Voice quality dimensions of pitch accents

Britta Lintfert, Wolfgang Wokurek

Institute of Natural Language Processing
University of Stuttgart, Germany
`Britta.Lintfert; Wolfgang.Wokurek@ims.uni-stuttgart.de`

10.21437/Interspeech.2005-48

Abstract

Acoustic and electroglottographic (EGG) measurements were used to examine voice quality parameters during the production of the rising and falling pitch movements in German. The vowels /aː/ and /ɛ/ were studied in a single-speaker speech corpus. The acoustic measurements comprised an automatic spectral analysis of the glottal parameters open quotient (OQ), glottal opening (GO), skewness of glottal pulse (SK), rate of closure (RC), amplitude of voicing (AV) and completeness of closure (CC). OQ and AV seem to be the only acoustic parameters influenced by pitch accent and not by word stress. From the electroglottographic measurements only open quotient parameters (OQI and OQII), two parameters of closing phase (SCV and SCA) and two parameters of opening phase (EOV and EOT) showed a significant difference as a function of pitch accent type.

1. Introduction

This work is part of a research project which proposed an extension and generalization of Guenther's and Perkell's speech production model [1, 2, 3] from the segmental to the prosodic domain [4, 5]. According to Guenther and Perkell, there is a unique phonetic target region in auditory-temporal space for each phoneme of a given language. We have recently suggested to integrate the model with an exemplar-theoretical view by asserting that accumulations of exemplars implicitly define corresponding regions in perceptual space that serve as targets in the production of prosody [6]. Thus, the speaker has access to stored representations of prosodic events, including their tonal and temporal structure, serving as a reference in speech production.

The aim of the work presented in this paper was to perform a mapping of tonal acoustic parameters to articulatory gestures. For the purpose of computational modeling, the optimal mapping from invariant perceptual prosodic targets to variable acoustic or articulatory targets can be learned by means of supervised machine learning. But before we can perform a computational mapping we must know more about how the invariant perceptual target of a pitch accent will be produced by variable articulatory targets and variable articulator movements. Speakers and listeners have access to phonemic prosodic models that share all crucial properties with internal phonemic models on the segmental level [4, 6]. Phonemic prosodic models emerge during language acquisition [7, 8] and seem to be stable across dialect and clinical conditions [3, 9]. The speaker may rely on a set of acquired internal models (in our case of pitch accents) and select from this set a particular model depending on communicative and situative constraints and use variable acoustic and articulatory gestures for realization of the target pitch accent.

The invariant goal can be achieved by a number of different variable articulatory goals, e.g. adducing or slackening the vocal folds, generating sufficient subglottal pressure, producing a sufficient degree of mouth opening or/and enlarging the capacity of the vocal tract [10]. In practice there should be a fine-grained combination of these factors to produce stress and accents and various combinations of these gestures can contribute to the production of word stress and pitch accents. Word stress and pitch accent are two different linguistic constructs, and they have separate acoustic and perceptual correlates [11, 12, 13]. As stress is not just a weaker degree of accent, only correlates relating to the presence of an accent-lending pitch movement show weakening.

2. The speech corpus

The experiments reported in this paper are based on a large, phonetically and prosodically well-designed, single-speaker speech corpus [14] recorded for unit selection speech synthesis in an anechoic chamber at the Phonetics Department of the Institute of Natural Language Processing (IMS), University of Stuttgart. The data amounts to almost 160 minutes of speech comprising approximately 94 000 segments and 37 000 syllables, we analysed about 1100 syllables. The electroglottographic (EGG) signal was recorded simultaneously with the microphone signal.

3. Methods

Measurements of laryngeal and supralaryngeal articulatory gestures were not performed; instead, we relied on an indirect extraction of voice source parameters by measuring spectral correlates of voice source parameters [15], which is a promising technique [16, 13, 17]. Voice source parameters were also studied using the electroglottographic (EGG) signal recorded simultaneously with the speech signal [13, 18]. Acoustic and EGG-based measurements were made every 10 ms in realizations of the vowels [aː] and [ɛ] that occurred with the labeled pitch accents "H*L", "L*H" and no pitch accent ("NPA") in the segmental context of sonorants (SON) and voiced obstruents (VOB). Because the analysis tools are dependent on the vocal formants, we have to separate the analyzed speech samples by vowel identity. For this work we have analyzed only the segmental context of sonorants and voiced obstruents. The statistical sample size for each pitch accent type was in the range of 400–4000.

3.1. Acoustic measurements

For the acoustic analysis the automatic signal analysis procedure was based on amplitude and frequency measurements at harmonic spectral peaks [15]. The procedure provides spectrum-based correlates of voice quality parameters.

The voice quality parameters open quotient, glottal opening, etc. are defined in time domain. The notion of finding correlates of them in frequency domain was introduced by Stevens and Hanson [19]. These frequency domain measurements we use are rough estimates of the temporal voice quality parameters and we use their names for convenience. Further developement of the spectral estimation formulas are still in progress and the following equations represent a snapshot in this development. The correlations between the following paramters may still be reduced. The dependency of the voice quality parameters on the vowel quality may also be reduced in future.

- **Open quotient (OQ)**
  The open quotient indicates the time during which the glottis is open, defined in the time domain as a fraction of the total period. The primary acoustic manifestation of a narrow glottal pulse, i.e. of a decrease in open time, is a reduction of the amplitude of the fundamental component in the source spectrum relative to adjacent harmonics. A correction is being made for the effect of the formants on the amplitude of the first and second harmonic. The presence of that correction is denoted by appending a tilde (~) to the varibale name. The resonance gain of each of the four formants is substracted.

  If either the spectral peak of fundamental frequency is under 70 Hz and thus too low, or the octave distance between the fundamental wave and the second harmonic varies for more than 10%, the first error condition ERR1 is produced. In this case the harmonic structure is not reliable represented in the actual spectrum and the affected values were removed from further analysis.

  ```
  OQ~ = ( H1~ - H2~ )
  ```

- **Glottal opening (GO)**
  Degree of opening over the entire glottal cycle. The amplitude of F1 (A1) is influenced by the glottal aperture during the open phase (H1-A1). A correction has is being made for the effect of F1–F4 on H1 and F2–F4 on A1 [19].

  If the frequency of the second harmonic is too close to or above the first formant, the second error condition ERR2 is produced. The affected values were removed from further analysis.

  ```
  GO~ = ( H1~ - A1~ )
  ```

- **Skewness (SK)**
  The abruptness of glottal closure influences the spectrum of the glottal waveform at mid and high frequencies.

  ```
  SK~ = ( H1~ - A2~ )
  ```

- **Rate of closure (RC)**
  Duration of the closing part, which directly influences the skewness of the glottal pulse.

  ```
  RC~ = ( H1~ - A3~ )
  ```

- **Amplitude of voicing (AV)**
  Controls the height of each glottal pulse. The height of H1, corrected for the influence of F1, gives information about AV.

  ```
  AV = H1~
  ```

- **Completeness of closure (CC)**
  Energy loss in the F1 range, adding significantly to the F1 bandwidth (B1) when the glottis is not completely closed during phonation.

  ```
  CC = B1
  ```

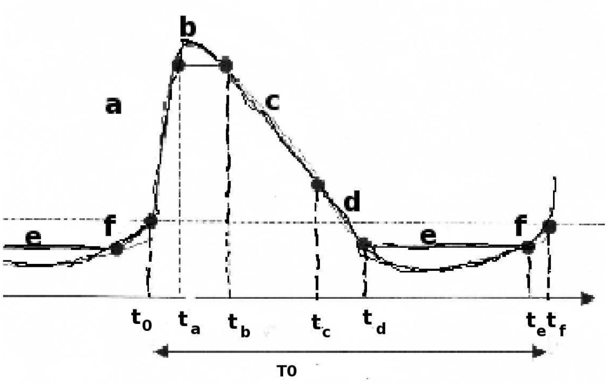

Figure 1: Description of the EGG signal shape using 6 straight segments and timing of the phases, see text for explanation of segments labeled with letters a–f

3.2. Articulatory measurements

In the EGG signals the pitch period is usually defined as the duration between maximum positive peaks in the differentiated EGG waveform. Not only the relative duration of the different phases but also the skewness of the glottal waveform seem to be important for articulatory realization of pitch accents.

For the articulatory analysis the EGG signal was segmented and described using a set of timing and shape parameters [18]. Every period of the EGG signal was segmented according to the temporal intervals (*a*–*f*) and temporal instances (t_a–t_f) illustrated in Fig.1.

The maximum-contact phase (close phase) *b* was defined to lie above 90% of peak-to-peak amplitude, while the no-contact phase (start of opening phase) *e* was below 10% of peak-to-peak amplitude. The beginning of the closing phase (*a*) is determined by the position of the peak change of the current flow (the steepest of the EGG rise signal). The opening phase (*c*) and (*d*) is converted into two parts of the opening instant. The open segment (*e*) lies below 10% of the peak-to-peak amplitude and the last segment (*f*) connects the open phase with the point of steepest increase of the EGG signal.

The following parameters are of sustained interest for analysis:

- **Open Quotient I (OQI)**

$$OQI = 100 * \frac{t_f - t_c}{T0} \quad (1)$$

- **Open Quotient II (OQII)**

$$OQII = 100 * \frac{t_e - t_d}{T0} \quad (2)$$

- **Parameters of closing phase**
  SCV = variability at the beginning of closing movement (t_e to t_f);
  ECV = variability at the end of closing movement (t_0 to t_a).

- **Parameters of opening phase**
  EOV = variability at the end of the opening movement (t_c to t_d);
  EOT(%) = relative duration of the ending phase of opening movement (t_c to t_d).

4. Results

Statistical analysis was carried out using SPSS version 12.0. First, multivariate analysis of variance were performed to test the parameters' ability to distinguish the pitch accent types. The measurements were analyzed in four groups, splitting the data by segmental context (SON, VOB) and vowel identity (/aː/, /ɛ/) because there was a significant influence of the proceding context on accent types. There was no significant influence of the following context on voice quality parameters distinguishing pitch accent types. The post-hoc Waller-Duncan t-test (harmonic mean, unequal groups) showed a significant differentiation of pitch accent types H*L, L*H, and NPA ('no pitch accent') in preceding segmental contexts SON vs. VOB, on the basis of the acoustic (Table 1) and articulatory (Table 2) parameters.

	OQ	GO	SK	RC	AV	CC
SON, [aː]	*	*			*	
VOB, [aː]	*	*		*		
SON, [ɛ]	*			*	*	*
VOB, [ɛ]	*	*	*	*	*	*

Table 1: Acoustic correlates: significant parameters for H*L, L*H and NPA

	OQI	OQII	SCV	ECV	EOV	EOT(%)
SON, [aː]	*	*	*	*	*	*
VOB, [aː]			*	*	*	*
SON, [ɛ]	*	*	*		*	
VOB, [ɛ]	*	*	*	*	*	*

Table 2: Articulatory correlates, significant parameters for H*L, L*H and NPA

The acoustic parameter OQ was significant for SON [aː] (i.e., vowel [aː] preceded by a sonorant), SON [ɛ], VOB [aː] (i.e. [aː] preceded by an obstruent) and VOB [ɛ]. GO was significant for SON [aː], VOB [aː] and VOB [ɛ] but not for VOB [aː]. SK was only significant for VOB [ɛ]. RC was significant for VOB [aː], SON [ɛ] and VOB [ɛ] but not for SON [aː]. It is important to note that the acoustic parameters GO, SK and RC are also known to be correlates of German stress [17]. AV was significant for SON [aː], SON [ɛ] and VOB [ɛ] but not for VOB [aː]. CC showed significant differences for SON [ɛ] and VOB [ɛ], which may contribute to the findings that stressed syllables show more glottal leakage than unstressed syllables. This effect has been found to be independent of accentuation [11].

As a consequence, stress and accent appear to have confounding effects on the acoustic parameters measured in this study, with the exception of OQ and AV, which are the only acoustic parameters influenced by accent and not by stress. This result corresponds well with the literature [16, 11, 12, 13, 17].

The most informative articulatory parameters influenced by accent seem to be the open quotient and the additional modulation of muscular tension shown by the significant parameters OQI and OQII for SON [aː], SON [ɛ] and VOB [ɛ]. SCV was significant for all observed contexts and vowels. ECV was significant for SON [aː], VOB [aː] and VOB [ɛ], but not for SON [ɛ]. EOV was significant for all observed contexts and vowels and EOT(%) was significant for SON [aː], VOB [aː] and VOB [ɛ] but not for SON [ɛ]. The additional modulation of muscular tension affects the skewness of the waveform at the beginning and ending of the closing phase (SCV and ECV) and at the ending of opening phase (EOV, EOT(%)). The other possible parameters described in [18] did not show such significant differentiation of pitch accents. Increasing fundamental frequency causes an increase of the open quotients, greater SCV, EOV and EOT(%) but lower ECV. For the variation of pitch connected with fundamental frequency movement, fine-grained modifications of the vocal fold tension is needed.

5. Discussion

The main laryngeal correlate of pitch accents seems to be an additional modulation of muscular tension shown in significant differences of the Open Quotients. An increasing fundamental frequency requires an increase of vocal fold tension or a rise of subglottal pressure, or both. The pitch level can be altered through the intentional change of activation of the intrinsic laryngeal muscles. An increase of subglottal pressure and/or vocal fold tension affects the skewness of the airflow waveform (SCV, ECV, EOV). A higher muscular tension also supports an increase in the glottal airflow duty ratio, which in turn affects the relative duration of the ending phase of the opening movement (EOT(%)). Thus, EOT(%) and the skewness of the airflow waveform are found to be significant for differentiating the pitch accents and appear to be articulatory correlates by means of which pitch accents in German can be distinguished. Pitch accent in German is controlled by different phonatory mechanisms, and these differences are measurable in the EGG domain.

The methods of analysis presented in this paper contribute to the analysis of pitch accents with the help of acoustic and articulatory parameters. Both applied analysis tools provide a characterization of voice quality in the physiological domain, viz. the glottal waveform. The acoustic and articulatory parameters facilitate a selective distinction of the different accent variations. Altogether, an additional muscular tension seems to be responsible for the change of pitch. Subglottal pressure, however, ties up with accentuation. Being put under more pressure, the stimulation rises and the parameters describing the skewness of the waveform are affected in their variability by increasing or decreasing pressure.

A tendency appears to emerge according to which the acoustic parameters best describe the opening stage of articulation and the parameters derived from the EGG signal best describe the closing phases (Pützer, personal communication). The acoustic parameters that depend on the time of opening (OQ) and height of glottal pulse (AV) seem to be influenced by accent and not by stress. The articulatory parameters describe stages of adduction or relative closure of the vocal folds. These parameters depend on the shape of the F_0 curve of the target pitch accent. Higher subglottal pressure must be built up for accent than for stress, which is reflected by a longer time of opening and greater height of the glottal pulse. But higher subglottal pressure also requires a careful adjustment of the vocal folds to achieve the appropriate pitch height. An increase of the subglottal pressure correlates with the linguistic marking of word stress, and causes a higher skewness of the glottal pulse. Additional muscular tension is needed for marking accents which in turn contributes to the prosodic function of sentence intonation. Higher tension of the vocal folds causes an increase of fundamental frequency, which is the best correlate to mark pitch accent.

The invariant perceptual target of a pitch accent will be produced by variable articulatory targets and variable articulatory movements. It seems that the control over subglottal pressure is important for marking intonation. Pitch accents seem to need additional muscular tension to control the skewness of the vocal folds. The skewness of the vocal folds affects the shape of the F_0 curve of the target pitch accent. Pathological voices with impaired control of vocal fold settings seem to have a reduced control of variation of pitch accents compared to healthy voices [18].

6. Conclusion

The aim of this work was to perform a mapping of tonal acoustic parameters to articulatory gestures. For a computational mapping more data are required, above all from different speakers, to find out more about of how pitch accents are produced in German. An additional aspect is how phonemic prosodic models are acquired and at which stage of vocal development such fine-grained control of pitch accents appears that is found in adult speech. This question is part of subsequent research.

Both methods of analysis presented in this paper are suitable to analyze language data for parameters of pitch accents in German. It can be assumed that the acoustic analysis of the spectrum provides an adequate representation of the opening phase of the vocal folds and the EGG waveform provides an adequate representation of the closing phase of the vocal folds. Since the two methods relate to different stages of articulation, both methods of analysis should be used for a complete mapping from invariant perceptual targets of pitch accents to variable articulatory targets and variable articulator movements that produce these pitch accents. There is no overlapping or redundance of the parameters; the acoustic and articulatory parameters do not describe parallel but complementary voice quality features of the examined pitch accents.

7. Acknowledgments

We thank Manfred Pützer for helpful discussions and Manuel Kountz for technical support. This experiment is part of the project *Target Oriented Production of Prosody* funded by the German Research Foundation (DFG, Grant DO 536/4-1).

8. References

[1] Guenther, F.H., "A modeling framework for speech motor development and kinematic articulator control", Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm, Vol. 2, p 92–99, 1995.

[2] Guenther, F.H., Hampson, M. and Johnson, D., "A theoretical investigation of reference frames for the planning of speech movements", Psychological Review, 105:611–633, 1998.

[3] Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. and Zandipour, M., "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss", J. Phon., 28(3):233–272, 2000.

[4] Dogil, G. and Möbius, B., "Towards a model of target oriented production of prosody", Proceedings of the European Conference on Speech Communication and Technology (Aalborg, Denmark), Vol.1, p 665–668, 2001.

[5] Möbius, B. and Dogil, G., "Phonemic and postural effects on the production of prosody", Speech Prosody 2002 (Aix-en-Provence), p 523–526, 2002.

[6] Schweitzer, A. and Möbius, B., "Exemplar-based production of prosody: Evidence from segment and syllable durations", Speech Prosody 2004 (Nara, Japan), p 459–462, 2004.

[7] Beckman, M.E., "Input Representations (Inside the Mind and Out)", In Garding, G. and Tsujimura, M., (eds), WCCFL 22 Proceedings, p 70–94, Somerville, MA: Cascadilla Press, 2003.

[8] Curtin, S.L., Representational richness in phonological development, PhD thesis, University of Southern California, 2002.

[9] Claßen, K., "Realisation of nuclear pitch accents in Swabian dialect and Parkinson's dysarthria: A preliminary report", Speech Prosody 2002 (Aix-en-Provence), p 223–226, 2002.

[10] Jessen, M., "Phonetic implementation of the distinctive auditory features [voice] and [tense]", Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), Vol. 6, p 11–62, University of Stuttgart, 2000.

[11] Sluijter, A.M.C., Phonetic Correlates of Stress and Accent, PhD thesis, University of Leiden, 1995.

[12] Jessen, M., Marasek, K., Schneider, K., and Claßen, K., "Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German", Proceedings of the 13th ICPhS (Stockholm), Vol. 4, p 428–431, 1995.

[13] Marasek, K., Electroglottographic description of voice quality, Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS), Vol.3, University of Stuttgart, 1997.

[14] Schweitzer, A., Braunschweiler, N., Klankert, T., Möbius, B., and Säuberlich, B., "Restricted unlimited domain synthesis", Proceedings of Eurospeech 2003 (Geneva), p 1321–1324, 2003.

[15] Wokurek, W. and Pützer, M., "Automated corpus based spectral measurement of voice quality parameters", Proceedings of the 15th ICPhS (Barcelona), p 2173–2176, 2003.

[16] Hanson, H.M., "Glottal characeristics of female speakers: Acoustic correlates", J. Acoust. Soc. Amer., Vol. 101:466–481, 1997.

[17] Claßen, K., Dogil, G., Jessen, M., Marasek, K., and Wokurek, W., "Stimmqualität und Wortbetonung im Deutschen", In: Linguistische Berichte, Vol. 174, p 202–245, Westdeutscher Verlag, 1998.

[18] Pützer, M. and Marasek, K., "Differenzierung gesunder Stimmqualitäten bei Rekurrensparese mit Hilfe elektroglottographischer Messungen und RBH-System", Sprache, Stimme, Gehör, 24:154–163, 2000.

[19] Stevens, K.N. and Hanson, H.M., "Classification of glottal vibration from acoustic measurements", in Fujimura, O. and Hirano, M. (eds), Vocal Fold Physiology: Voice Quality Control, p 147–170, Singular, San Diego, 1994.