: *Percentage of /z/ devoicing (DEV) and single peaks of the first derivative within the closing phase.*

GLOBAL	L		N			S
LOCAL	+F +S	+F +S	+F -S	-F +S	-F -S	+F +S
Devoice %	0	0	87	50	0	50
/z/-single %	18	42	75	96	96	82
/l/-single %	13	66	92	92	100	87

4. Discussion and Conclusion

The aim of this study was to compare the local effects of focus and word stress with global changes of vocal effort. Generally the results indicate that changes due to a higher level of prominence were in the direction of loud speech. One of the desiderata of this preliminary study is to find independent parameter sets for the two linguistic factors word stress and focus. It has been suggested that sentence accent is produced mainly by an active control of the fundamental frequency whereas increased vocal effort should contribute to the perception of contrastive word stress (see e.g. [12]). This could only be confirmed for the dimension word stress: when unfocussed stressed and unstressed vowels differed only in

RMS not in f0. The affected parameters in this case were the slopes of glottal closing and opening as well as the closing and opening peaks of the first derivative. These parameters were also suggested as determining word stress by Marasek. For the dimension focus the results are much more complicated: besides parameters such as OQ and SQ the slopes also increased significantly in focused words. Following Marasek's simulations steeper slopes of glottal adduction and abduction are caused by an increased subglottal pressure and shouldn't change for a controlled f0 variation. Obviously, data on more subjects are needed for obtaining more conclusive results.

5. References

[1] Alku, P.; Vintturi, J.; Vilkman, E. 2001. Evidence of the significance of secondary excitations of the vocal tract for vocal intensity. *Folia Phoniatrica et Logopaedia* 53, 185-197.

[2] Dromey, C.; Stathopoulos, E.; Sapienza, C. 1992. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *Journal of Voice* 6, 44-54.

[3] Fant, G. 1959. Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics* No. 1.

[4] Gendrot, C. 2003. EGG and spectral investigations on final focalized positions in French. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 547-550.

[5] Henrich, N.; D'Alessandro, C.; Doval, B.; Castellengo, M. 2004. On the use of the derivative of electro-glottographic signals for characterization of non-pathological phonation. *Journal of the Acoustical Society of America* 115, 1321-1332.

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

[7] Howard, D.; Lindsey, G.; Allen, B. 1990. Towards the quantification of vocal efficiency. *Journal of Voice* 4, 205-212.

[8] Ladefoged, P. 1967. Stress and respiratory activity. In: *Three Areas of Experimental Phonetics*, London: Oxford University Press, 1-49.

[9] Marasek, K. 1997. Electroglottographic description of voice quality. *Arbeitspapiere des Instituts für maschinelle Sprachverarbeitung*, Stuttgart, 3(2).

[10] Pierrehumbert, J. 1997. Consequences of intonation for the voice source. In *Speech Production and Language: In Honor of Osamu Fujimura*, S. Kiritani, H. Hirose and H. Fujisaki (Eds.). Berlin: Mouton de Gruyter, 111-130.

[11] Sawashima, M. 1970. Glottal adjustments for English obstruents. *Haskins Status Reports on Speech Research* 21/22, 187-200.

[12] Sluijter, A. & van Heuven, V. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471-2485.

[13] Titze, I.; Sundberg, J. 1992. Vocal intensity in speakers and singers. *Journal of the Acoustical Society of America* 91, 2936-2946.