Acoustic correlates of linguistic stress and accent in Dutch and American English

Agaath M.C. Sluijter[1] en Vincent J. van Heuven[2]

[1] KPN Research, Groningen, the Netherlands
[2] Holland Institute of Gen.Linguistics / Phonetics lab. Leiden University

1. INTRODUCTION

In the literature the same acoustic correlates of stress and accent have been established for Dutch and English, i.e. F0 movement, duration, intensity and vowel quality. Sluijter and Van Heuven (1996a) showed that F0 movement and overall intensity in Dutch differentiate only between accented and non-accented syllables, rather than between stressed and unstressed. The most reliable acoustic correlates of stress were duration and high-frequency emphasis. Vowel quality differed significantly only in lexical items, but was only a weak correlate in reiterant speech copies. In this study we reconsider the acoustical correlates of stress and accent in American English (AE) and compare the results with the Dutch results. We offer an analysis of the discriminating strength of the parameters in an attempt to optimally distinguish initial and final stressed tokens by machine, using LDA.

Sluijter and Van Heuven (1996b) determined F0 contours and formant values, and examined spectra of stressed and unstressed vowels spoken with and without an accent using AE minimal stress pairs and their reiterant speech-copies. The results show that accent-lending F0 movements only occur on focused targets, whereas non-focused targets were only realized with minor F0 changes, probably due to segmentally conditioned intonation and declination.

As for vowel quality, they found that stressed vowels are characterized by a fuller vowel quality than unstressed vowels. Furthermore, focused constituents (marked by a pitch accent) have a fuller vowel quality compared with vowels in unfocused constituents.

They also investigated differences due to stress in the glottal vibration pattern, inferring glottal parameters directly from the audio signal radiated from the mouth. The inferences of the glottal parameters indicate that glottal pulses are more sinusoidal in unstressed syllables: high-frequency emphasis is weaker, indicating smoother and slower vocal fold closing movement. Focused constituents had more high-frequency emphasis than their unfocused counterparts. Counterintuitively, glottal leakage, reflected in the bandwidth of F1, was found to be larger for stressed than for unstressed vowels. This effect was independent of accentuation. Accented stressed vowels are additionally characterized by a considerable increase in the amplitude of voicing (AV) and a slightly increased open quotient (OQ).

In order to compare *all* the acoustic correlates of stress and accent in AE, we will study the following parameters: (1) F0, (2) duration, (3) overall intensity, (4) source parameters (components of spectral balance): OQ, AV, closure rate/skewness of the glottal pulse, glottal leakage, and (5) filter parameters, i.e. F1 and F2′ as acoustic correlates of vowel quality.

The parameters 1, 4 and 5 were investigated in Sluijter and Van Heuven (1996b). In addition, we need to measure duration and overall intensity in the same corpus. Both duration and overall intensity have been reported as reliable acoustic correlates of stress in AE (Beckman, 1986 and references mentioned there). However, Sluijter and Van Heuven (1996a) showed that in Dutch overall intensity appeared to be a correlate of accent rather than stress. Therefore, we investigate whether overall intensity would still provide a reliable acoustic correlate of stress in AE. Previous research on AE was generally hampered by covariation of stress and accent. In our AE corpus unfocused targets were not pronounced with an accent-lending pitch movement, so that its influence is removed. Of course, we need not be surprised if intensity variations should turn out to provide only a marginal stress cue in AE as well. Duration, on the other hand, proved to be reliable in Dutch even out of focus. We expect it to be a reliable correlate in AE as well.

A great deal of research has concerned the acoustical realization of stress and the relative strength of these parameters in separating stressed from unstressed tokens (Beckman, 1986). Much of this research, however, suffered from covariation of accent and stress. Now that it has been proven that F0 is not a reliable correlate of stress, it seems reasonable to include other parameters. Given the effects of stress on the glottal parameters, we predict that a combination of these parameters should yield a more successful separation than overall intensity and F0. Moreover, we will investigate whether the separation of initial and final-stressed tokens on the basis of glottal parameters in AE is also better than on the basis of vowel quality, since it is widely held that AE is more sensitive to vowel reduction than Dutch. We also expect that the glottal parameters and duration are close in strength as predictors of linguistic stress in line with the Dutch results. Finally, we will examine whether the relative strengths of the five stress correlates interact with the presence versus absence of a pitch accent. Once we have established the hierarchy of stress correlates in AE, we will be able to compare Dutch and AE as to the importance of each of the correlates

2. METHODS

2.1 Speech corpus and measurements

We used the existing speech corpus consisting of four noun-verb minimal stress pairs and three different reiterant speech-copies. Targets were produced with and without focal accent in fixed carriers by six AE speakers (Sluijter and Van Heuven, 1996b). *Overall intensity* values were obtained by first multiplying the speech signal with a 40 ms Hamming window for male speakers and a 25.6 ms Hamming window for female speakers and then

computing a 512-pt DFT. All measurements were made at the F1 maximum in each target syllable, i.e., when the mouth is maximally open. Speech intensities were expressed in dB_{SPL} (determined by using a reference tone).

Syllable boundaries were determined by visual criteria described in Slujter (1995) using the oscillograms and the spectrograms of the words

2.2 Statistical analysis

Duration and overall intensity
To examine the effects of stress and accent on duration and overall intensity, we ran five-way analyses of variance on the measurements in both speech conditions with *focus condition*, *stress*, *sex* and *vowel type (word type*: for lexical speech data) as fixed factors and *speaker* as a random factor nested within *sex*. Word type, repetition and syllable position were used as repeated measures in the reiterant speech condition. Repetition was used as repeated measure in the lexical condition. Syllable duration and overall intensity were used as the dependent variable in the analyses. Since the main interest is directed towards differences due to stress and/or accent, we will not discuss differences due to speaker, sex and vowel type in this paper.

Statistics to determine the strength of the acoustic cues
Following Beckman (1986), from each of the different types of acoustic measurement some sort of ratio was derived for each token. For the F0 measurements the location of the F0 peak was used to determine the stress position of each token.

 The ratio for the duration values was the difference in ms between the two syllables per token, computed as the logarithm of the quotient of the two syllable durations (in ms), as in equation (1):

$$\text{log duration ratio} = \log (\text{duration}_{\sigma 1} / \text{duration}_{\sigma 2}) \qquad (1)$$

The ratio will be positive for tokens with longer initial syllables and negative for tokens with shorter initial syllables.

Comparable ratios for the various intensity measurements were computed by simply subtracting the measured value (in dB) of the second syllable from that of the initial syllable. Negative values indicate an increase in value from the first to the second syllable and positive values a decrease, as shown in the following formulae (all values are in dB):

$$\text{overall intensity ratio} = \text{intensity}_{\sigma 1} - \text{intensity}_{\sigma 2} \qquad (2)$$
$$\text{OQ ratio} = (H1^*\text{-}H2^*)_{\sigma 1} - (H1^*\text{-}H2^*)_{\sigma 2} \qquad (3)$$
$$\text{AV ratio} = H1^*_{\sigma 1} - H1^*_{\sigma 2} \qquad (4)$$
$$\text{tilt A2 ratio} = (H1^*\text{-}A2^*)_{\sigma 1} - (H1^*\text{-}A2^*)_{\sigma 2} \qquad (5)$$
$$\text{tilt A3 ratio} = (H1^*\text{-}A3^*)_{\sigma 1} - (H1^*\text{-}A3^*)_{\sigma 2} \qquad (6)$$

In order for the glottal leakage ratio, i.e. F1 bandwidth ratio, to be of comparable type, it was computed as the logarithm of the quotient of the two bandwidths in Hertz:

$$\text{log bandwidth ratio} = \log (B1_{\sigma 1} / B1_{\sigma 2}) \qquad (7)$$

In order to obtain a relative measure for the F0 synchronization, we used the location of the F0 peak relative to the word-internal syllable boundary (in ms). Negative values indicate that the peak is located prior to the internal boundary, i.e. in the first syllable, positive values that it is located in the second syllable.

The vowel quality ratio was derived by computing the euclidean distance of each vowel to schwa as:

$$\text{distance} = \sqrt{((F1\text{-}F1_{neutral})^2 + (F2'\text{-}F2'_{neutral})^2)}. \qquad (8)$$

$F1_{neutral}$ was set to 4.86 and 5.77 Bark for male and female speakers, respectively. $F2'_{neutral}$ was set to 12.01 and 13.03 Bark for male and female speakers, respectively. A perceptually meaningful ratio for vowel quality could be computed by subtracting the computed distance of the second syllable from that of the first syllable, because the measurements were in Bark. Negative values indicate that the first syllable is more reduced than the second syllable, positive values indicate that the first syllable is less reduced, as shown in the following formula:

$$\text{distance}_{\sigma 1} \text{ (in Bark)} - \text{distance}_{\sigma 2} \text{ (in Bark)}. \qquad (9)$$

To determine how each of the acoustic measures can be applied to determine stress position, we carried out linear discriminant analyses (LDA) on the reiterant speech data in each focus condition separately. In all analyses the stress positions functioned as groups: *'bVbV* vs. *bV'bV*, with 144 data points (6 speakers * 8 repetitions * 3 vowels) per group. In the first analysis, we compared the strength of a combination of these parameters with the other known correlates of stress: F0, duration, overall intensity and vowel quality. In the second analysis we compared the strength of each of the glottal parameters: OQ (i.e. $H1^*\text{-}H2^*$), glottal leakage (B1), closure rate/skewness of the glottal pulse ($H1^*\text{-}A2^*$ and $H1^*\text{-}A3^*$) and AV ($H1^*$). Since not all the glottal parameters were measured for the vowel /i/ (i.e. OQ and B1), we determined the capacity to separate stress positions of each acoustic correlate and each glottal parameter (if possible) with and without the data of this vowel. The latter analysis consisted of 96 data points per group.

3. RESULTS ACOUSTIC ANALYSIS

3.1 Duration

In table 1 mean absolute syllable durations (in ms) are broken down by speech type, focus condition and stress position.

As can be seen in table 1 stressed syllables are generally longer than unstressed syllables [lexical: $F(1,338)=68.6$, p=.001, $\eta^2=.07$; reiterant: $F(1,1140)=64.2$, p=.001, $\eta^2=.3$]. Syntagmatically, in the lexical speech condition, the initial stressed words had longer unstressed syllables due to the segmental structure of the words. However, when compared paradigmatically, stressed syllables are indeed longer than unstressed ones.

The presence or absence of an accent lengthens the duration of both stressed and unstressed syllables. Accented words ([+F]) have longer syllables than unaccented words ([-F]) [$F(1,340)=9.2$, p=.038, $\eta^2=.03$; reiterant: $F(1,1140)=42.1$, p=.003. $\eta^2=.15$]. The temporal contribution of an accent is an almost linear time expansion of the entire word. Similar results were found for Dutch (Sluijter, 1995).

There was no significant interaction between focus and stress for the lexical data [$F(1,336)=7.2$, n.s.], but there was a significant interaction in the reiterant speech condition: accented stressed syllables lengthen somewhat more compared to their unstressed counterparts than unaccented stressed syllables [$F(1,1138)=78.6$, p=.001, $\eta^2=.01$].

We conclude that there is a considerable effect of *stress* on the syllable duration. An accent also affects syllable duration. The lengthening effect of an accent affects both the stressed and the unstressed syllable.

focus	stress	lexical			reiterant		
		σ1	σ2	ΔS	σ1	σ2	ΔS
[+F]	initial	**220** (65)	274 (55)	-54	**230** (36)	190 (31)	40
	final	160 (42)	**319** (71)	159	171 (26)	**238** (32)	67
	ΔP	60	45		59	48	
[-F]	initial	**184** (50)	236 (48)	-52	**189** (26)	157 (21)	32
	final	147 (44)	**272** (54)	125	159 (28)	**198** (33)	39
	ΔP	37	36		30	41	

Table 1: Mean syllable duration (in ms) of stressed and unstressed syllables per focus condition (-F and +F). Standard deviations in parentheses. The differences are presented syntagmatically (ΔS) and paradigmatically (ΔP).

3.2 Overall intensity

In figure 1 means of overall intensity (in dB_{spl}) of stressed and unstressed syllables are given per focus condition.
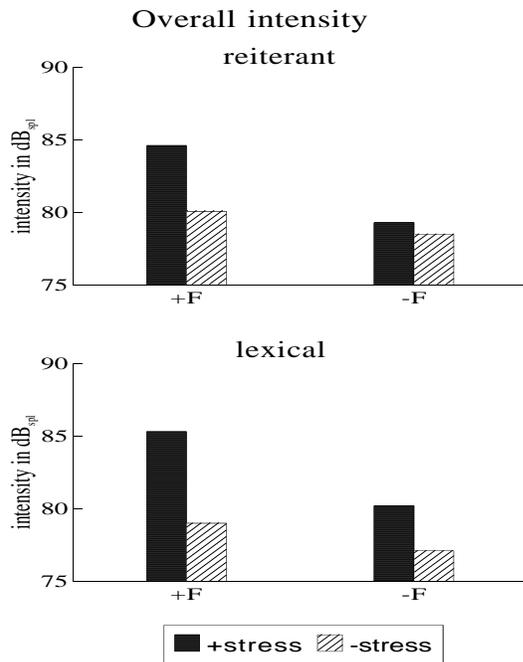
Figure 1: Mean overall intensity (in dB_{spl}) for stressed and unstressed vowels in condition [+F] and [-F].

As can be seen in figure 1 stressed syllables are characterized by an increased overall intensity [reiterant: F(1,1122)=59.9, p=.002, η^2=.06; lexical: F(1,185)=238.6, p$\leq$.001, η^2=.16]. Focus condition also led to an increase of overall intensity [reiterant: F(1,1122)=66.5, p=.001, η^2=.09; lexical: F(1,185)=52.7, p=.002, η^2=.09]. However, as can be seen in the table, the overall intensity of accented syllables increases more due to stress, than unaccented stressed syllables. The interaction between *focus* and

stress, however, was only significant for the reiterant speech data [reiterant: F(1,1120)=19.5, p=.012, η^2=.03; lexical: F(1,183)=4.6, p=.099, η^2=.02]. We conclude that there is a small effect of *stress* (1-3 dB) on overall intensity, but that there is a considerably larger effect of accent.

4 STRENGTH OF STRESS CORRELATES

We examined the capacity each of the acoustic correlates of stress in AE as an acoustic separator between initial and final stressed words for each focus condition separately. Moreover, we would like to know if the relative strengths of the five types of stress correlate interact with the presence versus absence of a pitch accent. In a LDA the differences between the two syllables of each of the parameters were used as the predictors to separate initial and final stressed *reiterant* tokens. Lexical tokens were not used in this analysis.

In figure 2 we compare the percentages of correct discrimination by LDA for each of the known correlates of stress and accent.

It can be observed that in condition [+F] vowel quality is the poorest correlate of stress position. A combination of the glottal parameters is a reliable correlate of stress, even stronger than F0 peak location and duration. As expected, overall intensity and F0 peak location perform reasonably well in condition [+F]. However, as was mentioned above, the higher overall intensity and the location of the F0 peak can be explained by the fact that in condition [+F] a rise-fall configuration, marking the accent on the stressed syllable, is realized on the stressed vowel. Therefore, overall intensity and peak location are only reliable correlates of accent. These two correlates lose their strength in condition [-F] in which the true correlates of stress are examined without the confounding influence of accent. Only 60 to 65% correct classification was reached for overall intensity and F0 peak location. As can be seen in figure 2, F0 is the weakest correlate of stress in this condition. Duration and glottal parameters are almost equal in strength. In figure 3 we compare the percentages correct discrimination by LDA for each of the glottal parameters. As can be seen in figure 3, closure rate/skewness of the glottal pulse (H1-A2 and H1-A3) and glottal leakage (B1) are the most reliable predictors of stress.

In condition [-F] OQ and AV hardly contributed to the correct separation. These measures were shown to be correlates of accent rather than stress. When the glottal parameters are combined, 100% correct separation is reached in condition [+F], and

89% in condition [-F]. When OQ and AV are omitted, only 79% correct separation is reached. They do contribute to correct separation in combination with more powerful discriminators.

5. CONCLUSIONS AND DISCUSSION

This study examined the hierarchy of the acoustic correlates of stress and accent in AE. In line with the research for Dutch, we were especially interested in comparing the acoustic correlates of stress without the intervening effect of accent. We compared all the known acoustic correlates of stress in two conditions, i.e. with a pitch movement on the stressed syllable and without a pitch movement on the stressed syllable in order to know to what extent the relative strength of the five types of stress correlate interacts with the presence versus absence of a pitch accent. Duration and overall intensity measurements had to be made since they were not available from preceding experiments. The duration measurements confirmed the Dutch results and much earlier research on this topic, i.e. stressed syllables are longer than unstressed syllables. Furthermore, it was found that the lengthening effect of accent affects the entire word.

The measurements of overall intensity supported our hypothesis that overall intensity is a correlate of accent rather than stress. At first sight, of course, these results do not agree with earlier studies on the acoustic realization of minimal stress-pairs in English in which F0 and overall intensity were reported to be reliable acoustic correlates of stress. However, all this research was hampered by the covariation of stress and accent. It is not correct to regard F0 and hence overall intensity as reliable acoustic correlates for stress.

We determined the discriminating strength of the acoustic ccorrelates in an attempt to optimally distinguish initial and final stressed tokens by machine, using LDA. The results do imply a hierarchy of cues to stress: duration, glottal parameters (i.e. high-frequency emphasis and glottal leakage, reflected in B1) and vowel quality are respectively the first, second and third best cue in condition [-F]. F0 and overall intensity have little or no cue-value in this condition.

In condition [+F], i.e. with a pitch accent, F0, overall intensity, OQ and AV are reliable correlates of the accent position. When comparing the results with the results for Dutch (Sluijter and Van Heuven, 1995a), we conclude that AE and Dutch do not differ greatly in the extent to which accent and stress are associated with acoustic patterns. Stress patterns in AE have somewhat more influence on vowel quality than in Dutch.

We conclude that in general the same hierarchy of the correlates is observed for the two languages, with the exception of vowel quality, which assumes a more prominent position in AE.

6. REFERENCES

1. Beckman, M.E. *Stress and Non-stress accent*. Foris, Dordrecht, 1986.
2. Sluijter, A.M.C. *Phonetic correlates of stress and accent*. Foris Dordrecht, 1995
2. Sluijter, A.M.C. and Van Heuven, V.J. "Spectral balance as an acoustic correlate of linguistic stress", *J. Acoust. Soc. Am.*, 1996a (to appear).
4. Sluijter, A.M..C. and Van Heuven, V.J. "Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in American English", *J. Acoust. Soc. Am.*, 1996 (submitted).
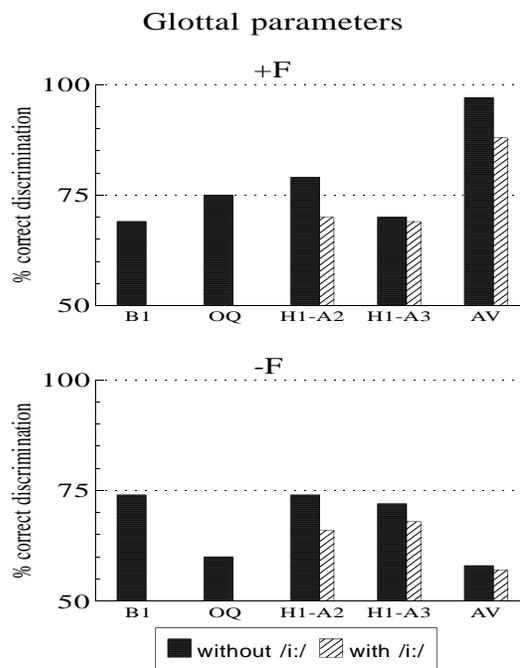
Figure 2: Percentages correct discrimination for each acoustic correlate (including /i:/: hatched bars, excluding /i/ black bars).
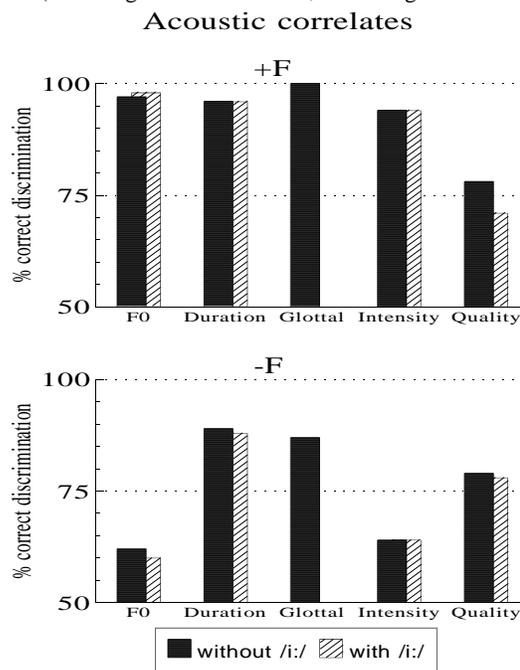

Figure 3: An overview of the percentages correct discrimination for each glottal parameter.