

# Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German<sup>a)</sup>

Christine Mooshammer<sup>b)</sup>

Haskins Laboratories, 300 George Street, Suite 900, New Haven, Connecticut 06511

(Received 30 December 2008; revised 23 November 2009; accepted 23 November 2009)

This study uses acoustic and physiological measures to compare laryngeal reflexes of global changes in vocal effort to the effects of modulating such aspects of linguistic prominence as sentence accent, induced by focus variation, and word stress. Seven speakers were recorded by using a laryngograph. The laryngographic pulses were preprocessed to normalize time and amplitude. The laryngographic pulse shape was quantified using open and skewness quotients and also by applying a functional version of the principal component analysis. Acoustic measures included the acoustic open quotient and spectral balance in the vowel /e/ during the test syllable. The open quotient and the laryngographic pulse shape indicated a significantly shorter open phase for loud speech than for soft speech. Similar results were found for lexical stress, suggesting that lexical stress and loud speech are produced with a similar voice source mechanism. Stressed syllables were distinguished from unstressed syllables by their open phase and pulse shape, even in the absence of sentence accent. Evidence for laryngeal involvement in signaling focus, independent of fundamental frequency changes, was not as consistent across speakers. Acoustic results on various spectral balance measures were generally much less consistent compared to results from laryngographic data. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3277160]

PACS number(s): 43.70.Fq, 43.70.Gr, 43.70.Aj [DAB]

Pages: 1047–1058

## I. INTRODUCTION

In Germanic languages, prosodic variation is associated with changes in duration, fundamental frequency, intensity, and articulatory precision (e.g., Fry, 1955; Lehiste, 1970; Beckman, 1986). Most types of prosodic variation can be seen as a local enhancement of the prominence of syllables or words relative to their contexts. 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  $f_0$  changes (e.g., Sluijter and van Heuven, 1996). However, a largely unresolved and controversial issue is whether lexical word stress and sentence accent changes are controlled physiologically by the same types of mechanisms. In this study, effects of linguistic prominence on acoustic and physiological measures are compared to effects of global vocal effort changes. The linguistic conditions studied here are lexical word stress with the two levels stressed and unstressed, and sentence accent with the two levels accented and unaccented. This latter variation was elicited by varying the focus.

This introduction is structured in the following way: first, general production mechanisms for vocal effort and linguistic prominence enhancements are reviewed. Then, laryngographic (henceforth Lx) and acoustic reflexes of these

mechanisms are discussed with respect to different types of prominence enhancement. Finally, the aims of this study are stated.

### A. Production of prominence enhancement

Raising the voice in order to speak more loudly can be achieved by two main mechanisms: increase in respiratory force yielding an increase in subglottal pressure (e.g., Ladefoged, 1967; Hixon, 1973) and increase in laryngeal activity, e.g., glottal adduction and adjustments of the length and stiffness of the vocal folds (e.g., Hirano *et al.*, 1969). The usage of these two mechanisms depends on several factors: As was shown by Finnegan *et al.* (2000), subglottal pressure makes a much larger contribution to vocal effort changes than laryngeal activity. Furthermore, a great deal of speaker-specific variability has been observed in the choice of strategies for effecting changes in vocal effort (Stathopoulos and Sapienza, 1993). Titze (1988) and Winkworth *et al.* (1995) suggested that the preference of one mechanism over the other might depend on the speech task: more global and longer-lasting increases in vocal effort are produced by the respiratory system, whereas shorter and more local changes such as word stress or focus are signaled by the much faster and more flexible laryngeal system. However, as reported by Finnegan *et al.* (2000), laryngeal muscle activity played a minor role to intensity regulation in sentence accent since the respiratory system was also capable of initiating more rapid changes in subglottal pressure. Stathopoulos and Sapienza (1993) showed that these two mechanisms are difficult to discern by airstream or Lx measurements since for higher levels of vocal effort, both laryngeal and respiratory mechanisms affected the shape of the Lx pulses with a quicker and more abrupt closing of the vocal folds and a longer closed phase.

<sup>a)</sup> Portions of this work were presented at the International Conference on Voice Physiology and Biomechanics, Marseille, August 2004, the Between Stress and Tone Conference, Leiden, June 2005 and the 7th International Seminar on Speech Production, Ubatuba, December 2006.

<sup>b)</sup> Electronic mail: tine.mooshammer@haskins.yale.edu

Several other parameters are also affected by increasing vocal effort: Fundamental frequency rises with subglottal pressure (e.g., Ladefoged, 1967; Stathopoulos and Sapienza, 1993). In addition, as found first by Schulman (1989), speakers changed the supraglottal articulation, e.g., using lower jaw positions and therefore increased first formant frequencies. These other manipulations can also affect loudness (by focusing spectral energy in regions where the ear is more sensitive).

Many studies have claimed that word stress is also primarily characterized by an increase in loudness or vocal effort. In his pioneering work in 1967, Ladefoged (1967) found that subglottal pressure not only contributed to global paralinguistic vocal effort changes but also to local variations in prominence, namely, word stress. His experiment showed that short contractions of the muscles activated during exhalation, measured by electromyography, cause an increase in subglottal pressure. However, apart from his later replication of this experiment (Ladefoged, 2005), none of the follow-up experiments directly measuring respiratory muscle activity could reproduce a significant stress effect. Marasek (1997) suggested in a modeling study that greater subglottal pressure alone underlies word stress, whereas sentence accent is primarily controlled by vocal fold tension. Indirect evidence for this position was provided in an acoustical study by Sluijter and van Heuven (1996), who found that the spectral slope was flatter in stressed vowels independent of their accentual status (see also Okobi, 2006 for a clear-cut lexical stress distinction in de-accented position). They attributed this difference in “spectral balance” to a different glottal configuration instead of different subglottal pressures.

There is, however, 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 distinctions in acoustic parameter changes for stress and accent were found by Campbell and Beckman (1997) and Hanson (1997a). Furthermore, in a study by Heldner (2003) focal accent was also produced with a flatter spectral slope. The Finnegan *et al.* (2000) study also varied focal accent rather than word stress and still found higher subglottal pressure values for focused words. All these studies suggest that there is no clear physiological distinction between stress and accent.

In order to explore this issue further, changes due to word stress, sentence accent (by varying focus), and changes due to increased vocal effort (by asking a subject to vary loudness) were analyzed by using laryngographic techniques and spectral characteristics of the speech waveform. The assumption was that if word stress changes can be attributed to changes in vocal effort, then a similar pattern of Lx and spectral changes should be found in both word stress and raised loudness, and both are expected to be different from the types of Lx and spectral change that accompany sentence accent. In the literature a number of acoustic and Lx parameters have been identified as being affected by changes in vocal effort, sentence accent, and/or stress. Only the parameters used in the current study will be reviewed here: the Lx

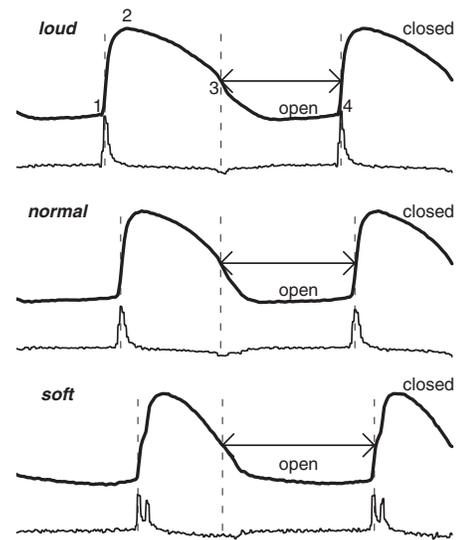


FIG. 1. Examples of Lx signals for loud, normal, and soft speech from speaker M01 at midvowel in stressed and accented position. Data printed as thin lines correspond to the first derivative of the Lx pulses (bold lines). In the upper panel landmarks are denoted by numbers: (1) onset of closing at maximum of the first derivative, (2) maximum contact, (3) moment of glottal opening, determined by a 3/7 threshold, and (4) next glottal closing. The length of the arrows corresponds to the open glottis interval.

parameters open quotient (OQ), skewness quotient (SQ), slopes of glottal closing and opening, as well as the acoustic parameters  $f_0$ , intensity, and the acoustic OQ, denoted as  $H1^*-H2^*$ , and spectral tilt, denoted as  $H1^*-A3^*$ . These will be defined precisely in Secs. I B and I C.

## B. Laryngeal reflexes

Laryngography or electroglottography has been very popular for recording phonatory behavior for the last 3 decades. This popularity can be attributed to the facts that laryngography is a completely non-invasive technique that does not interfere with normal articulation and to the simplicity of handling this instrument. However, since the laryngograph measures the time-varying impedance between the vocal folds and not the glottal area or the airflow, the signal is difficult to interpret which has led to several critical articles (e.g., Colton and Conture, 1990; Holmberg *et al.*, 1995; Sapienza *et al.*, 1998; Titze, 1990).

Figure 1 shows examples of Lx signals from the current study. The thicker lines represent the Lx pulses and the thin lines the first derivative for loud (top), normal (middle), and soft speech (bottom). In this figure the signal decreases with the amount of contact between the folds, i.e., low for no contact [numbers (1) and (4) in the upper signal] and high for closed glottis (2). By comparing endoscopic high speed images with Lx signals, Henrich *et al.* (2004) found that after the maximum (2) the glottis continues to be closed despite the decrease between (2) and (3). The moment of glottal closing (1) is usually labeled at the maximum of the derivative and most of the times is unambiguous and clearly defined. There are, however, occasionally cases with two peaks in the derivative, as shown in the signal derived from soft speech (bottom of Fig. 1) which are related to discontinuities in the glottal closing (Henrich *et al.*, 2004). Even more problematic is the detection of the moment of glottal opening

(e.g., Childers *et al.*, 1990). Whereas in airflow signals, glottal opening affects the measurements precisely and immediately, the change in Lx signals is much more gradual. Since clear negative peaks, indicating the glottal opening, are rarely found in the derivative (see Fig. 1), various thresholds are used instead. In Fig. 1, a threshold of 3/7 was applied as recommended by Howard *et al.* (1990) and Henrich *et al.* (2004). This difficulty comes about because mucous bridging and vertical phase differences between the lower and upper edges of the vocal folds often make the instance of glottal opening undetectable (see Sapienza *et al.*, 1998; Childers *et al.*, 1990; Colton and Conture, 1990).

Probably because of these problems, several parameters derived from airflow signals that showed variation with linguistic and non-linguistic prominence did not change when derived from the Lx signal. First, it will be considered how **vocal effort** affects different vibratory characteristics and their Lx correlates. The most obvious change due to vocal effort variation concerns the interval during which the glottis is open. The open quotient (ratio of the duration of open glottis to the entire period, henceforth OQ) derived from airflow data generally showed a significant decrease when speaking louder (e.g., Dromey *et al.*, 1992; Holmberg *et al.*, 1988 only for male speakers; Stathopoulos and Sapienza 1993; Sapienza *et al.*, 1998), because the open glottis interval decreases for loud speech. As shown in Fig. 1 by the length of the horizontal arrows, the no-contact interval in the Lx signal clearly increases going from loud to normal to soft speech. However, most studies using Lx data could not replicate this effect (Dromey *et al.* 1992; Sapienza *et al.*, 1998), presumably due to the noise introduced by the uncertainty for detecting the instant of glottal opening. Only Henrich *et al.* (2004) found a clear correlation between intensity and OQ based on the Lx signal, but this was for trained singers.

A second important characteristic varying with vocal effort is the symmetry of the pulse because the glottis closes more quickly at higher levels of vocal effort, which causes a less symmetric and more left-skewed pulse in the airflow signals. The symmetry of the glottal pulse has been quantified by the skewness or speed quotient SQ as the interval of the closing phase (see the raising slope of the pulse in Fig. 1) in relation to the opening phase (the falling slope of the pulse). Airflow SQ usually decreases for a higher vocal effort, indicating a more skewed pulse because of a faster glottal closing movement (Dromey *et al.*, 1992; Holmberg *et al.*, 1988; Sapienza and Stathopoulos, 1998). As for the OQ, this could not, however, be replicated with Lx data (see Dromey *et al.*, 1992; Sapienza *et al.*, 1998, who observed no significant changes for OQ and SQ based on Lx signals). It can be seen in Fig. 1 that Lx pulses are generally not symmetric but left skewed with a very steep raising slope which seems to limit the sensitivity to further changes. Apart from the OQ and SQ, the slopes of the closing and opening phases or the closing peak of the first derivative are also often successfully used to quantify the abruptness of the closure. To summarize, speaking louder decreases the OQ and increases the SQ, but consistent results could only be found for quotients based on airflow data.

The shape and time course of glottal vibrations might

also be affected by prosodic variation. Following Beckman (1986) and others, Sluijter and van Heuven (1996) promoted the view that sentence accent and word stress are produced with different laryngeal mechanisms: lexical stress by skewing the glottal pulse and sentence accent by increasing the rate of pulses and thereby producing a higher pitch. Up to now, many acoustic studies have addressed this issue (see Sec. I C) but only a very few laryngographic and airflow investigations exist. To our knowledge, no airflow data are available for **word stress**, but in a study by Marasek (1996) on German, word stress (confounded with sentence accent) had a significant effect on the steepness of the closing and opening slopes in the Lx signal with steeper closing and shallower opening slopes. However, neither the OQ nor the SQ was affected.

A higher  $f_0$  increased the airflow OQ for global  $f_0$  changes (e.g., Holmberg *et al.*, 1989), singing (Henrich *et al.*, 2004), and also for **pitch accents** (Pierrehumbert, 1997), probably because when the vocal folds are stiffer they close only at the outer edges, as a result of which the closed phase is shortened (e.g., Titze, 1992). The airflow-based SQ also increases with  $f_0$ , indicating a more symmetrical pulse at higher  $f_0$ , in global tone changes (Holmberg *et al.*, 1989) and for more local pitch accents (Pierrehumbert, 1997). With Lx data, however, only the effects on the OQ could be confirmed (Marasek, 1997 for pitch accents in German, and Gendrot, 2003 for French for focalized vs non-focalized vowels) whereas SQ based on Lx data did not vary systematically with  $f_0$ .

As mentioned above the discrepancy between measures based on airflow data and measures based on Lx data can probably be attributed to the problems with defining meaningful landmarks, especially the moment of glottal opening. In order to overcome these well-known difficulties with labeling landmarks in Lx signals, a more holistic approach was pursued in the current study by analyzing the shape of the Lx pulse as a whole. A similar approach was adopted by Mokhtari *et al.* (2003) who applied a principal component analysis (PCA) to the inverse filtered speech signals of Laver's (1980) recordings of several phonation types. The resulting components discriminated between a wide range of voice qualities. In the present study, function data analysis (henceforth FDA) was used to calculate functional versions of splines of time and amplitude normalized Lx pulses (for an introduction to FDA see, e.g., Ramsay and Silverman, 1997). After computing the spline functions, a PCA was applied to the data (for further details see Sec. II). The prediction was that Lx pulses taken from stressed syllables would exhibit a similar shape to pulses from loud speech and Lx pulses from unstressed syllables would pattern with softly spoken items.

### C. Acoustic reflexes

As shown above, very few studies looked at effects of prosodic variation on voice quality using Lx or airflow data. There are, however, many studies addressing this question by means of acoustic data. As mentioned above, the vocal folds close more rapidly with rises of subglottal pressure or changes of the laryngeal settings. This change is not well

captured by the acoustic measure of intensity because it is highly dependent on the distance between the speaker's mouth and the microphone. The more abrupt glottal closure boosts the energy of the higher harmonics in the frequency region above 2000 Hz and therefore decreases the spectral slope. Due to this non-uniform effect on the spectrum, most acoustic studies on word stress or vocal effort are based on a measure of the power spectrum's slope. There are several different measures for quantifying this change, e.g., the balance between the sum of the amplitudes within certain frequency bands ("spectral balance," [Sluijter and van Heuven, 1996](#)), the difference between the overall intensity and the intensity of a low-pass filtered signal ("spectral emphasis," e.g., [Traunmüller and Eriksson, 2000](#); [Heldner, 2003](#)) and the difference between the amplitude of the first harmonic and the third formant H1-A3, also termed rate of closure (see also [Fig. 2](#), discussed further in [Sec. II](#)). Lower values for this latter measure indicate a flatter slope with more energy in the frequency region of the third formant. [Holmberg et al. \(1995\)](#) demonstrated for 20 female speakers that loud speech was produced with a lower H1-A3 than modal speech. A similar result would be expected in comparing stressed with unstressed vowels. A severe problem for all of these measures is that they reflect not only changes in the glottal source but are also affected by the vocal-tract resonances, that is, these measures vary with vowel quality. In order to compensate for these effects, a number of correction algorithms have been suggested. For example [Hanson \(1997b\)](#) and [Hanson and Chuang \(1999\)](#) corrected the amplitude of the first harmonic H1 by the location of the first formant frequency F1 and the amplitude of the F3 by the first two formant frequencies of the neutral vowel. However, as shown convincingly by [Iseli et al. \(2007\)](#), these algorithms can be improved by also correcting for the effect of bandwidths of the formants. Corrected amplitudes are usually denoted by an asterisk, e.g., H1\* and A3\*.

With respect to the conditions investigated in this study, i.e., vocal effort, word stress, and focal accent, we would expect that H1\*-A3\* should progressively increase from loud to normal to soft speech. Stressed vowels should exhibit lower values than unstressed vowels (as confirmed in [Sluijter and van Heuven, 1996](#) for Dutch, although there were contradictory results in [Claßen et al., 1998](#) for German; [Campbell and Beckman, 1997](#); and [Hanson, 1997a](#) for English), and sentence accent should not affect the slope of the spectrum (but see [Heldner, 2003](#) for significant differences in spectral slope for focus variation in Swedish).

Speaking more loudly also increases the portion of the glottal cycle during which the glottis is closed because glottal closure is achieved more rapidly and more completely yielding a lower OQ of airflow. Differences in loudness and breathiness have also been found to be associated with amplitude differences between H1\* and H2\*. Louder and more modal voices are produced with lower values of the acoustic OQ, whereas softer and more breathy voices show increased values ([Holmberg et al., 1995](#)). Again, this measure is strongly affected by the vocal-tract resonances, especially the position of F1. Therefore, we applied the correction algorithm by [Iseli et al. \(2007\)](#), which has the advantage that it

continues to work if F1 approaches the second harmonic. Algorithms based on a correction for the F1 frequency (e.g., [Hanson, 1997b](#)) produce invalid data in these cases. Prior studies showed no significant effect for vocal effort ([Holmberg et al., 1995](#)), word stress, and pitch accent ([Sluijter, 1995](#); [Okobi, 2006](#) for English; [Claßen et al., 1998](#)), probably due to the prominent effect of vowel quality, even on corrected values (see [Iseli et al., 2007](#)). However, it has been found that accented vowels are produced with less breathiness than unaccented vowels ([Choi et al., 2005](#)). Since we wanted to compare the acoustic to the Lx OQ we included the acoustic OQ in our measurements.

## D. Aims

To summarize so far, the first aim of this study is to compare effects of vocal effort, word stress, and sentence accent on the acoustic measures  $f_0$ , H1\*-H2\*, and H1\*-A3\* and the Lx measures OQ, SQ, closing slope, opening slope, and FDA shape parameters. The hypothesis is that changes due to vocal effort and word stress affect the investigated parameters in a similar way whereas focus is produced mainly by changes in  $f_0$ . In particular, the goal is to identify a set of parameters that distinguishes stressed from unstressed syllables independently of sentence accent. Earlier studies often failed because they confounded stress and accent. In this current study stress and focal accent are varied orthogonally, and, additionally, vocal effort is introduced for comparison. Consequently, stress could be seen as localized vocal effort change. Variation in sentence accent should not have an effect on the same parameters as stress and vocal effort. If sentence accent affects mainly stressed syllables, then parameters changed by accent variation should show significant differences between stressed focused items on the one hand and all others (stressed unfocused, unstressed focused, and unstressed unfocused). From the literature we expect  $f_0$ , H1\*-H2\*, and OQ to vary in this way. Second, since the FDA shape analysis has not been applied to Lx data, the second aim of this study is to test whether FDA shape parameters are better suited for analyzing Lx pulses than conventional measures based on landmarks. Third, we are interested in the relationship between this new measure, the more conventional quotients, and the acoustic outcomes.

## II. METHOD

### A. Material and subjects

The test words were two-syllable words in German that varied in the position of the primary stress; for example, the words "Lena" (a woman's name) 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 /le/ 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 or an unaccented production because the information was already known [U]. The unfocused condition was constructed in such a way that no  $f_0$  movement during the test syllable was to be expected. Ques-

tions were pre-recorded and presented via headphones. The answers were presented on a computer screen. In order to get consistent realizations, the words that should be focused were printed in upper case letters. Examples of questions and answers are given below:

- (a) Focussed and strong: [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) Unfocused and strong: [U, s]  
 Q: Wie findest Du Lena? (*How do you feel about 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's?*)  
 A: Ich kaufe Lenor bei Schlecker. (*I buy Lenor at Schlecker*)
- (d) Unfocused 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 focal accent  $\times$  lexical stress combinations: [F, s], [F, w], [U, s], and [U, w], which were spoken at a comfortable vocal effort level (henceforth N). The terms *strong* and *weak* are chosen here in order to avoid confusion of *unstressed* with *unfocused*. A further condition was that all of the [F, s] combinations were produced by the speakers in either a loud (L) or a soft voice (S). In the loud condition, speakers were instructed to speak loudly without shouting; for the soft conditions, the instruction was to speak softly without whispering. In order to ensure consistent loudness levels the questions that were presented over headphones were pre-recorded in the three different loudness conditions. These six (lexical stress  $\times$  focus in normal loudness+soft and loud levels) possible combinations were repeated nine times for six of the speakers and eight times for the first speaker of this experiment. The items were in randomized order together with a second set of test items, which is not presented here.

Acoustic and Lx signals at a sampling rate of 16 kHz were obtained from seven male subjects between 20 and 35, speaking a northern variety of Standard German. Male speakers were preferred since due to their smaller thyroid angle and longer vocal folds Lx signals are more reliable (e.g., Colton and Conture, 1990) and since female speakers generally speak with a breathier voice quality which reduces the amplitude of the Lx signal (e.g., Holmberg et al., 1995). A dynamic stand microphone (Sennheiser MD 421) was positioned at a distance of 50 cm from the speaker's mouth at an angle of about 45°. Since intensity is also measured in this experiment we instructed the speakers not to move their heads. For a 50 cm mouth-microphone distance a change of  $\pm 5$  cm corresponds to +1.82/−1.66 dB for frequencies up to 117 Hz and +0.91/−0.83 dB for frequencies above.

All sentences were labeled according to the GToBI conventions (Grice et al., 2005). Six items of the strong unfocused condition had to be excluded because speakers 4 and 5 failed to deaccent 2 and 4 items, respectively. All other items in the unfocused condition had a prominent pitch accent on the word *hasse*, preceding the test word. Some items had to

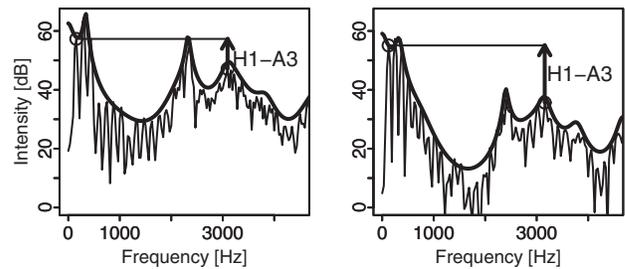


FIG. 2. Calculation of uncorrected spectral tilt H1-A3 for normal vocal effort (left panel) and soft vocal effort (right panel), based on data from speaker M01 during the vowel /e/ in strong and focused position. The solid bold lines display the LPC spectra, and the solid thin lines the narrow band DFT spectra. The length of the arrows corresponds to the difference between the amplitudes of H1 and F3.

be excluded because of mispronunciation (2) or because the Lx signal was distorted due to excessive vertical larynx movements during the test syllable (3). Altogether 361 items were analyzed.

## B. Measurements

*Acoustic measurements.* From the speech signal the rms energy,  $f_0$ , and formant frequencies for the first three formants were measured at the acoustical temporal midpoint of the vowel /e:/ from all the test words. The temporal midpoint was chosen in order to minimize coarticulatory effects of the consonant context. However, due to vowel reduction in unstressed position, the consonant context might affect weak vowels to a greater degree than strong ones (see Mooshammer and Geng, 2008 for German). The rms signal was calculated with a Hamming window of 50 ms and shifted by 5 ms. Formants were estimated by using 16 linear predictive coding (LPC) coefficients and a 25 ms Hamming window with 5 ms shift. For some items the number of coefficients had to be adjusted. The formant frequencies and  $f_0$  were used as reference values for the correction algorithms in order to calculate the acoustic OQ and the spectral tilt discussed below.

The acoustic OQ H1-H2 and the spectral tilt H1-A3 were obtained by applying the following steps: For the vowel interval two different kinds of spectra were calculated, a narrow-band discrete Fourier transform (DFT) spectrum with a frequency resolution of 40 Hz, a Hamming window of 32 ms and shifted by 5 ms, and a LPC spectrum with 22 coefficients and a pre-emphasis of −0.97. The frequencies and amplitudes of the first and second harmonics were detected by means of a peak-picking algorithm. The amplitude of the third formant, A3, was measured at the harmonics that was closest to the third peak in the LPC spectrum. Examples of the measurements before correction for changes in formant values are given in Fig. 2 for speaker M01 in normal speech (upper panel) and soft speech (lower panel). In order to compensate for the formant changes due to modifications of articulatory positions during the vowels, H1\*-H2\* and H1\*-A3\* were corrected by the procedure suggested by Iseli et al. (2007). Their approach is especially useful for correcting H1 of vowels with low frequency values of F1. Since in

TABLE I. Summary of analyzed parameters derived from the audio signal (upper part) and the laryngographic data (lower part).

Acoustic parameters	Description	Prediction for strong vs weak items	Prediction for focussed vs unfocused items
H1*-H2*	Acoustic open quotient	Lower	Higher
H1*-A3*	Spectral tilt	Lower	-
Lx parameters			
OQ	Open quotient	Lower	Higher
SQ	Skewness quotient	-	-
CSlope	Closing slope	Steeper (more positive)	-
OSlope	Opening slope	Steeper (more negative)	-

this study items with German /e/ with relatively low F1 values are analyzed, the correction suggested by Iseli *et al.* (2007) was applied:

$$H^*(\omega_0) = H(\omega_0) - \sum_{i=1}^N 10 \log_{10} \times \frac{(1 - 2r_i \cos(\omega_i) + r_i^2)^2}{(1 - 2r_i \cos(\omega_0 + \omega_i) + r_i^2)(1 - 2r_i \cos(\omega_0 - \omega_i) + r_i^2)}, \quad (1)$$

where  $r_i = \exp(-\pi B_i / F_s)$  and  $\omega_i = 2\pi F_i / F_s$ . The variables  $B_i$  and  $F_i$  are the bandwidths and frequency of the  $i$ th formant,  $F_s$  is the sampling rate, and  $N$  is the number of formants to be corrected for. In our case the amplitude of the first harmonic  $H(\omega_0)$  is corrected by the first two formant frequencies yielding the corrected amplitude  $H^*(\omega_0)$ . As shown in Table I, a smaller spectral tilt value would be expected for stressed than for unstressed vowels. Likewise, smaller spectral tilt would be expected for loud than normal and normal than soft vocal effort levels. The acoustic OQ has been found to vary with sentence accent. Accordingly, a higher value for H1\*-H2\* is expected for unaccented items and also for sentences spoken in a soft voice because of an increase in breathiness.

*Laryngographic measurements.* From the Lx signal the two medial pitch periods during the vowel were extracted. As beginning and end of the extracted Lx pulses, the predetermined 3/7 threshold of glottal opening was used (see Howard *et al.*, 1990). Since we were not interested in shape differences induced only by different period lengths, the two pulses of all items were linearly time-normalized to a uniform length of 1000 samples by linear interpolation.

Since the Lx signal cannot be calibrated, the data were also amplitude normalized to an amplitude of 1 for the first glottal closure. In order to compensate for vertical larynx movements a line connecting the minima of the first and

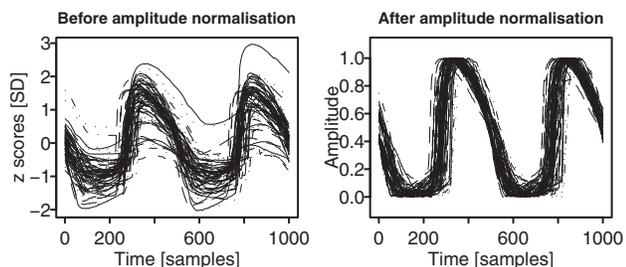


FIG. 3. Lx pulses after time normalization (left panel) and amplitude normalized pulses (right panel). For reasons of visibility only every fourth token is shown.

second periods was subtracted from all values. This effect of this amplitude normalization is illustrated in Fig. 3 showing the Lx pulses of about a quarter of all trials from all speakers. The left panel depicts the time-normalized  $z$  scores and the right panel the amplitude normalized pulses after subtraction of the line between the two minima. As can be seen the minima and maxima of the amplitude of the normalized data are aligned after this procedure and the amplitude induced variability is much smaller compared to the upper panel.

The following Lx parameters were computed for the medial pitch period of /e/ in all test conditions using EMU/R (Bombien *et al.*, 2006):

- (1) For calculating the OQ, the 3/7 threshold was used as the instant of glottal opening (as suggested by Howard *et al.*, 1990) as well as the peak in the first derivative as the instant of glottal closing (see Fig. 4, left). The OQ was then calculated as the percentage of the open glottis interval to the pitch period duration.
- (2) The SQ [using a 10% threshold as suggested by Marasek (1997), see Fig. 4, right] was computed as the ratio between the closing and the opening duration. This value decreases with a quicker closing movement.
- (3) The slope of glottal adduction was also computed as the quotient between the amplitude of the closing movement and its duration, both defined with a 10% threshold.
- (4) Similarly, the slope of the opening movement was defined as the quotient between the amplitude and the duration of the opening movement.

Table I shows the expectations on the way the laryngographic parameters should change based on simulations by Marasek (1997) and results from the literature.

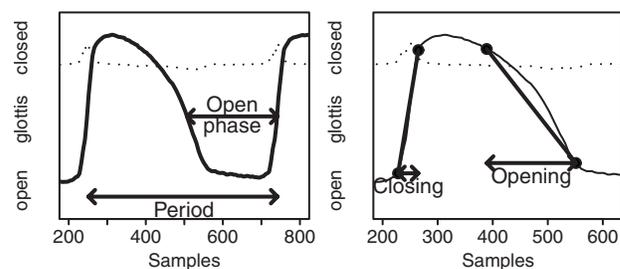


FIG. 4. Left panel: measurements of the open phase and the period duration for calculating the OQ, right panel: measurements of the closing and the opening movement, for calculating the SQ, bold lines indicate the closing and the opening slope. Since data were normalized, the time axis is in arbitrary samples as well as the y-axis in arbitrary amplitude units. Solid lines: normalized glottal pulses; dotted line: first derivative in arbitrary units.

In order to avoid the well-known difficulty with detecting specific landmarks in the Lx signal (see Sec. I), we analyzed the whole shape of the Lx pulse by means of the functional version of the principal component analysis (FPCA) using the R package FDA version 1.2.4 (for further details and formulas, see Ramsay and Silverman, 1997, 2002). Basis functions for the pre-processed medial two Lx pulses were computed by using the Fourier basis functions. Fourier basis functions are recommended for periodic data and involve the calculation of coefficients for the sine components of the waveform: in the current case the number of coefficients was set to 200. This number was necessary because a lower order modified the shape of the Lx pulses too much toward a sinusoidal wave. Smoothing of the resulting curves was obtained by a roughness penalty of the third-order time derivative with the smoothing parameter  $\lambda = 10^{-12}$ . The order of 200 and the smoothing values were determined by visual inspection of the results and used to ensure that important details of the original data are captured by the basis functions. In many studies FDA time registration is applied to the data prior to further analysis (e.g., Lee *et al.* 2006; Lucero and Koenig, 2000; Lucero and Löfqvist, 2005) for nonlinear time-warping. However, only the above-mentioned linear amplitude and time normalization were applied to the Lx pulses, because we were interested in the skewness of the Lx pulse and the steepness of the closing movement which also implies a shift of the maximum relative to the minimum. Dynamic time-warping tends to obscure these shape characteristics.

In order to identify the main sources of variability in the Lx pulses, a FPCA was applied after calculating the Fourier basis functions (Ramsay and Silverman 1997, 2002). In the case of Fourier basis functions, the FPCA is carried out on the covariance matrix of the smoothed Fourier coefficients. The resulting principal component weight functions are defined over the same range of time as the Fourier functions. Some smoothing was applied to the second derivative and only the first two factors were considered here. The resulting factor scores of the first two components PC1 and PC2 indicate at which time-stretches the Lx pulses show the largest variation as well as the extent of variation. The advantage of the PCA is that only two parameter values, the two factor scores, are needed for describing the range of Lx pulse shapes that occur. Standard statistics as described below were then carried out with these factor scores as dependent variables.

### C. Statistics

The cell means of the acoustic measures H1\*-H2\* and H1\*-A3\*, the Lx parameters SQ, OQ, closing and opening slopes, and the derived factor scores were analyzed by repeated measures analysis of variance (ANOVA) with speaker as a random factor and 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 focused items, two separate analyses were carried out with subsets of

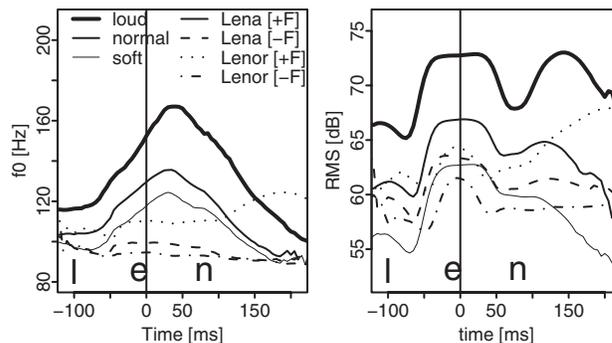


FIG. 5. Ensemble averages of  $f_0$  contours (left panel) and rms contours (right panel) during the test words *Lena* (initial stress) and *Lenor* (final stress). The vertical line indicates the mid-vowel measurement time-point. All contour are aligned to this time-point.

data: for linguistic prominence only data spoken in normal volume were used and tested for the two within-subject factors stress (levels: strong and weak) and focal accent (levels: focused and unfocused). For effects of vocal effort, only data from focused and stressed vowels were taken into account. The within-subject factor vocal effort had the levels loud, normal and soft. For testing speaker consistency, one-way ANOVAs were calculated for the two different data sets split by speaker.

## III. RESULTS

### A. Acoustic parameters

Before analyzing several parameters at the midpoint of the vowel, it was verified that the speakers realized the conditions as expected. Therefore, ensemble averages of  $f_0$  and rms tracks for the entire test word were calculated. Figure 5 shows the averaged  $f_0$  contours in the left panel and the rms contours in the right. The contours are aligned at the midpoint of the vowel, indicated by the vertical lines. This is approximately the time-point at which measurements were taken for the analyses in Secs. III B–III D. Since the averaged items had different lengths, the averages further away from the midpoint are less reliable and therefore more “bumpy.” Solid lines show variation with vocal effort with linewidth decreasing with effort, i.e., thick lines for loud speech and thin lines for soft speech and normal in-between.

As can be seen in the left figure, the items with varying vocal effort levels, which are all stressed on the first syllable and are focused, are produced with a steep rise on the first syllable followed by a fall toward a phrasal low tone. The range of  $f_0$  variation was largest for the loud condition ( $\bar{x} = 67$  Hz,  $s = 14$  Hz), in-between for normal ( $\bar{x} = 44$ ,  $s = 12$ ), and smallest in the soft condition ( $\bar{x} = 38$  Hz,  $s = 11$  Hz). This order was consistent for all speakers. The increase in  $f_0$  with vocal effort is in agreement with findings from the literature (e.g., Ladefoged, 1967). Within the normal condition the  $f_0$  range was largest for stress on the initial syllable (i.e., *Lena*) in the focused condition ( $\bar{x} = 44$ ,  $s = 12$ ). Pitch accents on the second syllable in *Lenor* in the focused condition were less extensive (dotted line in Fig. 5,  $\bar{x} = 32$  Hz,  $s = 9$  Hz) and the whole word was produced with a somewhat higher  $f_0$  as compared to the unfocused items (see dashed and dashed dotted lines in Fig. 5). Only the focused word with initial

TABLE II. Results of repeated measures ANOVA and pairwise t-tests for the dependent acoustic variables  $H1^*-H2^*$  and  $H1^*-A3^*$ , the Lx variables OQ, SQ, CSlope, and OpSlope, and the factors vocal effort, stress, focal accent, and interactions between stress and accent. Greenhouse–Geisser corrected degrees of freedom are given for vocal effort, for stress, accent, and the interactions the degrees of freedom are always 1 and 6. Significant effects ( $p < 0.05$ ) are printed in bold.

Variable		Vocal effort	Stress	Focus	Interaction
$H1^*-H2^*$	F	1.3	<b>6.9</b>	2.5	0.2
	df	1.8, 10.9	$s < w$		
$H1^*-A3^*$	F	<b>11.8</b>	1.1	2.2	1.1
	df	1.2, 7.1	L < S		
OQ	F	<b>12.7</b>	<b>12.8</b>	>0.1	2.1
	df	1.3, 7.6	L < S	$s < w$	
SQ	F	0.97	0.6	<b>6.6</b>	0.8
	df	1.1, 6.8		F > U	
CSlope	F	0.63	0.5	>0.1	0.4
	df	1.6, 9.7			
OpSlope	F	0.32	<b>8.0</b>	0.3	2.4
	df	1, 6.5	$s > w$		

stress showed a prominent  $f_0$  movement in the vicinity of the measurement point; for the other three conditions the contour remained flat. This was consistent for all speakers after excluding some of the items (see Sec. II A.)

The right panel in Fig. 5 presents the rms contours. A comparison between the solid lines gives clear evidence that, overall, the speakers managed to produce the items with the required levels of vocal effort with a mean difference over the test items between loud and normal of 6 dB and between normal and soft of 5 dB. The minimal difference was 3 dB, produced between soft and normal speech by two speakers. Prosodic variation also affected the rms contour with weak and unfocused items having a similar rms as items spoken with a soft volume. This can be attributed to the fact that rms rises with  $f_0$ . Therefore, the focused items, produced in a soft volume, had a higher rms during the  $f_0$  rise during the initial syllable, whereas during /l/ and for the final vowel rms was lower for the soft items than for all others. From these considerations, it can be concluded that the elicitation technique applied in this study was successful because speakers distinguished consistently between the three volume levels and generally produced the predicted  $f_0$  contours.

Table II gives the results of repeated measures ANOVAs for vocal effort and linguistic prominence (stress and focus) as independent variables and the dependent variables  $H1^*-H2^*$  and  $H1^*-A3^*$ , calculated at the vowel midpoint for the seven male speakers. Figure 6 shows the mean values and standard errors for the vocal effort and prosodic conditions. The grayscale of the bars denote our expectations on prominence measures, i.e., measures relevant for the stress distinction should pattern with vocal effort. Accordingly, darkness decreases in the following order: loud speech, stressed items in comfortable loudness condition, unstressed items in comfortable loudness condition and soft speech.

The acoustic OQ  $H1^*-H2^*$ , shown in the left panel of Fig. 6, was only affected by stress. On average, strong items had only a 1 dB lower acoustic OQ than weak items. Acoustic OQ was not significantly affected by focus or vocal effort.

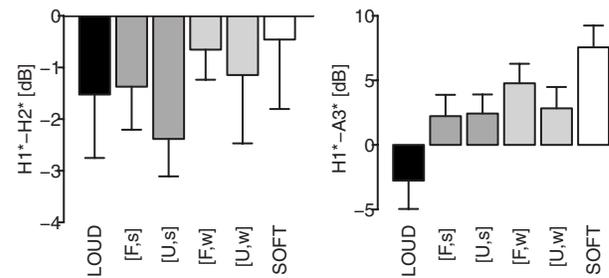


FIG. 6. Means and standard errors for the dependent acoustic variables  $H1^*-H2^*$  and  $H1^*-A3^*$  for vocal effort and linguistic prominence variations. F: focused; U: unfocused; s: strong (first syllable in *Lena*); w: weak (first syllable in *Lenor*).

One possible reason for the latter results might be the low F1 which was always close to the second harmonic for the vowel /e/. Furthermore, the correction algorithm reduced the OQ of the pitch-accented items by 6.3 dB (and most for the loud items) and items with a flat low  $f_0$  contour only by 3 dB. This non-uniform correction can be attributed to the fact that for the pitch-accented items  $f_0$  approaches the low first formant in German /e/. Therefore, the amplitude of  $f_0$ , i.e., H1, is boosted by the first vocal-tract resonance. Since the *Iseli et al. (2007)* algorithm includes the bandwidth, it corrects for these  $f_0$  changes. This explains why no focus and vocal effort effects were found for the acoustic OQ as would have been expected from the literature.

Spectral tilt  $H1^*-A3^*$  (see right panel in Fig. 6) was affected only by vocal effort with substantially lower values of 10 dB for loud vs soft levels. This effect was significant for six of the seven speakers. Contrary to our expectations, stress did not affect this value significantly; however, for the focused items at least, the differences were in the right direction, i.e., spectral tilt for strong vowels was more similar to loud speech and for focused weak vowels [F, w] spectral tilt was more similar to soft speech. The applied correction algorithm reduced the amplitude difference between H1 and A3 by about 8 dB without obvious differential effects for different items.

## B. Laryngographic measures

Results for the four Lx measures OQ, SQ, closing slope, and opening slope are given in Table II and Fig. 7. Table II indicates that OQ was significantly affected by vocal effort with the OQ for loud speech 6.6% smaller than for soft speech. There was a significant but smaller effect of stress: OQ was lower by 2.5% for strong items than for weak. Therefore, the glottis was closed for a longer duration for strong than for weak items and for loud than for soft items, which confirms the predictions stated in Table I.

Focus did not affect OQ, but it did affect SQ, resulting in significantly higher values (an average increase of 1.9%) for focused as compared to unfocused items. This means that Lx pulses in focused syllables were produced with a more symmetrical pulse, which is in agreement with data derived from aerodynamic measurements (*Holmberg, et al. 1989, Pierrehumbert, 1997*). According to *Marasek's (1997)* study, the slopes of the closing and opening movement should vary with stress. CISlope showed no significant differences for

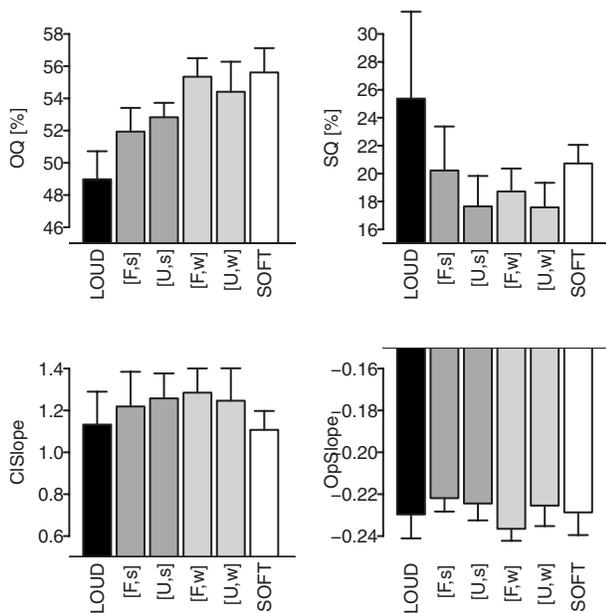


FIG. 7. Means and standard errors for the dependent Lx variables OQ, SQ, CISlope, and OpSlope for vocal effort and prominence variations. F: focused; U: unfocused; s: strong (first syllable in *Lena*); w: weak (first syllable in *Lenor*).

linguistic prominence or loudness. The steepness of glottal opening varied significantly for stress with steeper (negative) opening slopes for weak items. This result was contrary to our hypothesis, but as can be seen from Fig. 7 the range of variation for this parameter was very small ( $-0.28$  to  $-0.19$ ) and probably not very meaningful. Generally, the landmark-based parameters SQ, CISlope, and OpSlope were not consistently affected by stress and vocal effort. Only the OQ varied in the expected direction, i.e., the OQ of weak syllables resembled the OQ of softly spoken syllables, and there was a similar relationship between strong and loud syllables.

### C. Results from shape analysis of Lx pulses

Results of the FPCA are presented in Fig. 8. The upper two panels show the shapes of the Lx pulse for negative factor values indicated by minus signs, positive factor values by plus signs, and the mean curve by a dashed line. Shape differences as shown for the PC1 (left) and PC2 (right) explain about 71% and 17%, respectively, of the variance in Lx pulse shapes. Positive values of the first factor were characterized by a longer open phase (as can be seen from the downward hump before the glottis closes) compared to the mean curve. Negative factor values had a shorter open phase and a higher rise. The second factor showed variation in the opening slope, with curves for negative values exhibiting an earlier opening with a steeper opening slope and therefore a longer open phase and a more symmetrical Lx pulse. In gross terms 71% of the variation in Lx pulses can be attributed to the closing movement and the shape of the open phase and 17% to the opening movements. These changes, a later closing movement for positive values of factor 1 and an earlier opening for negative values of factor 2, increased the duration of the open glottis.

In the two panels below, mean and standard errors of PC1 and PC2 are displayed. The higher these values are, the

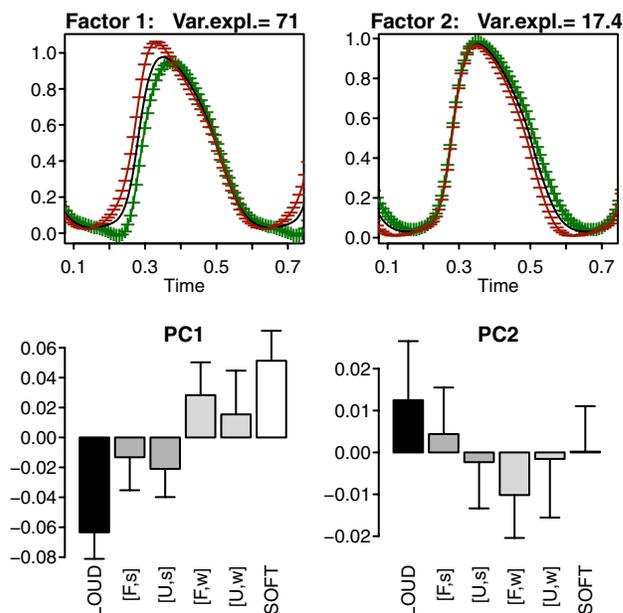


FIG. 8. (Color online) Upper panels: the shapes of the Lx pulse for negative factor values indicated by minus signs, positive factor values by plus signs, and the mean curve by a solid line for the first factor (left) and the second factor (right). Lower panels: means and standard errors of the factor scores for vocal effort and prominence variations. F: focused; U: unfocused; s: strong (first syllable in *Lena*); w: weak (first syllable in *Lenor*).

greater the approximation to a “positive” Lx pulse shape. Hence, for the first factor, Lx pulses extracted from loud speech exhibited a “negative” pulse shape with a short open phase and a steep and short closing phase. Soft speech was more similar to the positive Lx pulse and involves a longer open phase. According to our predictions, Lx pulses for strong items should pattern with loud speech and for weak items with soft speech. Scores of the first principal component, shown in Fig. 8, and results from the repeated measures ANOVA in Table III, confirmed this prediction. PC1 differed significantly, both between loud and soft items and also between strong and weak items. One-way ANOVAs split by subjects showed significant stress effects for five of the seven speakers and significant vocal effort effects for six speakers.

For the second factor, there seemed to be a tendency for items bearing a pitch accent to pattern together with slower glottal opening and shorter open phases. The unfocused items and the focused, weak items with a flat  $f_0$  contour were produced with an earlier opening and a longer open phase. This was reflected in a significant interaction between

TABLE III. Results of repeated measures ANOVA and pairwise t-tests for PC1 and PC2 and the factors vocal effort, stress, focal accent, and interaction between stress and focus. The variances explained by the first two factors are reported in the second column. For vocal effort the degrees of freedom are adjusted for sphericity violations, in all other cases the degrees of freedom are 1 and 6. Significant  $F$  values ( $p < 0.05$ ) are printed in bold.

PC	Var. expl.	Vocal effort	Stress	Focus	F × S
1	71.0	<b>11.7</b> 1.3 7.9 L < S	<b>14.3</b> s < w	1.0	0.1
2	17.7	2.6 1.2 7.2	5.7	0.5	<b>9.7</b>

TABLE IV. Correlation coefficients between factor scores and the glottal shape and acoustic parameters. Coefficients based on cell means were calculated for two subsets of data: for prominence the soft and loud items were excluded; for vocal effort the unfocused and the weak items were excluded. Only significant coefficients ( $p < 0.05$ ) are shown here. Highly significant coefficients are printed in bold.

	PC1 prominence	PC1 vocal effort	PC2 prominence	PC2 vocal effort
OQ	<b>0.67</b>	<b>0.81</b>	<b>-0.71</b>	<b>-0.77</b>
SQ			<b>0.61</b>	<b>0.51</b>
CISlope-			<b>0.68</b>	<b>-0.68</b>
OpSlope	-0.56		<b>0.61</b>	<b>0.47</b>
$f_0$		-0.51		
rms-		<b>0.61</b>		
H1*-H2*			<b>-0.64</b>	<b>-0.75</b>
H1*-A3*		<b>0.41</b>		

stress and focus. Pairwise  $t$ -tests, however, did not reach significance. The scores of PC2 were also not affected by loudness. Single speaker one-way ANOVAs were also highly inconsistent for the second factor.

In order to understand better the meaning of the factor scores, Pearson correlation coefficients were calculated for two subsets of the data (see Table IV): the prominence subset consists only of items in the normal loudness condition. The vocal effort subset contained data from strong and focused items in the three loudness conditions. For the prominence subset, only the OQ and the opening slope were significantly correlated with PC1. For the vocal effort variation, PC1 was correlated with OQ and the acoustic measures  $f_0$ , rms, and H1\*-A3\*. The second principal component PC2 was significantly correlated with all of the parameters derived from the Lx signal and also with H1\*-H2\*. There were only some minor differences in the extent of the correlation coefficients for the two different data subsets, but not in direction. However, apart from the OQ, variables significantly related to PC2 did not contribute to distinguishing different levels of prominence or vocal effort. Therefore, the shape differences, captured by SQ, CISlope, and OpSlope, and most of the variation in the second factor seemed to be irrelevant for the conditions varied in the current study.

#### D. Summary

In this study Lx pulse shapes, derived from Lx data, their parametrizations and acoustically derived parameters were compared for three levels of vocal effort, two levels of focal accent, and two levels of lexical stress. Based on the literature, it was predicted that strong items should be produced with more vocal effort than weak items.

The following parameters showed the same tendencies for lexical word stress and vocal effort: the Lx OQ and PC1. For vocal effort changes, there was a decrease in OQ from soft to loud speech. Word stress affected the OQ in a similar way, i.e., strong items had a lower OQ than weak. The OQ of the strong items was also more similar to the OQ of loud speech, and weak items resembled soft speech on this measure. This was also reflected by the global shape parameter PC1 (factor scores of the first principal component) that dis-

tinguished Lx pulse shapes of loud from soft items and strong from weak items. Again, the strong items resembled the loud condition and the weak items resembled the soft condition. Vocal effort and stress also affected  $f_0$  and rms in the same directions. However, these two parameters were also affected by focus. The acoustic parameter H1\*-A3\* was influenced by vocal effort but not by stress. The opposite was the case for the acoustic OQ (defined as H1\*-H2\*) and OpSlope which varied with stress but not with vocal effort.

Focus affected SQ with higher values in the focused condition than in the unfocused, meaning that the pulses were more symmetrical in the focused condition, independent of stress. PC2 (factor scores of the second principal component) indicated that strong focused items differed from the others by a shallower opening slope. Since the interaction was significant and only strong syllables showed a significant focus effect, this change might be linked to  $f_0$  variation.

#### IV. DISCUSSION

The hypothesis that lexical stress distinction is produced by the same voice source mechanisms as vocal effort was supported by two derived shape parameters of the Lx pulse, the OQ, and PC. For both parameters, the open phase was shortened for higher levels of stress and vocal effort. It is important to note that on the one hand these differences for strong and weak syllables were independent of focus, i.e., strong syllables were produced with a longer closed phase than weak syllables in both focused and unfocused positions. The Lx pulse for focused stimuli was on the other hand more symmetrical, as indicated by increase in skewness of the Lx pulse parametrized as SQ. Both loud and soft items were also produced with a more symmetrical pulse than the focused items, because the former were also produced with pitch accents. Since the Lx pulses were time-normalized before further processing, these effects of stress, focus, and vocal effort cannot be attributed to  $f_0$  differences. Therefore, vocal effort and stress on the one hand and focus on the other exert different influences on the shape of Lx pulses.

The relevant parameters here differ from Marasek's (1997) who found that stress influenced the steepness of the slopes. The slopes were not affected in the current study. Furthermore, in his study the realization of pitch accents increased the OQ whereas in our study the SQ was modified and not the OQ. Two explanations might account for these different empirical results. First, Marasek (1997) did not vary stress and accent orthogonally, i.e., the two factors were confounded because all strong items were produced with a pitch accent and the weak without. Related to this, the second difference is that Marasek (1997) did not time-normalize the data. As a consequence, the differences in the closing slopes he found for the stress distinction might be attributable to changes in period duration, e.g., steeper slopes for higher  $f_0$  in strong items.

Another striking difference that is less easy to explain comes from Marasek's (1996, 1997) modeling study. He attributed the unexpected OQ increase for a higher subglottal pressure to a more abrupt closing and opening. However, in our study we found an OQ decrease for higher levels of

stress and vocal effort, which is not only more plausible than his increase but also very well supported by results from airflow data (e.g., Dromey *et al.*, 1992; Holmberg *et al.*, 1988, Stathopoulos and Sapienza, 1993). Our results based on Lx data showing more symmetrical pulses for accented stimuli could be matched more closely to airflow data than those based on Marasek's (1997) more indirect modeling. Even though the data presented here are generally more in agreement with results based on airflow data, the very prominent and consistently affected parameter SQ did not significantly decrease for vocal effort increase in our Lx data. The lack of an effect on the symmetry of the Lx pulses has also been noted by others (e.g., Dromey *et al.*, 1992; Sapienza *et al.*, 1998). A change in symmetry in the direction of the prediction was clearly visible, however, from visual inspection of the extreme PCA factor scores in the present study. The shape of the negative factor value of the first PC in Fig. 8 was more asymmetrical than the positive. The difference, however, lies more in the relative duration of glottal opening than in the abruptness of glottal closing, which has been proposed as one of the major causes for the boost in energy in the higher frequency ranges (e.g., Stevens, 1977; Sluijter and van Heuven, 1996). Since for the modal voice of male speakers, the Lx pulse is already very left skewed compared to the airflow glottal pulse (e.g., Dromey *et al.*, 1992) or the glottal area (Childers *et al.*, 1990), an upper limit for the temporal sensitivity of the recording device might contribute to these negative results. A change in the opposite direction, i.e., more symmetry with shallower closing slopes, was associated with both an increase in  $f_0$  and variation in PC2.

In summary, the functional version of a PCA, applied to Lx pulses in this study for the first time, has provided a more holistic analysis of the shape of Lx pulses. The resulting factor scores can be seen as parameters of the pulse shape amendable to further exploration by traditional statistical methods. The major advantage of the FPCA is that it does not rely on the identification of often rather arbitrarily defined landmarks, such as the instance of glottal opening. This error-prone landmark definition for the SQ and the opening and closing slopes of the Lx pulse probably contribute to the absence of an effect of stress and vocal effort on these parameters in previous studies.

As argued above, the steepness of the glottal closing is assumed to be related to the spectral shape, i.e., the more abruptly the glottis is closed, the flatter the spectral slope. The measure for spectral slope used in the current study,  $H1^*-A3^*$  with the Iseli *et al.* (2007) corrections, was only affected by vocal effort, not by stress. There was only an insignificant tendency for a lower spectral tilt for strong syllables than weak syllables, but only when the latter were in focused position. This result is contrary to the very clear word stress effects found by Sluijter and van Heuven (1996) for Dutch and to Okobi (2006) for English. However, as was stated in the Introduction, there are several other studies that also did not replicate Sluijter and van Heuven (1995), e.g., Claßen *et al.*, 1998 for German; Campbell and Beckman, 1997; and Hanson, 1997b for English. However, across all these studies there is variation in the languages that were investigated, the vowels that were analyzed, and in the algo-

rithms that were implemented. It is premature to conclude that spectral tilt is an independent correlate of word stress taking into account that some studies were not able to find a measure of spectral slope that could be generalized across different vowel types, vowel realizations, and  $f_0$ . In German, as in many languages, the reduction in weak syllables comes about both because they are more strongly co-articulated with the adjacent sounds (e.g., Mooshammer and Geng, 2008 for German) and because they are more centralized in formant space. In our data, F3 was affected more by stress ( $/e/_{\text{strong}}=2713$  Hz,  $/e/_{\text{weak}}=2504$  Hz) than by vocal effort ( $/e/_{\text{loud}}=2761$  Hz,  $/e/_{\text{normal}}=2713$ ,  $/e/_{\text{soft}}=2726$ ) (see also Traunmüller and Eriksson, 2000). It is not clear whether our negative results for stress has been brought about because the Iseli *et al.* (2007) method applied here has compensated for stress. This would imply that spectral tilt as a correlate of word stress in other studies can simply be attributed to modification in the vocal-tract and not to changes in subglottal pressure or glottal configurations. However, since the present laryngographic study leads to opposite conclusions, more modeling studies are needed for investigating the effect of glottal pulse shape changes on the shape of the power spectrum, independently of the formant structure.

## V. CONCLUSIONS

In this study, effects of vocal effort, stress and focus on glottal and acoustic parameters were analyzed in order to identify one or more reliable correlates of stress, independently of other prosodic variations. The hypothesis was that changes due to word stress affect the same set of parameters as vocal effort changes. In order to test this hypothesis, seven speakers of German were recorded by means of a laryngograph processor. The most important finding was that strong syllables were produced with a longer closed phase and an Lx pulse shape that resembled the Lx pulses also observed during loud speech. Holistic Lx pulse shape differences were parametrized by applying a new method, the functional version of a principal component analysis. Only these two parameters, derived from the Lx signal, varied with stress and vocal effort in a similar direction independent of focal accent. Acoustic parameters were either affected by stress and focal accent together or by only one of stress or vocal effort. The negative results for spectral tilt, which have also been found in other studies, can probably be attributed to changes in formant frequencies due to vocal-tract modifications. In conclusion, the most reliable and consistent correlates of stress and vocal effort, OQ and glottal pulse shape, were derived from the Lx signal. Since they varied with stress in the absence of  $f_0$  changes, these glottal adjustments can be interpreted as an independent phonetic dimension for signaling lexical stress. Further research is needed in order to determine the actual causes of these changes, namely, subglottal pressure changes or modifications in the glottal configuration.

## ACKNOWLEDGMENTS

I would like to thank Jelena Krivokapic, Mark Tiede, Christian Geng, Christine Shadle, and Arthur Abramson for

their valuable suggestions, Jonathan Harrington for help with the corpus design and with EMU/R as well as for very helpful comments, Jennifer Schneeberg for data labeling, and Herbert Fuchs for assisting the recordings. This work was partially supported by NIH Grant No. NIDCD DC008780.

- Beckman, M. E. (1986). *Stress and Non-Stress Accent* (Foris, Dordrecht, The Netherlands).
- Bombien, L., Cassidy, S., Harrington, J., John, T., and Palethorpe, S. (2006). "Recent developments in the EMU Speech Database System," in Proceedings of the 11th Australasian International Conference on Speech Science, pp. 313–316.
- Campbell, N., and Beckman, M. E. (1997). "Stress, prominence and spectral tilt," in Proceedings of the ISCA Workshop on Intonation: Theory, Models and Applications, pp. 67–70.
- Childers, D., Hicks, D., Moore, G., Eskenazi, L., and Lalwani, A. (1990). "Electroglottography and vocal fold physiology," *J. Speech Hear. Res.* **33**, 245–254.
- Choi, J.-Y., Hasegawa-Johnson, M., and Cole, J. (2005). "Finding intonational boundaries using acoustic cues related to the voice source," *J. Acoust. Soc. Am.* **118**, 2579–2587.
- Claßen, K., Dogil, G., Jessen, M., Marasek, K., and Wokurek, W. (1998). "Stimmqualität und Wortbetonung im Deutschen (Voice quality and word stress in German)," *Linguistische Berichte* **174**, 202–245.
- Colton, R. H., and Conture, E. (1990). "Problems and pitfalls of electroglottography," *J. Voice* **4**, 10–24.
- Dromey, C., Stathopoulos, E. T., and Sapienza, C. M. (1992). "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," *J. Voice* **6**, 44–54.
- Fant, G., Kruckenberg, A., and Liljencrants, J. (2000). "Acoustic-phonetic analysis of prominence in Swedish," In *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic, Dordrecht), pp. 55–86.
- Finnegan, E. M., Luschei, E. S., and Hoffman, H. T. (2000). "Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech," *J. Speech Lang. Hear. Res.* **43**, 934–950.
- Fry, D. B. (1955). "Duration and intensity as physical correlates of linguistic stress," *J. Acoust. Soc. Am.* **27**, 765–768.
- Gendrot, C. (2003). "EGG and spectral investigations on final focalised positions in French," in Proceedings of the 15th ICPHS, pp. 547–550.
- Grice, M., Baumann, S., and Benz Müller, R. (2005). "German intonation in autosegmental-metrical phonology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by J. Sun-Ah (Oxford University Press, Oxford), pp. 55–83.
- Hanson, H. M. (1997a). "Glottal characteristics of female speakers: acoustic correlates," *J. Acoust. Soc. Am.* **101**, 466–481.
- Hanson, H. (1997b). "Vowel amplitude variation during sentence production," in Proceedings of the ICASSP, pp. 1627–1630.
- Hanson, H. M., and Chuang, E. (1999). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," *J. Acoust. Soc. Am.* **106**, 1064–1077.
- Heldner, M. (2003). "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," *J. Phonetics* **31**, 39–62.
- Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (2004). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," *J. Acoust. Soc. Am.* **115**, 1321–1332.
- Hirano, M., Ohala, J. J., and Vennard, W. (1969). "The function of laryngeal muscle in regulating fundamental frequency and intensity of phonation," *J. Speech Hear. Res.* **12**, 616–628.
- Hixon, T. J. (1973). "Respiratory function in speech," in *Normal Aspects of Speech, Hearing, and Language*, edited by F. Minifie, T. J. Hixon, and F. Williams (Prentice-Hall, Englewood Cliffs, NJ), pp. 73–125.
- Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (1988). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," *J. Acoust. Soc. Am.* **84**, 511–529.
- Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (1989). "Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch," *J. Voice* **3**, 294–305.
- Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., and Goldman, S. L. (1995). "Comparison among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *J. Speech Hear. Res.* **38**, 1212–1223.
- Howard, D. M., Lindsey, G. A., and Allen, B. (1990). "Toward the quantification of vocal efficiency," *J. Voice* **4**, 205–212.
- Iseli, M., Shue, Y.-L., and Alwan, A. (2007). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *J. Acoust. Soc. Am.* **121**, 2283–2295.
- Ladefoged, P. (1967). "Stress and respiratory activity," in *Three Areas of Experimental Phonetics*, edited by P. Ladefoged (Oxford University Press, London), pp. 1–49.
- Ladefoged, P. (2005). "Speculations on the control of speech," in *A Figure of Speech: A Festschrift for John Laver*, edited by W. J. Hardcastle and J. MacKenzie Beck (Erlbaum, Mahwah, NJ), pp. 3–21.
- Laver, J. (1980). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge).
- Lee, S., Byrd, D., and Krivokapic, J. (2006). "Functional data analysis of prosodic effects on articulatory timing," *J. Acoust. Soc. Am.* **119**, 1666–1671.
- Lehiste, I. (1970). *Suprasegmentals* (MIT, Cambridge, MA).
- Lucero, J., and Koenig, L. (2000). "Time normalization of voice signals using functional data analysis," *J. Acoust. Soc. Am.* **108**, 1408–1420.
- Lucero, J., and Löfqvist, A. (2005). "Measures of articulatory variability in VCV sequences," *ARLO* **6**, 80–84.
- Marasek, K. (1996). "Glottal correlates of the word stress and the tense/lax opposition in German," in Proceedings of the ICSLP, University of Stuttgart, Stuttgart, Vol. **96**, pp. 1573–1576.
- Marasek, K. (1997). "Electroglottographic description of voice quality." AIMS Working Papers Stuttgart, Vol. **3**.
- Mokhtari, P., Pfitzinger, H. R., and Ishi, C. Toshinori (2003). "Principal components of glottal waveforms: towards parameterisation and manipulation of laryngeal voice quality," in Proceedings of the VOQUAL, pp. 133–138.
- Mooshammer, C., and Geng, C. (2008). "Acoustic and articulatory manifestations of vowel reduction in German," *J. Int. Phonetic Assoc.* **38**, 117–136.
- Okobi, A. (2006). "Acoustic correlates of word stress in American English," Ph.D. thesis, MIT, Cambridge.
- Pierrehumbert, J. (1997). "Consequences of intonation for the voice source," in *Speech Production and Language: In honour of Osamu Fujimura*, edited by S. Kiritani, H. Hirose, and H. Fujisaki (Mouton de Gruyter, Berlin), pp. 111–130.
- Ramsay, J. O., and Silverman, B. W. (1997). *Functional Data Analysis*. (Springer, New York).
- Ramsay, J. O., and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. (Springer, New York).
- Sapienza, C. M., Stathopoulos, E. T., and Dromey, C. (1998). "Approximation of open quotient and speed quotient from glottal airflow and EGG waveforms: Effects of measurement criteria and sound pressure level," *J. Voice* **12**, 31–43.
- Schulmann, R. (1989). "Articulatory dynamics of loud and normal speech," *J. Acoust. Soc. Am.* **85**, 295–312.
- Sluijter, A., and van Heuven, V. (1996). "Spectral balance as an acoustic correlate of linguistic stress," *J. Acoust. Soc. Am.* **100**, 2471–2485.
- Stathopoulos, E. T., and Sapienza, C. M. (1993). "Respiratory and laryngeal function of women and men during vocal intensity variation," *J. Speech Hear. Res.* **36**, 64–75.
- Stevens, K. N. (1977). "Physics of laryngeal behavior and larynx modes," *Phonetica* **34**, 264–279.
- Titze, I. R. (1988). "Regulation of vocal power and efficiency by subglottal pressure and glottal width," in *Vocal Fold Physiology: Voice Production, Mechanisms, and Functions*, edited by O. Fujimura (Raven, New York), pp. 227–238.
- Titze, I. R. (1990). "Interpretation of the electroglottographic signal," *J. Voice* **4**, 1–9.
- Titze, I. R. (1992). "Phonation threshold pressure: a missing link in glottal aerodynamics," *J. Acoust. Soc. Am.* **91**, 2926–2935.
- Traunmüller, H., and Eriksson, A. (2000). "Acoustic effects of variation in vocal effort by men, women and children," *J. Acoust. Soc. Am.* **107**, 3438–3451.
- Winkworth, A. L., Davis, P. J., and Ellis, E. (1995). "Breathing patterns during spontaneous speech," *J. Speech Hear. Res.* **38**, 124–144.