ACOUSTIC CORRELATES OF STRESS AND ACCENT IN STANDARD AUSTRIAN GERMAN

*Dina El Zarka**
Barbara Schuppler†
*Carina Lozo**
*Wolfgang Eibler**
*Patrick Wurzwallner**

**Department of Linguistics, University of Graz*
†SPSC Laboratory, Graz University of Technology

Abstract

Previous studies on the acoustic correlates of stress as metrical word prominence and (pitch) accent have come to contradictory conclusions concerning the phonetic reality of stress. Some suggest spectral tilt (spectral balance, spectral emphasis) as a phonetic correlate of stress in the abstraction from the effect of intonational pitch accent (e.g., Slujter & van Heuven, 1996a, b), while others found no difference in spectral tilt between stressed and unstressed vowels in the absence of pitch accent (e.g., Campbell & Beckman, 1997). So far, all studies used an experimental design with narrowly focused versus deaccented target words to test their hypotheses. Drawing on a large corpus of read sentences of Standard Austrian German (cf. Moosmüller et al., 2015) as pronounced by 37 speakers from the eastern provinces of Austria (Schuppler et al., 2014), our study examines vowel quality, duration, and spectral tilt in comparable syllables in different positions and accent conditions, using different factors as independent variables in a mixed effects logistic regression analysis. We take this approach as a first step for a subsequent investigation of the acoustic correlates of prominence in conversational speech.

Bisherige Studien zu akustischen Korrelaten von Wortakzent im Sinne metrischer Prominenz und (Ton-)Akzent im Sinne intonatorischer Hervorhebung kommen zu unterschiedlichen Ergebnissen im Hinblick auf die phonetische Realität von Wortakzent. Während einige Untersuchungsergebnisse nahelegen, dass ein Wortakzent auch ohne intonatorischen Tonakzent phonetisch durch spectral tilt (spectral balance, spectral emphasis) realisiert wird (z.B. Slujter & van Heuven, 1996a, b), fanden andere Studien keinen Unterschied im spectral tilt zwischen unbetonten und betonten Vokalen, wenn diese nicht auch durch einen Akzentton hervorgehoben waren (z.B. Campbell & Beckman, 1997). Alle bisherigen Studien verwendeten für die Untersuchung experimentelle Daten mit eng fokussierten versus deakzentuierten Zielwörtern. Die vorliegende Untersuchung stützt sich hingegen auf ein großes Korpus gelesener Sätze des Österreichischen Standarddeutsch (vgl. Moosmüller et al., 2015), gesprochen von 37 Sprechern aus den östlichen Regionen Österreichs (Schuppler et al., 2014). Untersucht werden Vokalqualität, Dauer und spectral tilt in vergleichbaren Silben in verschiedenen Positionen und unter unterschiedlichen Akzentbedingungen, wobei die verschiedenen Parameter in einem gemischten logistischen Regressionsmodell mit mehreren Faktoren als unabhängigen Variablen statistisch evaluiert werden. Die vorliegende Studie stellt einen ersten Schritt hin zur Untersuchung der akustischen Korrelate von Prominenz in konversationeller Spontansprache dar.

INTRODUCTION

Linguistic prominence is usually referred to as *stress* and *accent* in phonologically and phonetically oriented work. These two terms and their specific denotations, however, have created a great deal of confusion because what is called accent in one line of research is referred to as stress in another. The problem, however, is not merely terminological. Differences in terminology mostly also point to different underlying concepts.

From a perceptual point of view, *prominence,* in the sense it is used here, refers to an event in the speech flow that ‘stands out to the ear’. From an articulatory viewpoint, prominence is characterized by increased vocal effort. Acoustically, it is an admixture of phonetic features such as fundamental frequency, duration, intensity, spectral composition and spectral tilt or emphasis that collaborate in a not yet fully understood way to produce prominence.

Older phonetic work on accent/stress made a distinction between expiratory accent or dynamic stress (*Druckakzent*) and musical accent (*melodischer Akzent*) (Sweet, 1906). However, experiments by Fry (1955, 1958) showed that an accent-lending pitch movement was the most reliable cue to prominence in English minimal pairs, such as *pérmit* vs. *permít*, while intensity was only a weak cue. Beckman's seminal study (1986) essentially corroborated these findings. Beckman, however, was able to identify clear differences between what she called a *pitch-accent* language (Japanese) and a *stress-accent* language (English). Her study clearly showed that other acoustic correlates of prominence ('accent'), among them a special type of overall intensity measure, are much more salient in English than in Japanese. The problem with these studies, however, is that the target words always carried an intonational pitch accent and thus did not permit to answer the question whether word stress had any reliable acoustic correlates in the absence of a pitch accent.

In most recent work on linguistic prominence a distinction is made between word-level *stress* and phrase-level *accent.* However, it is still a matter of extensive debate whether word stress has phonetic correlates or whether it is merely a structural position that may or may not be given prominence in actual articulation. Thus phonetic studies on that topic may roughly be divided into two major lines. The first one views prominence as a uni-dimensional scale with three or four levels (e.g., Beckman & Edwards, 1994), whereas other scholars try to identify the acoustic cues that are specifically associated with *word stress* and differentiate between the latter and the post-lexical phenomenon1 of *accent* that has pitch, i.e. fundamental frequency, as its main phonetic correlate (e.g., Slujter & van Heuven, 1996a). On the assumption that "stress is [...] determined by the language system, and accent by language behaviour", Slujter and van Heuven (1996: 2471) argue that the older view was essentially correct in suggesting that stress is signaled by loudness. As we have seen above, previous studies have shown that overall intensity is only a weak cue to prominence. It has, however, been established in more recent work that overall intensity is not the only acoustic measure of loudness. Thus, Slujter and van Heuven suggested "that the acoustical correlate of greater physiological effort is a decrease of negative spectral tilt, or even a positive tilt." (p. 2472). Their conclusion was that *spectral balance* is indeed a reliable acoustic correlate of stress in the absence of pitch accent in Dutch, but also in American English (Slujter, 1995; Slujter & van Heuven, 1996b). A number of studies investigating the issue in various languages followed suit, delivering re-

---

[1] This is, of course, only true for intonation languages, which do not exploit pitch accents lexically, in contrast to pitch accent languages such as Swedish or Japanese that employ pitch differences on the lexical level.

sults that pointed in the same direction (Astruc & Prieto, 2006 for *Catalan*; Prieto & Ortega-Llebaria, 2006 for *Spanish* and *Catalan*; Gordon & Nafi, 2012 for *Tashelhiyt Berber*). However, other studies, relying on a similar methodology, came to contradictory conclusions for the languages investigated (Campbell & Beckman, 1997 for American English; Sadeghi, 2011 for Persian). Campbell & Beckman (1997: 70) interpret the missing difference in spectral balance measures between stressed and unstressed vowels in the absence of a pitch accent as lending support to Bolinger's (1958) view that stress (in English) is merely a structural position with no phonetic reality unless the "stressed" syllable is also "accented".

The studies mentioned above do not only differ regarding their conclusions, but also methodologically. It is, for instance, not at all clear what the best measure of spectral tilt or spectral emphasis is. In the next section, we will look into the different methods that have been suggested in more detail. Another methodological issue is the experimental design. All above mentioned studies used an experiment designed to investigate identical bi-syllabic target words under narrow focus versus deaccentuation conditions, which evidently is a promising method to find out whether stressed syllables can be identified in the absence of pitch. However, narrow focus and deaccentuation are two extreme prominence conditions that may yield effects that differ from 'neutral' articulation.

The ultimate goal of our work is the investigation of phonetic correlates of stress and accent in spontaneous speech. As an intermediate step between controlled production data and spontaneous speech, our present study draws on a large corpus of read sentences of Standard Austrian German. We chose to examine the acoustic correlates of specific syllables in different positions at word and phrase level and under different metrical prominence conditions, i.e. unstressed, stressed and accented. The different phonetic parameters are extracted and a number of variables are annotated. These are social variables, such as speaker and gender, and phonological variables, such as position, syllable type, segmental composition of the syllable, and metrical prominence. Additionally, we rely on a perceptual judgment from a small-scale perception task as another factor. All factors are included as independent variables in a mixed-effect logistic regression analysis that shall identify whether word stress and intonational accent can be predicted by the individual acoustic cues.

With our study we aim at increasing our knowledge concerning the following questions:

- Does stress have a phonetic reality distinct from accent?
- If yes, are the acoustic correlates associated with stress related to ‘loudness’ and thus compatible with the traditional notion of *expiratory/dynamic stress*?
- Or are the results rather compatible with the view that prominence is a uni-dimensional system and stress is just a structural position?

STANDARD AUSTRIAN GERMAN

Several studies in the literature describe the phonetic and phonological characteristics of Standard Austrian German as compared to Standard German spoken in Germany (e.g., Wiesinger, 2009; Klaaß, 2008; Bürkle, 1995; Moosmüller & Dressler, 1988; Mossmüller et al., 2015). The number of acoustic phonetic studies, however, is relatively small. In the following, we will provide a short summary of characteristics mentioned in the literature. The focus will be on vowel realizations and stress patterns, since these aspects are most relevant for the current study.

With respect to the realization of consonants, front fortis stops tend to be lenited, especially in intervocalic position, such that a word like *weiter ‘further’* with the citation form [vaɪtɐ] would be pronounced as [vaɪdɐ]. Moreover, lenis plosives are frequently realized as voiced fricatives with the same place of articulation (i.e., spirantization) (cf. Moosmüller & Ringen, 2004; Moosmüller, 2011). Other characteristics are low aspiration of all fortis plosives, frequent vocalizations of [R], the devoicing of alveolar fricatives in all word positions, and the pronunciation of the word-final syllable *-ig* as [ik] (instead of [iç], as usual in the northern German varieties, cf. Bürkle, 1995).

With respect to the realization of vowels, diphthongizations were reported for some regions (Hornung & Roitinger, 2000: 91-92). For the Viennese dialect, however, monophthongization is typical. This process seems to spread to other regions of Austria (Moosmüller, 1998).

Austrian German also differs from German German with respect to vowel duration. It has been argued that in Standard German German duration is a concomitant feature of vowel quality, i.e. tense vowels are pronounced significantly longer than lax vowels. Moosmüller (2007: 72) argues on the basis of careful production experiments that duration is not distinctive in Standard Austrian German.

Since the current study analyzes exclusively syllables with the nucleus [a], it is relevant to mention that only little difference has been found between tense and lax a-vowels for Standard Austrian German. Iivonen (1996), for instance, found that whereas for German speakers, F1 and F2 differ significantly for short and long /a/, for Austrian speakers, F1 and F2 are nearly the same. Similar observations have been reported earlier by Moosmüller (2007).

Duration is also a correlate of stress in German (Jessen et al., 1995). From a structural point of view, word stress in German is variable and mostly morpho-lexically governed. There is, however, a tendency for heavy syllables to be stressed. Furthermore, primary stress usually falls within the three final syllables of a word with a clear preference for the penultimate syllable (cf. Wiese 2011: 79). German has a high number of sometimes very long and complex compounds, the parts of which may be characterized by different degrees of prominence. In metrical terms, it is usually the first constituent of the compound that carries primary stress and the second part that is secondarily stressed. Additional stresses follow rhythmical principles. In the example from our corpus illustrated in Figure 1, main stress is thus on the first syllable and secondary stress on the first syllable of the second constituent, other stress positions are phonologically governed.

x							
x		x					
x		x		x		x	
x	x(x)	x	x	x	x	x	x
Vi	deo	ge	gen	sprech	an	la	ge

Figure 1: Primary and secondary stress in German compounds. Example from the GRASS corpus: *Video-Gegensprechanlage* ‘video interphone’.

In the field of speech technology, attempts have been made to classify different varieties of German automatically by using exclusively prosodic features. Hagmüller (2001), for example, suggested an approach which makes use of a set of features including those from Fujisaki's model (e.g., Fujisaki, 1983) and the F0 percentiles. With this set of prosodic features, he reached an accuracy of 70% in differentiating read speech produced by speakers from Germany and Austria.

Materials

The speech material used is taken from the recently collected Graz corpus of Read And Spontaneous Speech (GRASS). In total, it contains 62 phonetically balanced sentences, 20 commands elicited with pictures and one hour of free conversation between friends or family members produced by 38 speakers of Austrian German. As the corpus was collected with speech technology applications in mind, it fulfills the requirements for automatic processing: the recordings took place in a soundproof studio with both head-mounted and large-membrane microphones at 48kHz. The orthographic transcriptions were created with PRAAT and they contain detailed annotations of hesitations, disfluencies and other vocal and non-vocal noises. The speakers are between 20 and 60 years old, with a balanced number of male and female speakers. They grew up in one of the eastern provinces of Austria (Salzburg, Upper and Lower Austria, Vienna, Styria, Burgenland, east of Carinthia), but they all lived in Graz at the time of the recordings. All speakers were either university students or have completed university-level education.

From the GRASS corpus, we extracted all utterances with words containing the vowel [a] in a predominantly sonorant environment with a nasal or lateral as the onset consonant of the target syllable. In order to obtain a sufficiently large data set, we included a number of syllables with an obstruent, mostly a velar stop in the coda (e.g. in *machen* [maxən] 'to make' or *maximal* [maksima:l] 'maximal') or as the onset of the subsequent syllable (e.g. Lage [la:gɛ] 'situation'). The target syllables used in this study are *la, skla, ʃla, lam, ma, man, maʃ, max, tsvan, van, vant*.

The syllables occurred in different positions, both with respect to the target word and the utterance. At the same time, they varied with respect to word stress and pitch accent. The syllable /man/, for instance, appeared in the monosyllabic function word *man* 'one' which is always unstressed and unaccented, in the target words *Mantel* [ˈmantəl] 'coat' and *Mannschaft*

[ˈmanʃaft] ‘team’ where it is stressed and potentially also accented, and in *Hilfsmannschaft* [ˈhilfsmanʃaft] where word stress is on the previous syllable. Another example is the syllable /lam/ with word stress and potential pitch accent in the target word *Lampe* [‘lam.pe] ‘lamp’, unstressed in *Stehlampe* [ˈʃte.lam.pɛ] ‘floor lamp’ and with secondary stress in *Nachttischlampe* [ˈnaxt.tiʃ.ˌlam.pe] ‘beside lamp’. Thus one more important aspect that distinguishes our data from the data used in previous studies is the occurrence of primary and secondary stresses due to the fact that our data includes words that are longer than three syllables, as opposed to the two-syllable words investigated in previous studies.

Some of the utterances had to be excluded because of disfluencies or creaky voice, leaving 575 utterances from 37 different speakers for analysis.

Methods

Labelling of the data and a preliminary acoustic analysis was partly done in an advanced Phonetics/Phonology class at the Department of Linguistics (Graz University) taught by the first author in collaboration with the second author. After this first pilot study, the authors of this paper decided on the final design of the experiment and checked and unified the old labels and added new ones. The annotation procedure will be described in more detail below.

Acoustic Parameters

In this study, we extracted various spectral tilt measures suggested in the literature to identify those parameters which contribute to predicting different phonological categories. The acoustic parameters that were automatically extracted from the labelled sound files were: utterance duration; vowel duration; vowel formants (F1, F2, F3); spectral tilt measures, i.e. the difference in amplitude between the first and the second harmonic (H1-H2), and the difference in amplitude between the first harmonic and the amplitude peaks in the vicinity of the first three formants (H1-A1, H1-A2, H1-A3).

The target words, syllables and segments were labelled with Praat 5.3.57 (Boersma & Weenink, 2013) and formant and spectral tilt values as well as duration was extracted using a script developed and made available to us by Bert Remijsen. Formants were calculated at the midpoint of the vowel [a], using ‚To formant (burg)‘ and a formant tracker. Parameters were set according to speaker sex. Spectral tilt measures were also calculated by the same script. Duration values were determined based on normalized speech rate, for which the number of syllables per second within the utterance containing the test word free of pauses were calculated (cf. Torreira & Ernestus, 2009).

ANNOTATIONS

In addition to the measurements, the data was annotated by the third, fourth and fifth author, three undergraduate participants of the course after an intensive training period. Cross-annotation of the whole data set was carried out by the first author, a phonetically trained linguist. The annotated variables were the following:

Annotated variables	Values
(1) position of target word in phrase	initial, medial, final
(2) number of syllables in utterance	
(3) number of syllables in target word	
(4) Syllable Type	open, closed
(6) Word Stress (WS)	yes, no
(7) Pitch Accent (PA)	yes, no
(8) Metrical Stress (MS)	primary, secondary, none
(9) Perceived Prominence (PP)	primary, secondary, none
(10) Tonal Contour	LH, HL, H, L

Table 1: Structural and phonetic variables used in the annotation of the data.

To facilitate the annotation process and to achieve higher inter-rater reliability we annotated only *intonation phrases (IP)* instead of differentiating between an IP and a smaller phrase type, such as *prosodic phrase* or *intermediate phrase* (variable 1). The annotation of variables (1-6) was straightforward. A consonant following a lax (short) /a/ was either a coda consonant as in *Mantel* [mantəl] *'coat',* or it was classified as ambisyllabic as in *machen* [maxən] *'to make'*. Thus, syllable type (4) fully correlated with tenseness (length), as the data did not contain any word final syllables with a distinction between a the two vowels (such as Beet [be:t] and Bett [bɛt], for example.

In annotating variable (7) we faced a number of problems. In class we had started out with a German ToBI (GToBI) annotation of the pitch accents. Given the heterogeneity of our data, however, it soon turned out that GToBI is neither easy to implement, nor were the available categories sufficient to describe what we actually found in the data. For instance, the many compound words in the data seemed to contain a second strong prominence (see Figure 1 above) with or without tonal correlates, which in GToBI would not be analyzed as a pitch accent, but at best as a secondarily associated phrase tone, i.e. a phrase accent (cf. Grice et al., 2000). The considerable number of very long words also implied that we were confronted with a higher number of prominences per word than just one metrically strong position of 'word stress' would predict. We thus decided to keep *word stress* and *pitch accent* as binary variables, following the guidelines of GToBI for the annotation of the latter and included three more variables (8-10).

Metrical Stress (8) is a structural variable including secondary stress positions as illustrated in Figure 1 above. *Primary* and *secondary stress*, however, are phonological categories that do not necessarily coincide with the actually perceived prominences, nor with their phonetic correlates. Qualitative analysis of the data provided clear evidence of such mismatches, such as the target word *Parkanlagen* which sometimes contained two intuitively equal prominences. Looking at the f0 trackings corroborated our perceptual impression that both prominences often correlated with a clearly identifiable pitch rise.

Due to these facts, we decided to add to the structural variables (6-8) two phonetic ones, one on a purely auditory basis (9) and the other one on an auditory and acoustic basis (10). An informal perceptual discrimination task with three prominence values (very prominent, prominent and not prominent) was carried out by the four annotators individually, but at the same time, in a normal classroom upon listening to each utterance twice without headphones. A target syllable was classified as 'very prominent' when it was perceived as the strongest element within the phrase. A syllable was classified as 'prominent' when it was clearly prominent, but less prominent than the strongest element. Note that it was not always possible to identify the strongest syllable in a domain, sometimes more than one position in an IP were classified as strongest. However, as expected, complete consensus among the four annotators was not reached in all cases, but only in 42%. The best results were obtained in cases involving the undoubtedly unstressed function word *man* 'one', followed by cases of nuclear accents at the end of rather short IPs that could be easily identified as 'strongest prominence' of the IP. A common value was assigned to the consensus cases plus the ones that involved three equal annotations. As a next step, we looked at the remaining cases again in a second listening session and determined a common value.

For the tonal annotation, we decided to differentiate between a rising (LH), falling (HL), low (L) and high (H) contour as identified within the vowel segment in the f0 track. A tone was classified as rising or falling as opposed to level on a combined auditory and acoustic basis, i.e. when a movement was perceptible and when the difference between the highest and lowest points was at least approx. two semitones. A level tone was identified as low when it was estimated to be within the lower 25% of a speaker's pitch range (cf. the GToBI guidelines; Grice et al., 2005).

RESULTS AND DISCUSSION

Statistical Analysis

For our estimation of how strongly word stress, pitch accent, metrical position and perceived prominence contribute to predict duration, vowel quality and spectral tilt, we decided to use mixed-effects linear regression models. Previous work on acoustic correlates of accent has mainly used ANOVAs (e.g., Slujter & van Heuven, 1996; Heldner, 2003), which is an appropriate statistical method for controlled production experiments. In our case, however, we draw the data from a large corpus of a large number of speakers and the target syllables thus appear in different words in different contexts. Jaeger (2008) has shown, that for such data, mixed effects models are more reliable than ANOVAs. Mixed effects models allow to incorporate random effects, which cover the variation originating from different speakers or different word types. Furthermore, the independent variables can either be continuous or discrete.

For building the mixed effects models, we used the *languageR* library for the R software for statistical computation (Dalgaard, 2002). In all models, interactions between the variables were tested both two-way and as a set. Furthermore, significantly correlating factors were orthogonalized (i.e., when two variables correlated, one variable was substituted by the residuals of a linear model in which it was predicted by the other variable). Finally, predictors and interactions that were not statistically significant were removed from the models (i.e., backward selection). We did so by using the *step-function*, which is part of the *lmerTest* library. Furthermore, we tracked the AIC value (Akaike information criterion): The lower the AIC value of a model, the better it represents the data.

The models presented in the results section are those models with the lowest AIC value and only contain the remaining significant predictors.

In the following sections, we present the final mixed effects regression models with each of our acoustic measures (Duration, F1, F2, F3, H1-H2, H1-A1, H1-A2 and H1-A3) as dependent variable and word type and speaker as random factors. We included the independent variables shown in Table 2 into the models to estimate their effects on the acoustic measures.

1. Random factors	
1.1 Word	man, Wanderer, Klimaanlage, Mantel, etc.
1.2 Speaker	37 different speakers
2. Independent variables	**Values**
2.1 Gender	male, female
2.2 Position of target word in phrase	initial, medial, final
2.3 Phrase-boundary after target	yes, no
2.4 Position of syllable in word	1st, 2nd, 3rd...
2.5 Number of syllables in target word	min = 1; max = 9; mean = 2.54
2.6 Speech rate	min = 0.021; max = 0.719; mean = 0.203
2.7 Syllable type	open, closed
2.8 Tonal contour	LH, HL, H, L
3. Prominence-related variables	**Values**
3.1 Word Stress (WS)	yes, no
3.2 Pitch Accent (PA)	yes, no
3.3 Metrical Stress (MS)	primary (1), secondary (2) , unstressed (0)
3.4 Perceived Prominence (PP)	main (1), secondary (2), none (0)

Table 2: Random factors, independent variables and accent and stress related variables used in the mixed effects linear regression models.

WHICH ARE THE PREDICTORS FOR VOWEL DURATION?

RESULTS

In this section, we present our findings concerning the predictors for vowel duration. Table 3 provides an overview of the mean vowel durations for the different syllable types.

As expected, vowels in accented syllables are longer than in unaccented syllables. Also vowels in open syllables are longer than in closed syllables, pointing to a significant difference in duration between tense and lax vowels in the data.

	open	closed	WS yes	WS no	MS 1	MS 2	MS 0	PA yes	PA no	PP 1	PP 2	PP 0
DURATION												
Total	0.14	0.08	0.11	0.08	0.11	0.11	0.06	0.11	0.08	0.12	0.11	0.07
Phrase medial	0.13	0.07	0.10	0.07	0.10	0.10	0.06	0.11	0.07	0.11	0.10	0.06
At boundary	0.14	0.10	0.12	0.11	0.11	0.12	0.10	0.12	0.11	0.12	0.12	0.09

Table 3: Average vowel durations separately for syllables at a phrase boundary and in phrase medial position and separately for the different prominence measures (WS = Word Stress, PA= Pitch Accent, MS = Metrical Stress, PP = Perceived Prominence).

In order to estimate which of the observed durational differences are statistically significant and which are the predictors for the observed differences, we fitted a mixed-effects linear regression model with Speaker and Word as random factors for the complete set of tokens (Model DUR, N= 575). The dependent variable was Duration and the independent variables were those presented in Table 2. Table 4 shows the final model containing only the significant predictors.

Predictor	**beta**	**t-value**	**p-value**
Intercept	0.081	15.47	< .0001
Percept. prominence (1)	0.016	5.33	< .0001
Percept. prominence (2)	0.014	5.40	< .0001
Tonal contour (HL)	0.008	1.56	> .05

Tonal contour (L)	0.009	1.09	> .05
Tonal contour (LH)	0.012	4.52	< .0001
Speechrate	0.052	3.49	< .0001
Phrase boundary (no)	-0.010	-4.29	< .0001
Syllable Type (open)	0.025	6.36	< .0001
GenderM	-0.007	-3.00	< .001

Table 4: Significant predictors of the model for Duration (DUR) which reached the lowest AIC (= -2786).

As expected, our model shows that vowel duration is affected by speech rate, that vowels are significantly longer in open than in closed syllables, at phrase end (mean_dur = 0.11) as opposed to a phrase-medial position (mean_dur = 0.08) (cf. Table 3) . Also the effect of tonal contour is going into the expected direction: vowels associated with a rising movement (LH) are significantly longer than the reference values for high level tones (H). Interestingly, however, a downward movement does not have the same effect. This suggests that rising and falling movements have different functions regarding accentuation in the data. While a rise more frequently than not cues a pitch accent (N=92 vs. N=17), a falling contour is equally often just interpolating between a higher and a lower target (N=37).

Somewhat unexpectedly, we found an effect of gender: men (mean_dur = 0.08) tend to produce shorter /a/ vowels than women (mean_dur = 0.10), and this even though the overall speechrate is significantly lower for men (mean = 0.19) than for women (mean = 0.21, t= 3.09, p < .01).

Having incorporated all of these factors, which affect vowel duration, we can perform the analysis concerning the main question of this section: Is duration an acoustic correlate of stress and accent? For this purpose, we added the variables Pitch Accent (model DUR_PA), Word Stress (model DUR_WS), Metrical Stress (model DUR_MS) and Perceived Prominence (model DUR_PP) separately to the model DUR (we did so separately because they are highly correlating). We found that Pitch Accent (beta = -0.006, t = -2.27, p < .01), Word Stress (beta = -0.009, t =-2.75, p < .001), and Perceived Prominence (primary - none: beta = 0.016, t = 5.33, p < .0001; secondary - none: beta = 0.014, t = 5.40, p< 0.001) significantly predict vowel duration. Metrical Stress (secondary - primary: beta =

-0.010, t = -1.94, p < .05, none - primary: beta = -0.007, t = -1.53, p > .05) appeared to be only a marginally significant predictor for vowel duration. Looking at the estimate (beta) and the significance levels suggests that Perceived Prominence has the greatest contribution to the model for vowel duration. To further investigate whether Perceived Prominence is the strongest predictor, we compared the AIC-values of each model and found that indeed, the AIC value of the model DUR_PP is the smallest, and thus that this model is the one best fitting the data. ANOVA analysis of the models DUR_WS, DUR_MS, DUR_PA and DUR_PP showed significantly that DUR_PP is the best model (AIC = -2786, p < .0001 in ANOVA with DUR_WS, DUR_MS and DUR_PA). To conclude, our statistical analysis suggests that Perceived Prominence is (among the prominence variables discussed here) the strongest predictor for vowel duration.

Since the word *man* ‘someone’ is relatively frequent in our data set, we fitted the models DUR_PP, DUR_PA, DUR_MS and DUR_WS again for all tokens but *man* (N = 427). The resulting models have the same significant predictors as those of the complete data set. We may thus safely conclude that Perceived Prominence is the strongest predictor for vowel duration and Metrical Stress the weakest in this subset as well.

Finally, we performed additional analyses on a subset of targets that do NOT carry a pitch accent (N = 332) in order to investigate whether there still is an effect of Word Stress, Metrical Stress and/or Perceived Prominence on vowel duration in these cases. Interestingly, we found no effects of Word Stress (beta = -0.003, t = - 0.78, p > .05) or Metrical Stress (secondary - primary: beta = 0.001, t = 0.16, p > .05, none - primary: beta = 0.002, t = 0.3, p > .05) on vowel duration in this subset of the data. Perceived Prominence, however, proved to significantly affect vowel duration also in this subset of tokens. That means that tokens that do not carry a pitch accent but are perceptually prominent (whether with main or secondary prominence) are significantly longer (mean_dur = 0.13, beta = 0.026, t = 4.63 , p < .0001 for primary prominence and mean_dur = 0.11, beta = 0.016, t = 5.37, p < .0001 for secondary prominence) than the unstressed tokens (mean_dur = 0.07).

DISCUSSION

Although all accentual parameters (Word Stress, Pitch Accent, Metrical Stress and Perceived Prominence) had a significant effect on vowel duration, the results of the regression analysis show that Perceived Prominence is by far the best predictor for vowel duration in our data.

This fact was further corroborated by the second analysis that excluded all tokens that carry a pitch accent (according to the conventions of GToBI). In this subset of the data neither Metrical Stress nor Word Stress proved to have a significant effect on vowel duration.

This last result suggests two conclusions: First, contra to findings in a number of recent studies, albeit on different languages, (Slujter & Van Heuven, 1996a, 1996b; Jessen et al., 1995; Prieto & Ortega-Llebaria, 2006; Astruc & Prieto, 2006), but in line with the results of a perception study by Huss (1978) for English, duration does not seem to be a reliable correlate of Word Stress in the absence of Pitch Accent. It has to be noted, however, that we only investigated vowel duration while many of the cited studies refer to syllable duration. It is thus conceivable that duration measurements of the surrounding onset and coda consonants would yield a different result. But note that Perceived Prominence does have a significant effect on vowel duration, even in the absence of Pitch Accent, which brings us to the second conclusion to be drawn from our results, namely that duration is a correlate of actual prominence, whether brought about by a pitch accent or not and whether related to primary stress or not.

WHICH ARE THE PREDICTORS FOR F1, F2 AND F3?

RESULTS

In this section, we present our findings concerning the predictors for the formant values F1, F2 and F3. Table 5 provides an overview of the mean formant values for the two syllable types. As expected, our data shows that the average value for F1 is higher in open syllables than in closed ones, given that the a-vowel in closed syllables is usually less open and less constricted (Moosmüller 2007) than in open syllables. F2 and F3 show the opposite effect which indicates that a-vowels in open syllables are usually further back or more constricted than the a-vowel in closed syllables. Furthermore, all prominence parameters have a strong raising effect on F1 and a weak lowering effect on F2 and F3. Although our measurements yield slightly higher values than the ones given by Moosmüller, they are essentially in line with her measurements as far as the direction of the differences is concerned (Moosmüller 2007: 114-115).

	open	closed	WS yes	WS no	MS 1	MS 2	MS 0	PA yes	PA no	PP 1	PP 2	PP 0
F1 F	907	820	925	774	928	878	715	939	772	944	906	732
M	718	667	736	615	745	701	574	751	623	758	705	610
F2 F	1497	1525	1494	1539	1490	1533	1542	1496	1534	1488	1501	1550
M	1236	1307	1281	1304	1277	1281	1320	1277	1304	1294	1264	1312
F3 F	2796	2955	2873	2950	2880	2875	2984	2872	2946	2855	2889	2971
M	2465	2624	2584	2597	2590	2491	2653	2570	2606	2587	2552	2622

Table 5: Average formant values (in Hz) separately for male (M) and female (F) speakers for the different prominence measures (WS = Word Stress, PA = Pitch Accent, MS = Metrical Stress, PP = Perceived Prominence).

In order to determine which of the observed differences of the formant averages are statistically significant and in order to find the predictors for the observed differences, we fitted mixed-effects linear regression models for the complete set of tokens (N = 575) with Speaker and Word as random factors separately for each of the formant values F1, F2, or F3 as the dependent variables. The independent variables were again those presented in Table 2. Table 6 shows the final model for F1 containing the significant independent variables.

Predictors	**beta**	**t-value**	**p-value**
Intercept	822.1	49.10	< .0001
Tonal contour (HL)	-25.7	-1.38	> .05
Tonal contour (L)	-18.1	-0.99	> .05
Tonal contour (LH)	73.2	3.27	< .001
Gender (M)	-128.8	-7.56	< .0001
Syllable type (open)	66.2	3.88	< .0001
Gender (M) : Syllable Type (open)	-67.5	-3.37	< .0001

Table 6: Significant predictors of the final model for F1 (lowest AIC value of 2748). Variables linked with a ‘:’ show a significant interaction.

Contrary to vowel duration, F1 is neither significantly affected by speech rate, nor by the position within the phrase. F1 is significantly higher in open than in closed syllables (cf. Table 5), and this effect is larger for female speakers (significant interaction of Syllable Type with Gender). The mean F1 in vowels with a rising contour (LH) (F1_mean = 849) is significantly higher than in vowels with a level H-contour (mean_F1 = 725), but low level (L) (mean_F1 = 758) and falling (HL) (mean_F1 = 783) are not significantly different from high level (H).

Next, we performed our analysis to see whether F1 is an acoustic correlate of stress and accent. As we did for Duration, we added the stress and accent related variables Pitch Accent (model F1_PA), Word Stress (model F1_WS), Metrical Stress (model F1_MS) and Perceived Prominence (model F1_PP) separately to the model shown in Table 6. On the basis of their beta and the AIC value of the model, we evaluated which of these measures predict F1 best. We used *ANOVA* to determine whether the models differ significantly from each other. We found that all of the stress and accent measures significantly predict F1: Pitch Accent (beta = -42.5, t =-3.68, p<0.0001), Word Stress (beta = -49.1, t = -3.88, p<.0001), Metrical Stress (secondary - primary: beta = -28.4, t = -1.66, p<.05, none - primary: beta = -24.2, t = -3.800, p>.0001) and Perceived Prominence (primary - none: beta = 50.5, t = 3.96, p<.0.0001; secondary - none: beta = 31.9, t = 2.81, p<.0.001). Of these models, F1_PP reached the significantly (p > .0001) highest AIC value (F1_PP_AIC = 6742) and thus is the model which best fits the data (F1_PA_AIC = 6755, F1_WA_AIC = 6753, F1_MS_AIC = 6750). We thus assume that F1 is better predicted by Metrical Stress and Perceived Prominence than by Word Stress and Pitch Accent.

In a next step, we performed additional analyses on a subset of targets that do NOT carry a pitch accent (N = 332). We found that in this subset, Perceived Prominence still is a predictor for F1: vowels which are perceptually prominent (either main or secondary) have significantly higher F1 (F1_mean = 791 Hz) than those which are not prominent (F1_mean = 672 Hz). In this subset of tokens, Word Stress and Metrical Stress do not significantly predict F1.

In order to determine the predictors for F2, we fitted a model with Word and Speaker as random effects and the same independent variables as for F1. We found that neither Word Stress, nor Metrical Stress, nor Pitch Accent, nor Perceived Prominence are significant predictors for F2. The only significant predictors were, as expected, Gender (t = -7.33, p <.0001), Syllable Type (t = -3.08, p<.001), but also Tonal Contour (HL - H: t = -2.28, p< .01; L-H: t = -3.13, p<.0.001; LH - H: t = -2.05, p<.01).

The results for F3 are nearly identical to those for F2. For F3, the only significant predictors were again Gender (beta = -234.17, t= -5.81, p <.0001), Syllable Type (beta = -45.62, t = -7.18, p<.0001) and Tonal Contour (HL - H: beta = -36.43, t = -4.59, p< .0001; L-H: beta = -37.41, t = -4.89, p<.0001; LH - H: beta = -31.10, t = -2.56, p<.001). The models for F3 and F2 only differ in that the random effect Word does not significantly contribute to the F3-related model, while it does to the F2-related model.

Discussion

As the results reported in this section show, our data reproduce the well-known effects of Gender and Syllable Type on all formant values (cf. Table 5). The prominence-related measures Word Stress, Metrical Stress, Pitch Accent, and Perceived Prominence all had a highly significant effect on F1, but not on F2 and F3. This result suggests that vowel quality is not directly related to prominence. It rather is in line with results of other studies investigating vocal effort as a measure of increased communication distance in German (Traunmüller & Eriksson, 2000) or as a strategy used in German Lombard speech (Kirchhuebel, 2010) or as a correlate of linguistic prominence in Austrian Standard German (Moosmüller, 2007). Jessen et al. (1995) conducted a production task for German using read minimal pairs. They likewise found a significant effect of stress on F1 in a-vowels in their data, but also on F2. This difference between their results and ours may be due to the different type of data used (minimal pairs vs. heterogeneous targets; word lists vs. longer utterances) or due to the different language varieties used in these studies.

Furthermore, our results corroborate the findings of Moosmüller (2007) relating to the difference in vowel formants across stress conditions. Moosmüller (2007: 211) reports a significant difference in F1 between primary stress, secondary stress and no stress for almost all speakers, whereas the results concerning the effect of secondary stress on F2 and F3 are more diverse. As illustrated in the following ranking, adapted from Moosmüller (ibid.), our results show the same cline as hers, i.e. F1 is significantly higher in primary than in secondary stress position and again in secondary stress position than in unstressed position (calculated both for Metrical Stress and Perceived Prominence; significant differences are indicated by boldface type), whereas the other two formants only show a clear tendency, however without the differences reaching significance level.

F1 primary stress > F1 secondary stress > F1 unstressed
F2 primary stress < F2 secondary stress < F2 unstressed
F3 primary stress < F3 secondary stress < F3 unstressed

These results are interesting in at least two ways. Firstly, they show that secondary stress indeed may have a phonetic manifestation. Secondly, they suggest that metrical stress positions are only potentially prominent. As already shown for Duration, actually perceived prominence is again a better predictor of F1 than metrical stress, which is supported by the higher significance values for PP than for MS (cf. results section above).

Finally, we have to note that we only investigated the low a-vowels (in open and closed syllables) for which it has been suggested that they do not significantly differ in terms of quality features.

WHICH ARE THE PREDICTORS FOR THE SPECTRAL TILT MEASURES?

Results

In the following section, we present our analysis concerning the predictors for the variance observed for spectral tilt measures in the same conditions as discussed in the previous sections. Table 7 shows an overview of the average spectral tilt values for the different conditions.

		open	closed	WS yes	WS no	MS 1	MS 2	MS 0	PA yes	PA no	PP 1	PP 2	PP 0
H1-H2	F	0.34	2.67	1.33	2.69	1.60	0.60	3.66	1.46	2.52	1.61	0.87	3.18
	M	-0.69	-1.16	-0.98	-1.15	-1.01	-0.95	-1.19	-1.01	-1.11	-0.85	-0.83	-1.33
H1-A1	F	-7.79	-1.74	-4.76	-2.05	-4.86	-6.09	0.56	-5.28	-1.84	-5.87	-5.45	-0.11
	M	-7.52	-1.69	-4.14	-1.51	-4.12	-6.10	0.80	-4.55	-1.61	-5.16	-3.95	-0.62
H1-A2	F	-0.91	4.05	-1.9	5.99	-1.29	0.36	9.30	-1.69	5.99	-1.56	-1.51	8.51
	M	-0.37	4.45	0.79	6.32	-0.52	1.49	8.58	-1.12	6.14	-1.15	1.93	7.49
H1-A3	F	13.53	19.92	15.77	20.37	15.85	14.80	23.61	15.16	20.53	15.11	14.86	22.67
	M	15.24	22.32	19.06	22.78	19.48	17.27	24.98	18.19	22.85	16.43	20.20	24.15

Table 7: Average values for the spectral tilt measures (in db), separately for male (M) and female (F) speakers for the target syllables broken by WS = Word Stress, PA = Pitch Accent, MS = Metrical Stress, PP = Perceived Prominence.

In order to determine which independent variables significantly predict the observed differences of the spectral tilt measures, we fitted mixed-effects linear regression models for the complete set of tokens with Speaker and Word as random factors separately for each of the spectral tilt measures: H1-H2, H1-A1, H1-A2 and H1-A3. The independent variables were again those presented in Table 2. Since the variables Number of Syllables and Pitch Accent highly correlated, we orthogonalized them and used the residuals of Number of Syllables predicted by Pitch Accent instead of the original variable Number of Syllables.

RESULTS: H1-H2

The final model for H1-H2 revealed that neither Word Stress, Metrical Stress, Pitch Accent nor Perceived Prominence were significant predictors for H1-H2. As expected, Gender proved to be a significant predictor (beta = -3.21, t= -3.80, p < .0001), with lower H1-H2 values for male than for female speakers. Furthermore, the interaction between (the residuals of) Number of Syllables of the token and the position of the syllable within the word resulted to be a significant predictor (beta = -0.30, t = -4.3, p < .0001). Finally, we also found that Tonal Contour significantly predicts H1-H2 (HL - H: t = -1.29, p> .05; L-H: t = -2.77, p < .001; LH - H: t = -2.99, p< .001). The mean H1-H2 values are lowest for LH (mean = -0.47db), second lowest for L (mean = 0.16db). H (mean = 1.83db) and HL (mean = 0.61) do have the highest average H1-H2 values. This is an expected result, given that LH and L contours are very low in the speakers' pitch range and H and HL values both occur in a higher register.

RESULTS: H1-A1

We fitted a model with the same independent variables as we did for H1-H2 for the spectral tilt measure H1-A1 (= dependent variable). The model which best fitted the data (AIC = 3509) is shown in Table 8.

Predictors	**beta**	**t-value**	**p-value**
Intercept	-0.22	-0.23	> .05
Metrical Stress (2)	1.70	1.92	< .05
Metrical Stress (0)	-2.04	3.07	< .001
Tonal Contour (HL)	-0.01	-0.01	> .05

Tonal Contour (L)	0.59	1.08	> .05
Tonal Contour (LH	-2.67	-3.72	< .001
Syllable Type (open)	-3.57	-5.78	< .001
Number of Syllables	-1.12	-4.34	< .0001

Table 8: Significant predictors of the final mode for H1-A1 (AIC=3509).

For the spectral tilt measure H1-A1, Syllable Type resulted to be a significant predictor: the mean of H1-A1 for open syllables (H1-A1_mean = -7.68db) is significantly lower than the mean for closed syllables (H1-A1_mean = -1.67db). Furthermore, the number of syllables of the word resulted to be a significant predictor: the longer a word, the lower its average H1-A1 value (e.g., monosyllabic words: mean = 1.45db; 6-syllabic words: mean = -5.33db). As for all acoustic measures investigated so far, the tonal contour also significantly predicts H1-A1: H1-A1 is the lowest for LH (mean = -6.79db) and for HL (mean = -3.38db). The difference between L (mean = -2.27db) and H (mean = -1.73db) is not significant.

In the next step, we added our different stress and accent related variables separately (H1-A1_PP, H1-A1_PA, H1-A1_MS, H1-A1_WS) to the model and tested the effect sizes and AIC values for the four resulting models in order to be able to estimate which of the variables predicts H1-A1 best (following the same procedure as explained in the previous sections on Duration and F1). We found that the model with Metrical Stress has the lowest AIC value and best predicts the variation found for H1-A1 (see Table 8). Also Perceived Prominence resulted to significantly affect H1-A1 (primary - none: beta = -1.69, t =-2.69, p <.0.001; secondary - none: beta = -1.35, t = -2.32, p < .0.01, AIC = 3512), as did Word Stress (beta = 1.44, t = 2.56, p<.01, AIC = 3516) and Pitch Accent (beta = 1.35, t = 2.02, p < .0.01, AIC = 3518). An ANOVA of the models revealed that the one with Pitch Accent is the worst predictor for the variance found for H1-A2 (p < .001). The differences were significant.

In order to analyze, whether H1-A1 is also an acoustic correlate for Word Stress, Metrical Stress and Perceived Prominence in the absence of a pitch accent, we reran the models on the subset of tokens which do not carry pitch accent (N = 332). On this subset, neither PP, WA nor MS resulted to significantly affect H1-A1.

RESULTS: H1-A2

The model which best fitted the variation found for H1-A2 (AIC = 3509) is shown in Table 9. Unsurprisingly, this model revealed similar effects of Syllable type, Number of Syllables, position of the syllable in the word and Tonal Contour as the models for H1-H2 and H1-A1.

Predictors	**beta**	**t-value**	**p-value**
Intercept	3.37	1.86	0.0646
Perceived Prominence (1)	-4.10	-5.05	7.34e-07 ***
Perceived Prominence (2)	-3.07	-4.21	3.17e-05 ***
Tonal Contour (HL)	-0.03	-0.04	0.9725
Tonal Contour (L)	-0.58	-0.91	0.3621
Tonal Contour (LH)	-3.95	-4.66	4.08e-06 ***
Syllable Type (open)	-1.92	-2.31	0.0227 *
(residuals of) Number of Syllable	-0.95	-1.39	0.1714
Position of Syllable	1.85	0.86	0.0337 *
No.Syll : Pos.Syll	-0.31	-2.49	0.0145 *

Table 9: Significant predictors of the final model for H1-A2 (AIC = 3509). Variables linked with a ':' show a significant interaction.

The model with Perceived Prominence (shown in Table 9) significantly better models the data (p <.0001) than the models with Word Stress (beta = 2.70, t = 3.246, p <.001), Metrical Stress (secondary - primary: beta = 1.34, t = 0.88, p>.05, none - primary: beta = 2.58, t = 2.38, p>.01) or Pitch Accent (beta = 2.46, t = 3.22, p<.001) as independent variable.

As a next step in our analysis, we tested whether WS, MS or PP affect H1-A2 in the subset of tokens which do not carry pitch accent (N = 332). Whereas Word Stress and Metrical Stress did not prove to significantly affect H1-A2, Perceived Prominence did. Both, syllables carrying main perceptual prominence (H1-A2_mean = -1.38db, beta = -5.59, t = -2.71, p <.001) and secondary perceptual prominence (H1-A2_mean= 0.10, beta = -2.83, t = -2.89, p<.001) have significantly lower H1-A2 values than not prominent syllables (H1-A2_mean = 8.03).

Results: H1-A3

Finally, we fitted a model to estimate the predictors for H1-A3. The resulting best model (AIC = 3810) is shown in Table 10. Like all previous spectral tilt measures, also H1-A3 is predicted by Syllable Type, Number of Syllables, and Tonal Contour.

Predictors	beta	t-value	p-value
Intercept	24.72	19.253	< .0001
Perceived Prominence (1)	-4.04	-4.800	< .0001
Perceived Prominence (2)	-2.78	-3.591	< .0001
Tonal Contour (HL)	1.21	1.298	> .05
Tonal Contour (L)	0.59	0.846	> .05
Tonal Contour (LH)	-2.56	2.753	< .001
Syllable Type (open)	-2.77	-3.210	< .001
Number of Syllables	-1.08	-4.086	< .0001

Table 10: Significant predictors of the final mode for H1-A3 (AIC = 3810)

Similarly to our results on H1-A2, all prominence related variables significantly affect H1-A3, but the model with Perceived Prominence (shown in Table 10) is a substantially better fit for the data (AIC = 3810, p < .0001) than the models with Word Stress (AIC = 3818, beta = 3.34, t = 4.23, p < .0001), Metrical Stress (AIC = 3823, secondary - primary: beta = 2.45, t = 1.84, p < .05, none - primary: beta = 2.39, t = 2.32, p < .01) or Pitch Accent (AIC = 3819, beta = 3.06, t = 4.08, p < .0001) as independent variables.

Next, we tested whether WS, MS or PP have an effect on H1-A3 also in the absence of pitch accent (N = 332). Again, we observe similar tendencies as for H1-A2: Both Word Stress and Metrical Stress do not significantly affect H1-A3. Perceived Prominence, however, is a significant predictor for H1-A3, and the resulting model also is significantly better than the others (tested with ANOVA, p < .001). In contrast to H1-A2, where both the difference between main perceptual prominence and no perceptual prominence as well as the difference between secondary prominence and no prominence was significant, for H1-A3 only the difference between main perceptual prominence and no perceptual prominence proved significant (beta = -5.29, t = -2.57, p < .01).

Discussion

Our analysis shows that amplitudinal differences between the lower and higher spectral partials are clear correlates of prominence (the difference between the two adjacent values H1 and H2 was not significant), and H1-A2 and H1-A3 are better than H1-A1, i.e. a flatter tilt indicates more energy in the higher frequency bands (around F2 and F3) and steeper tilt indicates less energy in the higher frequency bands. This has been shown by many studies to be a clear correlate of vocal effort and thus also linguistic prominence (stress/accent). In that respect, the results of our analysis are well in line with the findings of all other different studies cited in the Introduction. However, our results differ from the findings of those studies that identified clear acoustic correlates for word stress in the absence of a pitch accent (Slujter & Van Heuven, 1996a, 1996b; Astruc & Prieto, 2006; Prieto & Ortega-Llebaria, 2006; Gordon & Nafi, 2012). Our results are, however, in line with the findings by Campbell & Beckman (1997) and Sadeghi (2007) who also did not find any evidence for the hypothesis that spectral tilt is a correlate of word stress.

Similarly to the conclusions drawn for Duration, our spectral tilt results also point in the direction that spectral tilt is a correlate of actual prominence. Although our results are not fully comparable to the results of other studies as we investigated longer words that contained potential secondary accents, we may still conclude that neither primary stress position (i.e. Word Stress), nor secondary stress position itself is a good predictor for a flatter or even negative spectral tilt. In the absence of a pitch accent, we not only found no significant effects of Word Stress, but also not of Metrical Stress, neither for H1-A2, nor for H1-A3, whereas Perceived Prominence proved to be a highly significant predictor of spectral tilt.

Among the two significant parameters, H1-A2 seems to be the better diagnostic, yielding even higher significance and effect sizes than H1-A3, although this measure is also very highly significant. Furthermore, the H1-A2 even permits to distinguish between main prominence and secondary prominence, whereas H1-A3 only permits a binary distinction between prominent and not prominent.

GENERAL DISCUSSION AND CONCLUSIONS

Summarizing the results of this study, we found that duration and spectral tilt are strong correlates of prominence in our data of Standard Austrian German. We also found that prominence has a strong effect on F1, whereas the effect on F2 and F3 was considerably smaller. This is most probably due to the fact that we only investigated syllables with an a-vowel as the nucleus. It is a well-established fact that the difference in quality between the two low vowel phonemes is much smaller than within the high and mid vowel spaces. It has also been suggested that, contrary to the situation with high and mid vowels, length rather than the tense/lax opposition is the distinctive feature between low vowels in German, which is corroborated by our data.

The conclusions drawn from these preliminary results of the investigation of acoustic correlates of ‘stress’ and ‘accent’ in Standard Austrian German allow us to give a preliminary answer to the questions posed in the introduction.

- Does stress have a phonetic reality distinct from pitch accent?

  In a sense: **yes** - if we take the question to mean: Is there a phonetic reality of prominence that can be measured acoustically and that is not brought about by pitch variation. In another sense: **no** - if we understand the question as whether the two structural categories *Word Stress* and *Phonological Pitch* accent have different phonetic realizations manifested in different phonetic cues.

- Are the acoustic correlates associated with stress related to ‘loudness’ and thus compatible with the traditional notion of *expiratory/ dynamic stress*?

  Again the answer is not straightforward. As we have already pointed out, we do not find any evidence for the differentiation between stress and accent on the basis of the phonetic cues associated with them. That means that the answer as posed in the introduction based on the work of the different studies cited there, is **no**.

  However, our results from the spectral tilt measures also clearly show that even without the contribution of fundamental frequency variation perceived prominence does have acoustic correlates, among which energy distribution over the spectrum is highly important. Our analysis revealed that pitch movement is not necessarily accompanied by other acoustic measures that could be shown to correlate with the prominence measures, which means that pitch change on its own does not necessarily lend prominence to a segment or a syllable. It rather seems that this pitch change has to be carried out voluntarily and be accompanied by other acoustic features, such as duration and spectral tilt. This may be interpreted as supporting the view that pitch, at least in English or German, is not necessarily 'prominence-lending', but rather 'prominence-cueing' (Ladd 2008: 54). This fact also speaks in favor of the assumption of an expiratory/dynamic stress as suggested in the traditional view. In that sense, the answer is clearly **yes.**

- Are the results compatible with the view that prominence is a uni-dimensional system and stress is just a structural position?

The answer is **yes**. The results of this study clearly show that it is the actually perceived prominence that affects the acoustic parameters rather than some kind of phonological entity. In other words, the results support Bolinger's distinction of word stress as an abstract structural position that may be made prominent by accentuation. The results also give evidence of the view that prominence is uni-dimensional and graded. We found acoustic evidence for three degrees of prominence, especially among the spectral tilt measures, but we also found duration and pitch to be strong correlates of prominence. The upshot of these facts is that prominence is a perceptual phenomenon reflecting articulatory vocal effort the acoustic correlate of which is a "bundle of phonetic features" (Peters et al., 2005).

ACKNOWLEDGEMENTS

The work of Barbara Schuppler was funded by a Hertha-Firnberg grant (T572-N23) from the Austrian Science Fund (FWF). We like to thank Bert Remijsen for his friendly advice and for providing us with the Praat script we used for most automatic measurements and Hans Grassegger for discussions of the acoustic measurements and finally the students of a Phonetics/Phonology class (winter 2013/2014) for annotating some of the variables.

REFERENCES

Astruc, Lluisa and Prieto, Pilar 2006. Stress and accent: Acoustic correlates of metrical prominence in Catalan. *ITRW on Experimental Linguistics, 28-30 Aug 2006, Athens, Greece*.

Beckman, Mary. 1986. *Stress and non-stress accent*. Dordrecht: Foris Publications.

Beckman, Mary, Edwards, Jan. 1994. Articulatory evidence for differentiating stress categories. In Keating, P.A. (ed.), *Papers in Laboratory Phonology* Cambridge: Cambridge University Press, 3, 7-33.

Boersma, Paul and Weenink, David 2013. Praat: doing phonetics by computer [Computer program]. Version 5.3.57, retrieved October 2013 from http://www.praat.org

Bolinger, Dwight. 1958. A theory of pitch accent in English. *Word 14, 109-149.*

Bürkle, Michael. 1995. Österreichische Standardaussprache: Vorurteile und Schibboleths. In: Rudolf Muhr, Richard Schrodt, and Peter Wiesinger (eds). *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen,* Vienna: Verlag Hölderer-Pichler-Tempsky, 236–249.

Campbell, Nick and Beckmann, Mary. 1997. Stress, prominence, and spectral tilt. *ESCA Workshop on Intonation: Theory, Models and Applications,* 67-70.

Dalgaard, Peter 2002. *Introductory Statistics with R*. New York: Springer.

Fry, Dennis Butler. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustic Society of America* 27, 765-768.

Fry, Dennis Butler. 1958. Experiments in the perception of stress. *Language and Speech* 1, 126-152.

Fujisaki, Hiroya. 1983. Dynamic characteristics of voiced fundamental frequency in speech and singing. *The production of speech*. Berlin: Springer.

Gordon, Matthew and Nafi, Latifa. 2012. Acoustic correlates of stress and pitch accent in Tashlhiyt Berber. *Journal of Phonetics* 40, 706-724.

Grice, Martine, Ladd, D. Robert, and Arvaniti, Amalia. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17/2, 143–185.

Grice, Martine, Baumann, Stefan, and Benzmüller, Ralf 2005. German Intonation in Autosegmental-Metrical Phonology. In: Jun, Sun-Ah (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, 55-83.

Hagmüller, Martin. 2001. *Recognition of Regional Variants of German using Prosodic Features*. Master Thesis. Graz University of Technology.

Heldner, Mattias. 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of spectral emphasis in Swedish. *Journal of Phonetics* 31, 39-62.

Hornung, Maria and Roitinger, Franz. 2000. *Die Österreichischen Mundarten: Eine Einführung*. Vienna: Öbv & hpt VerlagsgmbH & Co KG.

Huss, V. 1978. English word stress in post-nuclear position. *Phonetica* 35/2, 86-105.

Iivonen, Antti 1996. Unterschiede des Vokalismus und der Intonation in der ostmitteldeutschen und wienerdeutschen Standardsprache. In: Eva-Maria Krech, Eberhard Stock, Ursula Hirschfeld, and Lutz Christian Anders (eds.), *Deutsches Ausspracheworterbuch.* Berlin/New York: de Gruyter, 225-240.

Jaeger, Tim Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59, 434–446.

Jessen, Michael, Marasek, Krzyzstof, Schneider, Katrin and Claßen, Kathrin. 1995. Acoustic Correlates of Word Stress and the tense/lax Opposition in the Vowel System of German. *Proceedings of the 13th International Congress on Phonetic Sciences. Stockholm,* 428-431.

Kirchhuebel, Christin 2010. The effect of Lombard speech on vowel formant measurements. *Meeting of the Acoustical Society of America*, Cancun, Mexico (poster presentation)

Klaaß, Daniel. 2008. Untersuchungen zu ausgewählten Aspekten des Konsonantismus bei österreichischen Nachrichtensprechern. Duisburger Papers on Research. *Language and Culture* 74, 7–277.

Ladd, D. Robert. 2008. *Intonational phonology* (Cambridge studies in linguistics 119), 2nd edition. Cambridge [u.a]: Cambridge Univ. Press.

Moosmüller, Sylvia. 1998. The process of monophthongization in Austria (reading material and spontaneous speech), *Papers and Studies in Contrastive Linguistics*, 1998, 9–25.

Moosmüller, Sylvia. 2007. *Vowels in Standard Austrian German*. An Acoustic-Phonetic and Phonological Analysis. Habilitation thesis.

Moosmüller, Sylvia. 2011. Aussprachevarianten in Österreichischen Standarddeutsch. In: Ines Bose, and Baldur Neuber (eds.), *Interpersonelle Kommunikation: Analyse und Optimierung.* Frankfurt: Peter Lang, 193-199.

Moosmüller, Sylvia and Dressler, Wolfgang U. 1988. Hochlautung und soziophonologische Variation in Österreich. *Jahrbuch für Internationale Germanistik 20/2, 82–90.*

Moosmüller, Sylvia and Ringen, Catherine O. 2004. Voice and aspiration in Austrian German plosives. *Folia Linguistica* 38, 43–62.

Moosmüller, Sylvia, Schmid, Carolin, and Brandstätter, Julia. 2015. Standard Austrian German. *Journal of the International Phonetic Association* 45/2.

Peters, Benno,.Kohler, Klaus J., and Wesener, Thomas. 2005. Melodische Satzakzentmuster in pro*sodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion. („Pitch accent patterns in prosodic phrases of German spontaneous speech ? Statistical distribution and linguistic function“).* In Klaus J. Kohler, Felicitas Kleber, and Benno Peters (eds.), *Prosodic Structures in German Spontaneous Speech* (AIPUK 35a), Kiel: University of Kiel, 7–54.

Prieto, Pilar and Ortega-Llebaria, Marta 2006. Stress and Accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance. *Proceedings of Speech Prosody*, Dresden, Germany, 337-340.

Sadeghi, Vahedi. 2011. Acoustic correlates of lexical stress in Persian. *ICPhS XVII*, 1738-1741.

Schuppler, Barbara, Hagmüller, Martin, Morales-Cordovilla, Juan and Pessentheiner, Hannes. 2014. *GRASS: the Graz corpus of Read and Spontaneous Speech.* Accepted for presentation at LREC, Iceland.

Slujter, Agaath. 1995. *Phonetic correlates of stress and accent*. Dordrecht: Foris.

Slujter, Agaath, and van Heuven, Vincent. 1996a. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustic Society of America* 100(4), 2471-2485.

Slujter, Agaath, and van Heuven, Vincent. 1996b. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSLP,* 96. Philadelphia: Applied Science and Engineering Laboratories, Alfred I. DuPont Institute, 630-633.

Sweet, Henry. 1906. *A Primer of Phonetics*. Oxford: Clarendon.

Torreira, Francisco, and Ernestus, Mirjam. 2009. Probabilistic effects on French [t] duration. *Proceedings of the 10th Annual Conference of the International Speech Communication Association,* 448-451.

Traumüller, Hartmut, Eriksson, Anders. 2000. Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America* 107/6, 3438-3451.

Wiesinger, Peter. 2009. Die Standardaussprache in Österreich. In: Eva-Maria Krech, Eberhard Stock, Ursula Hirschfeld, and Lutz Christian Anders (eds.), *Deutsches Aussprachewörterbuch*. Berlin/New York: de Gruyter, 229-258.