SUPRALARYNGEAL RESONANCE AND GLOTTAL PULSE SHAPE AS CORRELATES OF STRESS AND ACCENT IN ENGLISH

Agaath M.C. Sluijter[1], Stefanie Shattuck-Hufnagel[2], Kenneth N. Stevens[2], Vincent J. van Heuven[1]

[1]Dept. Linguistics/Phonetics Lab., Leiden University/ Holland Institute of Generative Linguistics, The Netherlands
[2]M.I.T., Research Lab. of Electronics, Cambridge MA, USA

ABSTRACT

In a production experiment, it is shown that unstressed syllables have a smoother and slower vocal fold closing movement than stressed syllables. As a result the spectrum of stressed syllables is characterized by an increase in high-frequency emphasis. Accent, but not stress, is additionally characterized by a slightly increased open quotient and an increased amplitude of voicing.

INTRODUCTION

Sluijter and Van Heuven [1,2] showed that high-frequency emphasis is a powerful acoustical and perceptual correlate of linguistic stress in Dutch. They assumed that high-frequency emphasis arises because of the way the vocal folds and the glottis are configured during phonation when producing stressed syllables ("stressed" here refers to the main-stressed syllable of a word; "unstressed" means secondary stress or lower). Increased vocal effort, as is needed to produce stressed syllables, generates a more strongly asymmetrical glottal pulse, vocal fold abduction is faster, the maximum amplitude of movement following the opening time is greater and the closing phase gets shorter, so that the trailing flank of the glottal pulse is steeper. Certain of these differences are manifested in the spectrum at low frequencies, in the vicinity of the lowest three harmonics, whereas other differences modify the spectrum at mid and high frequencies. The low-frequency part of the spectrum is determined by the gross shape of the waveform. The relation between higher harmonics and the lowest few mainly depends on the speed of glottal closure. The faster the glottis closes, the more pulse-like the excitation signal will be, resulting in a harmonic spectrum with an increased tilt at high frequencies [3]. A more gradual pattern of glottal closure, as Sluijter and Van Heuven assumed to be the case for unstressed syllables, however, yields a steeper negative spectral slope (i.e. low-frequency emphasis).

In the present study we tried to replicate these results for American English, examining the possible cause of the high-frequency emphasis in more detail. Is it brought about by a change in the glottal pulse (e.g. longer glottal closure, faster vocal cord adduction) or by a change in the supralaryngeal tract (wider mouth opening, leading to an upshift of F1)? We investigated differences in the glottal vibration pattern for stressed and unstressed syllables, inferring glottal parameters from selected characteristics of the audio signal [3,4].

METHOD

Subjects, material and procedure

Six speakers of American English (three male, three female) produced four noun-verb minimal stress pairs, *'export~ex'port, 'uplift~up'lift, 'compact~com'pact, 'digest~di'gest* as well as their reiterant mimicries */bi:bi:/, /bɛbɛ/ and /ba:ba:/* with and without focal accent in fixed carrier phrases. In (1) an example is given of the condition with a pitch accent on the target (+F), in (2) without a pitch accent (-F), (target words in italics, accent position in bold face).

1. Please produce ***'compact*** for him again.
   Please produce ***'baba*** for him again.
   Please produce ***'bibi*** for him again.
   Please produce ***'bɛbɛ*** for him again.
2. Please produce *'compact* for him **again**.
   Please produce *'baba* for him **again**.
   Please produce *'bibi* for him **again**.
   Please produce *'bɛbɛ* for him **again**.

Each response type was recorded twice, the second time with the items in reversed order. This procedure yielded 640 utterances. Only the initial syllables of the lexical items, <u>ex</u>*port* and <u>up</u>*lift* and the final syllables of *di*<u>*gest*</u> and *com*<u>*pact*</u> were used for further analysis (where underline indicates the syllable that was analysed, in both its stressed and unstressed forms). Of the reiterant items, we will present only the data of the vowel /a:/ and /ɛ/.

Measurements

a) Effects of the filter function of the vocal tract: To control for differences in the shape of the vocal tract due to stress and/or accent, we measured formant frequencies (F1..F3) of stressed and unstressed vowels. Only F1 was used in the present study; the difference, ΔF1, was calculated between F1 for each individual utterance and mean F1 across speakers and conditions, for each vowel (mean F1: /a/ 760 Hz, /ɛ/: 605 Hz, lexical material: 637 Hz).

b) *The open quotient*, expressed as a percentage of the total period, determines the time during which the glottis is open. The primary acoustic manifestation of a narrow glottal pulse, i.e. a decrease in open time, is a reduction of the amplitude of the fundamental in the source spectrum relative to adjacent harmonics [5]. The amplitude difference between the first two harmonics (H1-H2), therefore, is an estimate of the open quotient (OQ). The stronger H1, the larger OQ [3]. H1 and H2 were corrected for the influence of F1, yielding the measure $H1^*$-$H2^*$ (see [3]).

c) Completeness of closure (bandwidth of F1): The amount of minimal flow (i.e. glottal leakage) varies over loudness conditions [6]. Louder voices tend to have a smaller minimal flow than soft voices. We therefore suggest that minimal flow also varies with stress. When the glottis is not closed during phonation, glottal resistance can contribute to energy losses in the F1 range, adding significantly to the F1 bandwidth (B1). B1 is estimated from the amplitude decay rate during the first two cycles of the F1 oscillation. To reduce interference by higher formants, the waveforms were filtered in a 600 Hz frequency band centered around F1.

d) Degree of opening over the entire glottal cycle: The amplitude of F1 (A1) depends on the degree of opening over the entire glottal cycle, i.e., A1 is influenced by OQ and the glottal aperture during the open phase, whereas B1 is not. We measured the difference between $H1^*$ and A1.

e) Skewness of glottal pulse, rate of closure: The spectrum of the glottal waveform at mid and high frequencies is influenced primarily by the abruptness of the glottal closure. There are two ways in which glottal closure can differ. If the closure is nonsimultaneous along the length of the vocal folds, there is a more gradual cutoff of airflow. A more abrupt closure introduces less negative spectral tilt in the higher frequency region. Another way in which glottal closure can differ is related to the duration of the closing portion, i.e. rate of closure (RC), which directly influences the skewness (SK) of the glottal pulse. As the slope of the closing phase gets steeper (keeping OQ constant) the amplitudes at mid and high frequencies increase relative to amplitudes at low frequencies. We derive information on the skewness of the glottal pulse and the rate of glottal closure by taking the difference between the amplitude values of the first harmonic ($H1^*$) and of F2 and F3 (A2 and A3). Both A2 and A3 are corrected for their dependence on F1, and F1 and F2, respectively, yielding the measures $H1^*$-$A2^*$, and $H1^*$-$A3^*$ respectively (see [3]).

f) Amplitude of voicing ($H1^*$): When intensity increases, subglottal airpressure will also increase, which directly influences amplitude of voicing. One of the main acoustic effects of an increase in subglottal airpressure is an increase in H1. We hypothesize that stressed and unstressed syllables differ in the amplitude of the glottal pulse.

g) Overall intensity: In addition we measured the overall intensity value of the stressed and unstressed syllables. We expect no differences between stressed and unstressed syllables in condition -F, whereas we do expect differences in +F.

An overview of the physiological

	Physiology	Acoustics
Filter	a) shape of vocal tract	F1, F2, F3
Source	b) open quotient (OQ)	H1*-H2*
	c) completeness of closure, glottal leakage	B1
	d) degree of glottal opening	H1*-A1
	e) skewness of glottal pulse (SK), duration of closing portion (RC)	H1*-A2* and H1*-A3*
	f) amplitude of voicing (AV)	H1*

dimensions in which glottal pulses of stressed and unstressed syllables can differ, and the acoustic parameters from which these differences can be derived, is presented at the top of this page.

All measurements were made at the F1 maximum in each target syllable, i.e., when the mouth is maximally open. The resulting measures were averaged over speakers and over vowels, both reiterant and lexical. We compared the averaged values of stressed and unstressed vowels *paradigmatically* for each focus condition separately.

RESULTS AND CONCLUSIONS

Figure 1 shows the differences between stressed and unstressed syllables (solid and hatched bars, respectively) in condition +F (in focus, i.e. with a pitch accent) and condition -F (out of focus, i.e. without a pitch accent) of selected parameters. H1*-A1, F2 and F3 were not significantly influenced by accent and/or stress, and are therefore not presented in the figure.

Results indicate that glottal pulses are more sinusoidal in unstressed syllables: high-frequency emphasis (SK and RC) is weaker, indicating smoother and slower vocal fold closing movement. Counter-intuitively, B1 was found to be wider for stressed than for unstressed vowels. This effect is even stronger for accented, stressed syllables. We assume that this effect is caused by the increased subglottal pressure with which stressed syllables are produced. Due to this pressure, the arytenoid cartilages remain abducted throughout the cycle, whereas the glottis is entirely closed over a part of the cycle of vibration when producing unstressed syllables. Accented stressed vowels are additionally characterized by an increased AV (H1*) and a slightly increased OQ (H1*-H2*). The transfer function of the vocal tract differs only in the mouth opening dimension (F1) showing an overall tendency towards greater opening for stressed vowels, irrespective of accentuation.

We investigated to what extent the glottal shape parameters, intensity and ΔF1 by themselves could be used to discriminate stressed from unstressed, as well as accented from unaccented vowels in linear discriminant analyses across speakers, conditions and vowels. Table 1 gives an overview of the percentages correct classification.

Table 1 Correct classification (%) of stressed and unstressed syllables (stress), and of accented and unaccented syllables (accent) in condition +F and -F, and across conditions (all), using a supralaryngeal parameter (ΔF1), glottal parameters and overall intensity separately and in combination.

	accent	stress		
	all	+F	-F	all
1. ΔF1	68	69	64	65
2. glottal	88	90	73	75
3. 1+2	88	91	72	75
4. intensity	69	64	53	61
5. 1+2+4	88	90	74	76

Almost 90% correct classification of (pitch) accented syllables was reached using only glottal parameters as predictors. Intensity and ΔF1 are less powerful predictors by themselves; moreover adding them to the glottal parameters did not significantly improve the results.

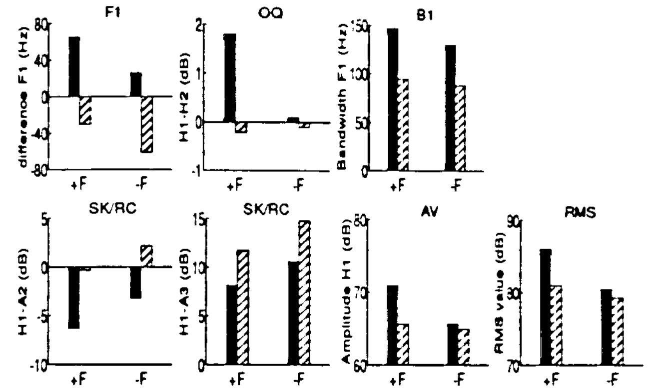

Figure 1. Effects of stressed (solid bars) and unstressed (hatched bars) syllables in condition +F (with pitch accent) and condition -F (no accent) on selected acoustic parameters (see text).

In condition [+F], percent correct classification of accented syllables is somewhat higher than in the above mentioned analyses. The relative strength of glottal versus other parameters, however, is virtually identical.

When separating stressed (both accented and unaccented) from unstressed syllables across conditions, almost 75% correct classification was reached. In this case, just as in condition [-F] (separating stressed, unaccented from unstressed syllables) especially the predictive strength of the glottal parameters decreases. Nevertheless, even in condition [-F], the contribution of glottal parameters always outweighs that of the other parameters, which means that glottal differences are the most reliable correlates of stress.

In this paper we studied the correlates of stress and accent other than F0 contour and duration, i.e. concentrating on the distribution of spectral energy. In this domain of correlates we conclude that accent and stress are mainly characterized by differences in the shape of the glottal pulse, rather than differences in the supralaryngeal tract.

Future research is needed to determine if stressed and unstressed syllables will sound more natural in synthesized speech if we also take aspects of vocal fold vibration and their effect on the spectral balance into account. Also, our results potentially contribute towards more accurate identification of accented and/or stressed syllables in automatic speech recognition systems.

REFERENCES

[1] Sluijter, A.M.C. and Heuven, V.J. van (1995). "Spectral balance as an acoustic correlate of linguistic stress," *J. Acoust. Soc. Am.* (submitted).
[2] Sluijter,A.M.C, Heuven, V.J. van and Pacilly, J.J.A. (1995)."Spectral balance as a cue in the perception of linguistic stress," *J. Acoust. Soc. Am.* (submitted)
[3] Stevens, K.N. and Hanson H.M. (1994) Classification of glottal vibration from acoustic measurements, in O. Fujimura, M. Hirano (eds) *Vocal fold physiology: Voice quality control*, San Diego: Singular, pp. 147-170.
[4] Fant, G., Liljencrants, J. and Lin, Q. (1985), "A four-parameter model of glottal flow", *Speech Trans. Lab. Q. Prog. Stat. Rep. 4*, Royal Institute of Technology, Stockholm, pp. 1-13.
[5] Klatt, D.H. and Klatt, L.C. (1990), "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Am. 87*, pp. 820-857
[6] Holmberg, E.B., Hillmann, R.E., and Perkell, J.S. (1988), "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice", *J. Acoust. Soc. Am. 84*, pp. 511-529.