Notes on the phonetics of word prosody

Heuven, V.J. van; Sluijter, A.M.C.; Goedemans, R.; Hulst, H. van der; Visch, E.

Chapter 7
Notes on the phonetics of word prosody

Vincent van Heuven
Agaath Sluijter

7.1 Introduction

7.1.1 Defining phonetic prosody

By prosody we mean the ensemble of properties of manifested language, e.g. speech, which cannot directly be derived from the properties of the mere sequence of smallest distinctive linguistic units (segments, i.e. vowels and consonants in speech). If, for instance, we consider two phonemically identical segment strings in English, we may notice a difference in pronunciation between these two strings that cannot be pinned down to one single segment; yet this difference is responsible for these phonemically identical segment strings to signal different, and completely unrelated meanings, e.g. *forbear:* /ˈfɔːbɛə/ 'ancestor, forefather' versus /fɔːˈbɛə/ 'to endure'. In this example all the segments belonging to the first syllable are pronounced with greater effort when the segment string means 'ancestor' but the segments in the second syllable are pronounced more forcefully when the segment string is used as the verb 'to endure'.

To give just one other example, this time from a non-western language, the segment sequence /bɑ/ in Mandarin Chinese has different meanings depending on the speech melody with which this syllable is spoken (see figure 1). Again, the difference between these four words cannot be located in any one segment; rather the distinguishing pitch characteristic extends over the entire course of the syllable, and is therefore prosodic in nature.

The term prosody dates back to the days of the ancient Greeks, and is a compound of the words προς 'with' and ωδειν 'to sing', i.e. that which goes together with the singing: the accompaniment. The suggestion being made is that the vowels and the consonants make up the words of the song, and the prosody is the instrumental backing. Obviously, the implication is that the segments and the prosody live on separate tiers, since different words can be sung to the same melody and/or rhythm, and the same words can be sung with different melodies and/or rhythms: prosody cannot be predicted from segmental structure and vice versa.

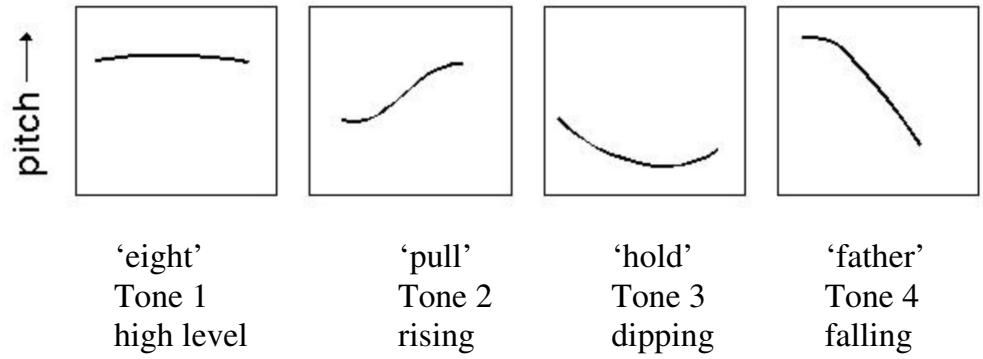


Figure 1: four different pitch patterns on the syllable /bɑ/ in Mandarin Chinese. The meaning of each word is indicated below the fundamental frequency tracing of each word (after van den Berg, 1986).

In most textbooks on phonetics the authors are satisfied with simply listing those properties of speech that can generally be considered to be prosodic in nature, such as:

- melodic organisation: pitch, tone, intonation
- temporal organisation: length, duration, tempo and tempo variation, and pause
- dynamic organisation: loudness and loudness variation, stress and accent

Although such a taxonomy gets us a long way, matters are more complicated when we consider the fact that all individual vowels and consonants have intrinsic melodic, temporal and dynamic properties. Generally, for instance, closed vowels such as [i] and [u] are pronounced – all else being equal – with a higher pitch than an open vowel [a]. Conversely, open vowels naturally have greater intensity than closed vowels, if only because the mouth radiates sound pressure more effectively from an open (horn shaped) mouth than from a closed mouth (which functions as a funnel, stifling the sound). Similarly, open vowels tend to be longer than closed vowels, since opening the jaw widely (as for [a]) and closing it again takes up more time than just to open it a little and close it again (as for an [i] or [u]). Yet, these intrinsic properties, by their very nature, are fully predictable from the identity of the segments, and should therefore be excluded from the realm of prosody. To belabour this point a little further, there is also a class of properties of segments that is referred to by the term "co-intrinsic". This term subsumes those properties of a segment that are fully predictable from the identity of the neighbouring segments. For example, a vowel will last longer when it is

followed by a voiced obstruent than when it is followed – all else being equal – by a voiceless obstruent. Again, this part of the melodic, temporal and dynamic properties of sounds is predictable from the sequence of segments, is therefore no part of prosody. To uncover the true prosody of speech, some form of melodic, temporal, and dynamic decomposition is needed that extracts the inherent segmental properties first, and then factors out the co-intrinsic influences so that pure prosody remains.

7.1.2 Word prosody versus sentence prosody

Prosodic differences between identical segment strings can relate to the meaning of words. If they do, they are part of the system of word prosody. The examples in § 7.1.1 (English and Mandarin) address word prosody, since different words were signaled by differences in stress and tone, respectively. However, prosody may also be used to express differences in meaning of identical word sequences. If, in English, someone says *I beg your pardon* with falling pitch on *pardon*, the utterance expresses an apology ('I am sorry'); if this sentence is pronounced with a rising pitch on the final word, the speaker ostensibly has not understood his interlocutor ('What did you say?').

The present book is focused on word prosodic systems. For this reason we will restrict this chapter to the phonetics of word prosody. And indeed, we will consider the range of phonetic correlates of word prosodic contrasts only. However, we will have to make more than occasional excursions into the world of sentence prosody. The reason for this is that many characteristics of word prosody are latent, and become manifest only, or at least more clearly so, when words are being used in different sentences. As a case in point consider the effect of sentence prosody in English on the prosodic properties of a word embedded within the sentence. The position of the stressed syllable within the target word can be located through purely phonetic comparison of the target when pronounced in a sentence where it is used in contrastive focus with a token of the same word in an utterance where the target is out of focus:

Did you see an AIRPLANE or a CAR?
Did you SEE the airplane or did you HEAR it?

The most striking difference between the focused and the non-focused token of the word *airplane* is that the pitch tracing of the focused exemplar shows a large change in pitch (a rise, a fall or a rise-fall combination) on or quite near the first syllable. This pitch movement is much smaller, or even completely

absent, in the non-focused realisation of this word. When, in a language such as English, a single polysyllabic word is in focus, its stressed syllable will be associated with a conspicuous pitch movement (such a conspicuous pitch movement is often called a pitch accent, cf. Bolinger, 1958). This procedure would allow us to determine the lexical stress position in an arbitrary English word, even if we did not know the language.

In the remainder of this chapter we will discuss various word prosodic constructs developed by theoretical phonologists, and consider possible phonetic correlates of these constructs. The discussion will be limited to culminative (see below) word prosodic phenomena, i.e. to accent-related phenomena.

7.2 Some linguistic concepts

7.2.1 Accent and stress

So far there has been a lot of confusion surrounding the terms accent and stress, such that certain researchers reserve the term accent for phenomena that others call stress and vice versa.

Let us begin then by stipulating that all human languages have a prosodic structure that is hierarchical in nature. The number of levels in the hierarchy may differ across languages, but somewhere near the bottom of the hierarchy there will always be a unit called the syllable (which in turn consists of a vocalic nucleus and zero to several leading and trailing consonants). At some higher level, syllables will be gathered into a larger structure, let us say a word. When a language has the structural property of accent, at some level within the prosodic hierarchy, one constituent is felt (both by linguists and by native speakers) to be stronger than the other constituent(s) at the same hierarchical level. So, in the English word *window* the first syllable /wɪn/ is felt to be stronger than the second /dəʊ/. It is crucial to this definition of accent that the strong-weak relationship is not tied to one level in the prosodic hierarchy. When two (or more) words are combined into a short phrase, for instance, one word will be considered stronger than the other(s). In the English phrase *a clean window* the noun will be strong and the adjective weak.

In more general terms, one can consider the relationship between a strong and a weak constituent at some level in the prosodic hierarchy as an instantiation of head-dependent relationships that pervade human language. Head vs. dependent relationships exist also on levels below the syllable. Since syllables can exist without consonants, but not without vowels, the

vowel is the head of the syllable (for a determination of the vowel's headship at the syllabic level using experimental methods, cf. van Heuven 1994a). Yet it would seem improper to consider individual segments to be the potential bearers of accents: although it is possible to pronounce accents on abutting syllables, it is impossible to have multiple accents within a single syllable. There is general agreement, therefore, that the smallest linguistic unit that can be accented is the syllable[1].

It is crucial to the definition of accent that it is culminative (Trubetskoy, 1958), that is to say, that only one syllable within a word can be the strongest (and that within a word group only one word can be the head). In terms of linguistic structure the strength of the stressed syllable (i.e. the syllable that is typically accented at the word level) is reflected in its dominant character in the phonological organisation of the word. For instance, stressed syllables in Dutch and English allow a richer inventory of syllables (i.e. allowing more, and more different, phonemes to make up a syllable. Other (word) prosodic structures do not have the property of culminativity. Lexical tone, for instance, is not culminative: several syllables within a word can be spoken at either high or low pitch, in any combination, without any degree of prominence being associated with the high pitch in those cases where only one syllable happens to carry a high tone. In this chapter we will be concerned with culminative word prosody, i.e. stress and accent. We will not attempt to survey the phonetic correlates of lexical tone in this chapter.

[1] There are indications, however, that units (i.e. segments) below the level of the syllable can be a focus domain, with measurable and perceptually relevant acoustical correlates in the timing of the accent-lending pitch movement on the syllable containing the subsyllabic focus domain (van Heuven, 1994a).

We have not yet considered by what phonetic properties those units that are called accented, are indeed stronger than those that are weak. All that matters at this juncture is that accent, in principle, has measurable correlates in the phonetic domain. Stress, in contradistinction to the above, is an abstract property of a word, or of a larger prosodic unit, that specifies the default position of accent for the unit concerned. Thus, if the first syllable in *window* is called stressed, this means that in normal circumstances /wɪn/ is strong and /dəʊ/ is weak. However, it is possible to move the accent to the second syllable in the rather more unusual sentence: *I did not say winDY, I said winDOW*. Similarly, it is possible to pronounce the phrase *an old man* with a single accent on the adjective as in *I met an OLD man, not a YOUNG one*. Note that we use the term stress to refer to the default accent position of syllables in words as well as of words in larger units[2]. Generally, it seems that most linguists adhere to the view that stress refers to prominence on the abstract linguistic level, whereas accent bears on the phonetic realisation of prominence in speech; we have no reason to deviate from this majority view.

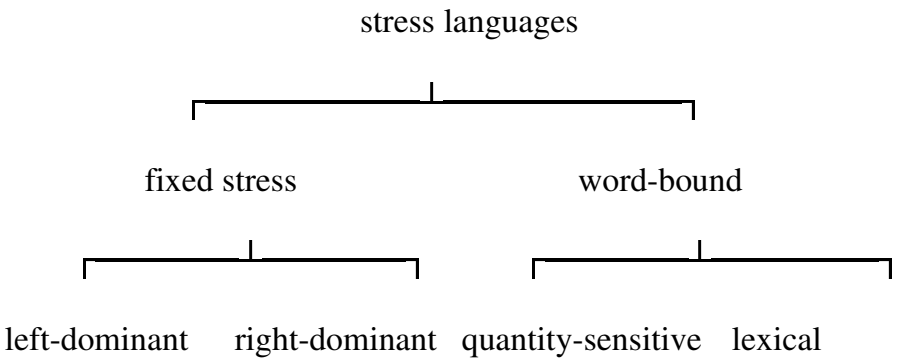
7.2.2 Word stress versus sentence stress

When a language has word accent marking one syllable within a word as stronger than the others, it will also use accent to mark certain words as being stronger than others at the level of the phrase or sentence. Word stress implies sentence stress. The converse, however, does not hold. Many languages use accent to highlight one word over an other, and yet do not use accent at the word level, either because all words are monosyllabic (as in Chinese), or because the accent may fall on any syllable within the word (possibly in Indonesian).

There exists a fairly well developed typology of word stress systems, as is crudely exemplified in the following tree (for a more sophisticated treatment of the subject see van der Hulst, to appear)[3]:

[2] By the default accent position at a particular level in the prosodic hierarchy we mean the so-called integrative accent position (Fuchs, 1984). An accent in this position has the capability to mark not just the accented unit itself as a [+focus] domain (i.e. a linguistic unit that the speaker wants to mark as communicatively important) but also all other units of the next higher constituent in the hierarchy that the accent unit is the head of. Thus the default accent position is semantically ambiguous when considered out of context: an accent in the default position may either indicate that the entire constituent that is headed by it, is in focus (so-called broad focus or integrative focus) or that just the accent unit itself is in focus (narrow focus). For further clarification see van Heuven (1994a, b) and references given there.

[3] Typological trees suggest that distinctions that apply to one branch in the tree, have no relevance in other subdivisions of the tree. This is clearly not the case, so that trees have generally been abandoned in favour of typological parameter systems in order to characterize



Languages with fixed stress have a single rule for stress placement at the word level, e.g. stress is always on the first (or second) syllable of a word (initial or left-dominant stress) or always on same syllable position counting from the trailing word edge (final or right-dominant stress). When the stress may vary from one word to the next, we have two options. In some languages the position of the word stress can be found by quantity-sensitive rules. In such languages (e.g. English and Dutch) the stress goes to the heaviest syllable within a domain at the left or right word edge (see chapter 1); the heaviest syllable is – as a first approximation – the syllable with the largest number of vocalic and postvocalic segments (see chapter 4 for a more extensive treatment of quantity and prominence sensitive stress rules). When no rules (quantity-sensitive or other) can be formulated for stress placement, and yet each individual word has a correct stress position (to the effect that stress marked on any other syllable in the word is judged to be incorrect by native listeners), the language has lexical stress. Dutch and English are, in fact, mixtures of quantity-sensitive and lexical stress. Langeweg (1988) showed that the stress position of some 85% of the Dutch mono-morphemic words is accounted for by a small number of (quantity-sensitive) rules; the remaining 15% is exceptional, hence has lexical stress.

No such typology exists as yet for sentence stress systems.

---

stress placement (cf. Hayes, 1980; van der Hulst, to appear).

7.2.3 Accent levels

Although accent may be a dichotomy at the level of immediate constituents, many degrees of accent are possible at the sentence level. Generally, the strongest accents will be signaled by pitch movements (as well as by a host of other acoustical markers); other, weaker accents lose the pitch movement, but will still bear some (or all) of the non-pitch correlates of accent such as longer duration, greater articulatory precision and loudness.

7.3 Phonetic correlates of accent

In this section we will review the phonetic correlates of accent. Correlates can be located within each of the three areas of phonetics, viz. in production (physiological correlates), in the acoustic signal (acoustical correlates), and in speech perception (perceptual cues).

7.3.1 Physiological correlates of accent

An accented syllable is produced with more physiological effort than its unaccented counterpart. The extra effort may be applied at each or all of the three stages of the speech production process, be it pulmonary, laryngeal or articulatory.

7.3.1.1 Pulmonary

The suggestion has been made that accented syllables are produced by a local contraction of chest muscles and/or the diaphragm so that more air is expelled from the lungs, increasing subglottal air pressure, and as a consequence of this, boosting both the intensity and frequency of the glottal pulse. The evidence for this mechanism is rather weak, however. Ladefoged (1967) reports electro-myographic data showing that generally no muscle contractions take place in the chest area during the production of accented syllables. There is one exception, viz. evidence of local muscle contractions was found coincident with the production of emphatic accent. Generally, collecting electro-myographic data is cumbersome (since it requires the enlistment of surgical help), so that little data is available on this issue.

7.3.1.2 Laryngeal

Extra laryngeal effort would be required in order to produce a louder glottal

pulse. Increasing the tension of the vocal cords may cause one or more of the following effects to occur:

- After that the vocal cords have been pushed open due to a sufficiently large subglottal air pressure (relative to the intra-oral air pressure), air will rush through the glottis at greater speed than would be the case during the production of an unaccented syllable. As a result of both the faster flow (causing the Bernoulli suction effect) and the greater elastic recoil forces in the tensed glottal ligaments, the vocal cords will snap back to their adducted position faster and more forcefully. The acoustic consequence of faster glottal adduction is a boost of the intensity of harmonics at higher frequencies (typically above 500 Hz).
- Due to the increased subglottal pressure the vocal cords will be pushed open again after a shorter time interval than would happen during an unaccented syllable. As a result, the repetition rate of the glottal pulses increases, generating a higher vocal pitch.

Note that we do not claim actively increased vocal intensity, since this would require increased effort of the pulmonary system, which is not normally the case during the production of an accent (see above). The reason that accents correlate with greater overall intensity would seem to be a passive consequence of the more efficient vibratory mode of the tensed vocal ligaments, cutting up the airflow into more, and at the same time sharper, pulses.

At this point it is hard to see how we can increase the speed of the glottal vibration without at the same time affecting the glottal pulse shape, especially the adduction phase. Yet, only stress-accent systems show this crucial correlation of high frequency intensity boost with changes in the repetition rate of the vocal cords. There is no indication that high frequency intensity is correlated with the high tone of Japanese accents (e.g. Beckman (1986) finds no effect of stress on intensity on Japanese minimal stress pairs[4]), and certainly intensity is not necessarily correlated with high tones in tone languages. There are two possible explanations why the correlation exists in stress-accent languages:

- There is actively increased subglottal pressure in normally accented syllables after all (in spite of Ladefoged's findings)
- The supraglottal system is configured during the articulation of an

[4] Recent data by Campbell (1995) and Campbell & Beckman (1995), however, cast doubt on this view; these authors report flatter spectral slopes in accented syllables, especially in spontaneous, i.e. non-read, speech.

accented syllable such that it caused less impedance to the glottal air stream, e.g. due to wider mouth opening. This configuration would result in a larger transglottal pressure drop, causing faster glottal vibration and more forceful glottal adduction.

It will be clear from the above that more research is needed to settle this issue.

As an aside, we have always been intrigued by the temporal accent marker in English that is commonly called aspiration. When an accented syllable in English begins with a voiceless plosive (/p, t or k/ not preceded by tautosyllabic /s/, the onset of voicing is delayed into the following vowel or sonorant by some 100 ms, generating a stretch of voiceless vowel (traditionally called aspiration) or a voiceless sonorant. Apparently, generating a voiceless vowel/ sonorant onset requires effort, mainly at the subglottal stage: voiceless vowel onsets (as well as voiceless /h/, which is the same thing) require an enormous expenditure of air: there will be a substantial loss of subglottal pressure after such a voiceless stretch, since the airflow rushing through the rather open glottis is not impeded in the oral cavity. Therefore there must be sufficient subglottal pressure at the beginning of a stressed syllable for this process to be executed. When the stressed syllable begins with a voiceless fricative followed by a plosive (only /s/ is allowed in this position in English) a lot of the subglottal pressure will have been lost during the articulation of the fricative, so that insufficient pressure is available after this fricative+stop cluster to produce a voiceless sonorant. This, of course, explains why no aspiration of voiceless sonorants is found in English accented syllables beginning with /s/ plus plosive. When the preceding /s/ is not tautosyllabic, there will be a longer time interval between the /s/ and the voiceless vowel onset (tautosyllabic /s/-es are shortened as there are more consonants in the onset cluster (Lindblom, Lyberg & Holmgren, 1981; Nooteboom & Cohen, 1984). Since the transglottal airflow during the production of a fricative is impeded by an obstruction in the oral cavity, the loss of subglottal pressure will soon be stopped; sufficient build-up of subglottal pressure will take place after an allosyllabic /s/ but not after the shorter tautosyllabic /s/.

7.3.1.3 Articulatory

As a consequence of greater articulatory effort, segments are generally pronounced more slowly during the production of an accented syllable, while the articulators approach their target positions more closely (by a process that can be called articulatory expansion). Generally, there seems to be a deceleration of the articulation rate from the onset consonant towards the

accented syllable's midpoint, which is followed by an acceleration from the midpoint towards the trailing edge. As a result, the timing differences between accented and unaccented tokens of the same syllable are located primarily in the vocalic nucleus rather than in the consonants. Although not much data is available on this subject, it seems that the temporal midpoint of the syllable is characterised by the greatest articulatory precision, so that the vocal organs reach their target positions most closely in the middle of an accented syllable. More or less in line with the above, it has been claimed as a characteristic of an accented syllable that articulatory precision (or expansion) spreads out from the temporal midpoint towards the syllable's edges more or less symmetrically. The opening gesture of the mouth at the onset of an accented syllable is faster, and takes place at the earliest possible moment. Similarly, the closing gesture is delayed towards the offset of the accented syllable, and when it takes place, the movement is faster[5].

The articulatory expansion can be observed in the amplitude of the movements of the articulators involved in the production of the accented (vowel) sounds. For instance, the degree of lip protrusion of an accented rounded vowel, say [u], tends to be stronger than that of its unaccented counterpart; similarly, the degree of mouth opening (as evidenced by the amplitude of jaw movement) is greater in an accented [a] than in an unaccented one. We will defer further comments on this mechanism to the discussion on the acoustical correlates of articulatory expansion.

[5] The lengthening effect of stress differs from that of preboundary position. When segments occur in domain-final position (e.g. at the edge of an intonational phrase) they are progressively lengthened such that the segments are stretched more as they occur closer to the boundary; the domain for preboundary lengthening is claimed, for English, to be the preboundary foot; Beckman & Edwards, 1991).

7.3.2 Acoustical correlates of accent

In sharp contrast to the small number of physiological studies there is an abundance of research on the acoustical correlates of accent. The research has concentrated on measuring the reflection of accented or degree of accentedness in five, possibly six[6], (sets of) acoustical parameters: fundamental frequency (F0), duration, intensity, formant frequencies, and spectral balance. The sixth correlate that has received some attention is the duration of aspiration after voiceless plosives in English accented syllables when these are not preceded by a tautosyllabic /s/; this latter correlate, of course, does not qualify as a universal accent marker. We will review each of these suggested correlates in turn in separate sections.

7.3.2.1 Fundamental frequency (F0)

Accented syllables are generally pronounced with a change in the repetition rate of the vocal cord vibration. The repetition rate is commonly expressed in terms of the number of cycles per second, or Hertz (Hz). Changes in repetition rate can either be expressed in absolute terms (Hz) or in relative terms, i.e. as a percentual change or as a musical interval (a 100% change, i.e. a doubling of the frequency, is called an octave, and comprises a range of 12 semitones, i.e. twelve compounded 6% increments. One semitone is the pitch interval that holds between two adjacent keys on a piano keyboard (white to black keys or vice versa)).

For the lower degrees of accent the change is a passive consequence of increased airflow through the glottis during the production of the vocalic nucleus and a decrease of flow at the syllabic edges, caused by the greater impedance to the flow due to the narrowing of the vocal tract during the articulation of a consonant. The result is small rise-fall contour. Such a contour will be present in any syllable, but its excursion size may be a little larger for higher degrees of accent. However, when a stronger degree of

[6] There is at least one study in which a multitude of potential accent correlates were examined, including the durations of subsyllabic events such as the rise time of the intensity envelop, i.e. the duration during which the intensity at the vowel onset rises monotonously (Lehto, 1969). The author found that this parameter, which she called the beat phase of the syllable, is the clearest and most stable correlate of English accent.

accent, typically expressing semantic focus, is being made, the pitch change that is associated with the accent is no longer the passive consequence of faster flow, but is brought about by a voluntary change in the tension of the glottal musculature. The acoustic result may be a rise in pitch, a fall, or a combination of the two, depending on which laryngeal muscles are being contracted and in what order. The magnitude of the voluntary pitch change is substantially larger than in the case of the passively generated rise-fall contour on non-focally accented syllables.

It is a popular belief that accent is carried by *high* pitch. Yet, this is obviously an oversimplification. Phonetically, accent is correlated with a *change* in pitch. In any pitch change there will be a (relatively) low pitch that changes to a higher pitch, or vice versa. Yet it is not invariably the case that the high-pitched portion of the movement marks the accent. In certain accent-lending pitch falls, it is demonstrably the low target of the movement that marks the accent, not the high-pitched starting point. In autosegmental accounts of tonally marked accents this is expressed by an explicit prominence marker that is associated with the low (L) constituent of accent-lending falls, e.g. HL* (where * indicates pitch prominence). It would, of course, make little sense for the H-part to carry the element of prominence, since it would not be tonally distinct from its preceding context.

Note, incidentally, that the centre frequency of the second formant, the resonance frequency that closely corresponds with the degree of backness of vowels (i.e. that reflects the length of the front or oral cavity in the vocal tract), is correlated with F0. This explains why it is possible to perceive intonation in whispered speech (Mayer-Eppler, 1957; Miller, 1961). The mechanism underlying this F0/F2 coupling is ill understood. For one thing we know that raising the tongue pulls up the larynx, which in turn causes stretching and tensing of the vocal cords, i.e. higher pitch (so-called vowel-intrinsic pitch, see above). Let us assume, then, that the larynx is raised when the speaker produces a high pitch. If the musculature of the larynx normally pulls the tongue body down, the tongue will be allowed to be raised and pushed forward more easily when the speaker adjusts the position of the larynx so as to generate a high pitched tone. The assumption being made here is that the tongue body - larynx coupling works symmetrically: when the tongue is raised, the larynx is pulled up (causing higher pitch), and when the larynx is pulled up so as to facilitate the generation of high pitch, the tongue body is given more leeway, which is manifested predominantly in allowing the tongue body to be shifted forward. This account is speculative, and needs to be backed up by further physiological evidence.

7.3.2.2 Temporal structure

Temporal structure of speech is commonly measured in absolute time, say milliseconds. The average duration of a syllable, across a wide range of languages, is in the order of 200 ms, and the average duration of a segment will not be in excess of 100 ms. Of course, syllable and segment durations will vary greatly within and between languages and speakers. For the purposes of accent marking it is often more revealing to abstract from absolute duration, and concentrate on relative durations, expressed for instance as the percentage of duration of a segment relative to the duration of a syllable, or syllable duration relative to word duration, etc. Changes in speaking rate, i.e. local accelerations and decelerations, may be represented as deviations from mean or normalised speaking rate; the latter presupposes a theory of nominal segment durations for a given language. When such a theory is not available, an experimental approximation of nominal segment durations can be obtained by measuring the duration of each segment in its original syllable as a focally accented monosyllabic nonsense word in a fixed carrier sentence (cf. Sluijter & van Heuven, 1995a). The result would be maximally elaborated segments in connected speech (this is in fact the type of speech that diphone building blocks are often excerpted from, for the purpose of speech synthesis systems, cf. Drullman & Collier, 1993).

The general claim is that an accented syllable is pronounced more elaborately, therefore more slowly, than an unaccented syllable. It follows from what was said under physiological correlates, that the accented syllable is not stretched linearly; rather, the middle portion of the vocalic nucleus is stretched more, with the effect tapering off towards the consonantal edges. However, the effect is somewhat asymmetrical in that the deceleration extends further into the postvocalic coda consonant(s) than into the prevocalic onset consonant(s). It has also been suggested that the difference in temporal structure between accented and unaccented syllables within a word increases as overall tempo is higher. This means that accented syllables retain more of their nominal duration than unaccented syllables, a characteristic that would be typical of stress-timed languages (Dauer, 1983). The experimental evidence for these claims is rather thin. Just to mention one example, Sluijter & van Heuven (1995a) presented evidence that the temporal contrast between accented and unaccented syllables in Dutch disyllabic words was statistically more sharply delineated when the onset consonants were excluded from the duration measurements of the syllables.

As far as we have been able to ascertain, relative duration of the syllables (rhymes) is the single most reliable correlate of accent. Whether a word is in focus or not, there will always be a duration difference in favour of the accented syllable – leaving aside for the moment the problem of normalising

out the obscuring effects of intrinsic and co-intrinsic duration of segments.

When a Dutch word is in focus on the sentence level (marked by focal accent) the entire word is stretched by some 10 to 15%. Unlike the non-linear stretching and shrinking of stressed and unstressed segments within a word as a result of overall changes in speaking rate, the time compression/expansion was found to be linear when the speaking rate was changed locally as a result of focal accent marking (Nooteboom, 1972; Eefting & Nooteboom, 1993; Sluijter & van Heuven, 1995a). As a result the *relative* durations of the stressed and unstressed syllables within the word remain unaltered, whether the word receives focal accent or not. It is not clear at this time what the lengthening domain of focal accent is. Is it the lexeme (as implied by the results of Eefting and Sluijter & van Heuven) or does some of the lengthening spill over from the prosodic head to the dependent word in compounds, as was suggested in van Heuven (1993)? At least one thing is clear, the lengthening domain of accent does not extend across (compound) word boundaries within phrases: when an entire word group containing two accentable words (e.g. *do you mean the fortress on the mountain (or the river?)*) was put in focus with just a single (integrative) sentence accent on its prosodic head (*mountain*), only the prosodic head was stretched but none of the other words within the focus domain (underlined, cf. Eefting & Nooteboom, 1993; van Heuven, 1995). Recently, it has been argued that the lengthening domain of focal accent is different in English than in Dutch. Turk & Sawush (1995) showed that the lengthening domain was bound to the within-word foot in English. It is unclear at this time whether Dutch and English are parametrically different in this respect or whether the Dutch experiments did not use the right kind of word materials to really distinguish between the entire word versus the within-word foot as competing candidates for the lengthening domain. The domain of focal lengthening has not been researched in any other languages than the two Western Germanic languages mentioned here. The field is obviously in need of further data here.

7.3.2.3 Intensity

The intensity of the sound pressure wave has long been considered as an acoustical correlate of accent. Intensity (or sound pressure) is proportional to the square of the amplitude of the speech waveform averaged over a moving time-window that is long enough to include two glottal pulses (typically with an integration time of 20 ms for the male voice range and 10 ms for a female voice). Absolute intensity is expressed in Watts per square centimeter. However, since in speech we are not so much interested in absolute sound pressures as in relative differences between sound pressures, intensities are usually expressed in decibels (dB). When two intensities differ in terms of

Watts by a 1:10 ratio, the stronger of the two has a 20 dB greater relative intensity; when the power ratio is 1:100 the relative intensity difference is 40 dB, when the ratio is 1:1000 the difference is 60 dB. So each time the absolute intensity difference is 10 times larger there is a 20 dB increase in relative intensity. The perceptual span between the weakest sound pressure that can be detected in silence (the threshold of hearing, axiomatically set at 0 dB) and the strongest sound pressure that can be tolerated without crossing the pain threshold is 120 dB.

Generally, the dynamic range of a spoken utterance is rather restricted, somewhere in between 55 and 75 dB above the threshold of hearing. When screaming, intensity levels can be increased to some 85 dB, and by whispering low intensities in the 40 to 55 dB range are afforded.

Intensities of speech sounds are unstable as they vary considerably (intensity drops in the order of 5 dB) when the speaker inadvertently turns his head or when some object momentarily intervenes between the speaker's mouth and the listener's ears. Intensity differences of similar magnitude have commonly been reported as correlates of accent. These differences are small but prove fairly reliable correlates (i.e. with little variability) for pitch accents (marking focus on the sentence level), but are even smaller and unreliable when lower degrees of accent are being signaled (cf. Lea, 1977; Beckman, 1986 for English; van Katwijk, 1974; Rietveld, 1984; Sluijter & van Heuven, 1995b; Sluijter, 1995 for Dutch).

In all these (and other) studies intensity was measured at the vocally most intense point during the syllable, the peak intensity, which is usually reached shortly after the vowel onset. Lea (1977) and Beckman (1986) suggested alternative correlates of accent, viz. the *intensity integral* (the summation of intensities throughout the accented vowel) or *average intensity* (as the preceding but normalised for vowel duration). The intensity integral proved a very stable correlate of accent, but it should be pointed out that the intensity and duration correlates are conflated here into one complex cue. Obviously, the combined correlate will be more successful than either of its components. As a general rule we advocate the use of multiple simplex correlates rather than singular complex indexes as the latter obscure whatever systematic interactions exist among the component correlates.

7.3.2.4 Spectral balance

Accent in Western Germanic languages has often been equated with the expenditure of vocal effort, which is correlated with perceived loudness. The most obvious acoustic correlate of physiological effort and perceived loudness, it was held, is vocal intensity. As was explained in § 7.3.1, increased pulmonary effort causes a larger volume-velocity of airflow

through the glottis. The result is not just the generation of larger glottal pulses but also, and more importantly, of a more strongly asymmetrical glottal pulse. Typically, the closing phase of the glottal period is shortened, yielding a smaller opening coefficient (the duty cycle of the glottal pulse, i.e. the proportion of the time the glottis is open relative to the period duration), and the trailing edge of the glottal period is steeper. The greater steepness of the glottal closure as well as its more abrupt ending, cause the generation of relatively strong higher harmonics in the glottal pulse. As a result the spectral tilt of vocalic sounds produced with greater vocal effort emphasizes the higher frequencies. The spectral tilt of the glottal period produced with average effort has a 12 dB/octave roll-off[7]. When speakers (or rather singers) were asked to produce sustained vowel sounds with great vocal effort, the spectral tilt proved less steep, due to the fact that there was a relative boost of frequencies between 500 and 2000 Hz (Gauffin & Sundberg, 1989). It has been shown recently that a similar phenomenon can be observed during the production of local vocal effort, i.e. during the production of an accented syllable (Sluijter & van Heuven, 1993, 1995b for Dutch; Sluijter, Shattuck-Hufnagel, Stevens & van Heuven, 1995 for American English; Fant & Kruckenberg, 1995 for Swedish; Campbell, 1995 for Japanese; see also Sluijter, 1995). Systematic comparison of the various acoustical correlates of accent in Dutch and English shows that the change in spectral balance (or tilt) is a much more reliable correlate of accent (not only of focal accent but also of lesser degrees of accent) than just overall intensity (Sluijter, 1995).

Measuring the spectral balance (or "tilt") is not without problems. Ideally, one needs to strip away the influence of resonances brought about by cavities in the supraglottal tract from the vocal output radiated from the mouth, so that the spectrum of the unfiltered glottal waveform is recovered. Once a clean glottal spectrum is available, the spectral tilt is a matter of fitting a simple linear regression function through the harmonics (plotted along a logarithmic frequency axis), and measuring its slope coefficient in dB/octave.

The process that has been developed to undo the resonance effects of the vocal tract, is called inverse filtering. Inverse filtering packages are in use at some of the more sophisticated phonetic laboratories, but they are not part of the standard techniques. In lieu of full-fledged inverse filtering, some fast-and-dirty approximations have been suggested by Stevens (1995) and applied

[7] When vowel sounds are radiated from the mouth some +6dB/octave is added to the spectral slope, so that the spectral tilt of an average vowel equals −12 + 6 = −6 dB.

to the description of accent levels by Sluijter et al. (1995), Sluijter & van Heuven (1995c) and Sluijter (1995).

When it is not necessary to know the absolute values of spectral tilt (e.g. when no comparison across different vowels is being made) a simpler approximation of spectral tilt is afforded by measuring intensity in four contiguous filter bands (one base filter from 0-0.5 KHz, and three contiguous octave filters: 0.5-1 KHz, 1-2 KHz, 2-4 KHz, cf. Gauffin & Sundberg, 1989; Sluijter, 1995). A linear regression line fitted through the four intensity levels at the filter bands' centre frequencies (plotted along a log frequency axis) yields the spectral tilt measure. In fact, we found that the intensity levels in the base and highest octave filter did not vary much as a function of accent level, so that a good substitute of spectral balance was obtained by just measuring mean vowel intensity (at the overall intensity peak) in the 0.5-2 KHz band (Sluijter & van Heuven, 1995b; Sluijter, 1995).

7.3.2.5 Vowel quality

Accented vowels have traditionally been equated with "clear" or spectrally expanded vowels featuring greater articulatory effort and precision, that is vowels lacking the spectral reduction that is characteristic of unaccented vowels. In order to discuss the effects of accent in vowel quality, we will first have to explain how vowel quality can be determined acoustically. It has been common practice in phonetics to express vowel quality with reference to a vowel space, which is most easily viewed as an articulatory space defined by:

- the position of the tongue in the vocal tract (the point where the tongue approaches the palate or pharynx most closely, dividing the vocal tract into a large back cavity (throat) and a smaller front cavity (mouth), and
- the width (area) of this narrowest passage between throat and mouth cavity;
- lip protrusion is the third articulatory parameter defining the vowel space;
- the fourth parameter, i.e. the width (area) of the nasal port, does not appear to play a role in vowel reduction/expansion under the influence of accent.

Simplifying somewhat, a perceptually adequate representation of the acoustic correlates of this articulatory vowel space can be obtained by a two-dimensional graph plotting the centre frequency of the lowest resonance, called first vowel formant or F1 against F2´ (F2-prime), which is some weighted average of the second, third and fourth lowest resonances: F2, F3, F4 (see Fant, 1973). F1 reflects the degree of mouth opening and F2´ reflects the degree of backness as well as lip protrusion (i.e. the effective length of the oral cavity). Distances in this two-dimensional space assume greater perceptual authenticity if the F1 and F2´ parameters are expressed in terms of Barks (number of critical bands[8]) rather than in Hertz. Using plots of this sort, accented vowels assume more peripheral positions, whereas unaccented vowels gravitate towards the centre of the plot, i.e. towards the position of the neutral vowel schwa. Degree of spectral expansion can then be expressed in terms of distance (in Barks) from the centre of gravity of the F1-by-F2´-plot[9].

Figure 2 illustrates the effects of word stress and sentence accent on the expansion/reduction of the long (tense) Dutch vowels /e:, o:, a:/ read by 15 male speakers. The position of the schwa (averaged over 300 tokens across consonant environments and speakers) may serve as a centre of gravity in the vowel space.

Spectral expansion is largest for vowels pronounced in isolation ('isol'). Some reduction is already visible when these vowels occur in the stressed syllable of focally accented words ('+S+A'). Considerable reduction is observed for stressed vowels in unaccented words ('+S−A') or for unstressed vowels in accented words ('S+A'). Severe spectral reduction is applied to the unstressed vowels of unaccented words ('S−A'): here the spectral distance to the centre of gravity /ə/ is minimal.

This account of *spectral* expansion/reduction appears to be an exact parallel of the account given above of the effects of stress and focal accent on *temporal* expansion/reduction.

---

[8] A critical band is a band of frequencies within which simultaneous tones do not interfere with one another; the size of the critical band at a particular frequency is a measure of the frequency selectivity (auditory acuity) of the human hearing mechanism in complex (e.g. speech-like) sounds (cf. Bladon & Lindblom, 1981).

[9] This spectral distance measure underestimates the degree of expansion of full rounded central vowels; more sophisticated measures in which the higher formants are kept separate, are called for in this case. In the practice of experiments, full rounded central vowels had better be avoided.

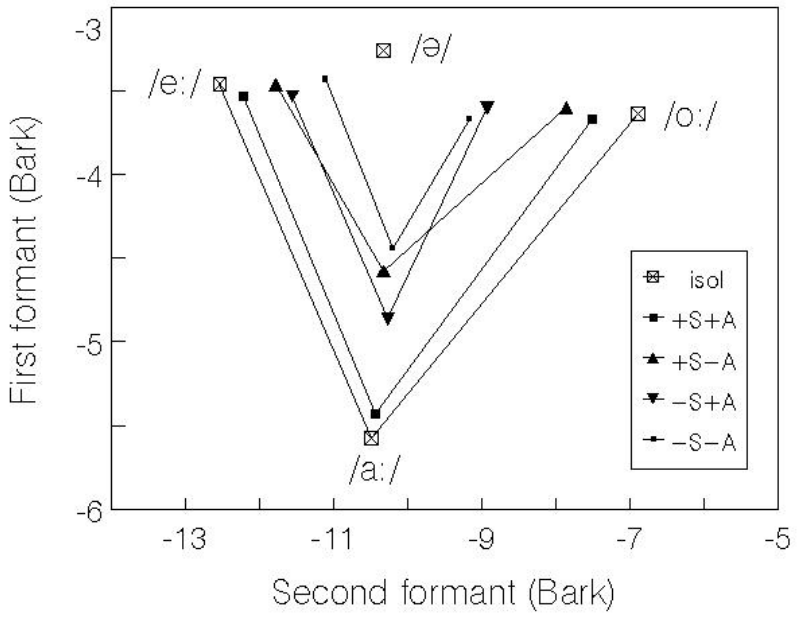


Figure 2. F1 (Barks) and F2 (Barks) of three Dutch peripheral vowels produced by 15 male speakers in five stress/accent conditions (see text, after van Bergem, 1993).

7.3.2.6 Resistance to coarticulation effects

One aspect of a spectrally expanded accented syllable is that it shows minimal influence of coarticulation with abutting syllables, which in turn are strongly influenced by the adjacent accented syllable. So characteristics of the accented syllable are anticipated in the preceding syllable, and perservere into the following syllable, but the accented syllable itself contains only minimal characteristics of the abutting unstressed syllables. Resistance to coarticulation was claimed to be the most important correlate of accent in Lithuanian by Dogil (to appear); see also Pakerys (1982, 1987).

It is not quite clear how resistance to coarticulation can be measured and quantified. One way in which this could be done is to locate the beginning and end of vowel-onto-vowel formant transitions (if the formants do not move in synchrony, study the behaviour of F2 only) from the preceding syllable into the accented syllable, and from the accented into the following syllable (cf. Öhman, 1966). Then determine the point along the time axis where half of the formant trajectory (i.e. half of the formant frequency difference between the consecutive vowels) from the unaccented to the accented vowel (and vice versa) has been covered. The coarticulatory

window of the stressed syllable is then expressed as the time interval between the preceding and following 50% points divided by the duration of the accented syllable. The larger the relative window size, the more resistant the syllable is to coarticulation.

7.3.2.7 Reiterant speech

As was discussed in the introduction to this chapter (§ 7.1.1), a proper perspective on prosodic variation can only be obtained if the non-prosodic effects of segmental structure have been normalised out of the speech materials. There is no adequate theory, however, for any language at all, that tells us how to factor out (co-)intrinsic pitch, intensity and duration from arbitrary segment sequences making up a stretch of connected speech. For this reason experimental phoneticians have devised a trick to circumvent the need for normalisation; this trick is called reiterant speech (Liberman & Streeter, 1978; Nakatani & Shaffer, 1978). The speaker is asked to pronounce a speech utterance while replacing every syllable (or only the syllables corresponding to some target word) by a single uniform syllable, e.g. /ma/ or /ba/. Since in this type of so-called reiterant speech, all the syllables have the same segmental make-up, there is no need for non-prosodic normalisation. The claim, which is insufficiently substantiated in the literature, is that the speaker dubs all (and only) the prosodically relevant variations onto the reiterant version of the original utterance. Although a good methodological study of the reiterant speech methodology is still lacking, researchers of prosody are well-advised to make recordings of relevant speech materials in two versions: normal and reiterant speech. We have invariably found cleaner results with reiterant speech than with the normal speech originals (cf. Sluijter, 1995).

7.3.3 Perceptual cues of accent

7.3.3.1 Methodological introduction

7.3.1.1.1 The need for speech (re-)synthesis

There is a respectable body of literature on the perceptual importance of the various acoustic correlates of accent dealt with in the previous section. In order to determine the perceptual importance of an acoustic difference between to speech utterances, say the difference between an accented and a non-accented version of a syllable, there should be no other differences between the two utterances than just the acoustic property that is being

studied. This condition cannot be met by using ordinary human speech, since it is not possible for a human speaker to produce two versions of the same word or syllable with just a single difference in F0 or in duration, while keeping all other acoustical properties of the two utterances identical. There is only one way out of this dilemma, which is the use of synthesized (or at least resynthesized human[10]) speech. The most acceptable sounding stimuli are obtained by resynthesized speech. For this purpose a human utterance is recorded and analysed into perceptually relevant parameters, e.g. in terms of five formant centre frequencies and their associated bandwidths (F1..F5; B1..B5), a determination of voicing (V/UV), F0 (for the voiced portions of the speech wave), and overall intensity. The parameter extraction is done either at fixed time intervals (say, every 10 ms) or for each single pitch period (pitch synchronous analysis). There are algorithms that perform this parameter extraction quickly and accurately[11]; the most widely used family of algorithms for the extraction of spectral parameters is called Linear Predictive Coding, or LPC (Atal & Hanauer, 1971); also there are over a hundred different algorithms for the extraction of F0, cf. Hess, 1983). The parametrised speech utterance contains only a fraction of the information that was contained in the original speech wave (and is therefore often used as an efficient coding scheme for information reduction for speech storage and telecommunication purposes, cf. Klein & Paliwal, 1995). The parametric representation can be used to regenerate an approximation of the original utterance. The result of the resynthesis is perceptually almost identical to the original (though some loss of quality is invariably incurred as a result of the data compression), but - crucially - all the linguistically relevant properties of the original utterance will be retained. The important thing is that the researcher now has the opportunity to change the measured parameter values of each and every parameter to his own liking. In an experiment on the importance of, say, the degree of spectral reduction for accent perception, only the centre frequencies of one or two selected formants can be changed,

[10] In synthesized speech, utterances are generated (by a computer program or by dedicated hardware) from scratch by recombining and concatenating pre-stored minimal units (segments, diphones, demi-syllables or other); all coarticulatry and prosodic adaptations have to be introduced by explicit rules in the synthesis system. In resynthesis the utterance is a integral copy of a human original, in which selected parameters may be edited before converting the parametric representation back to an audible waveform.

[11] Before the advent of computers, analysis and resynthesis were possible using analog apparatus such as the sound spectrograph and the Pattern Playback machine (or similar analog speech synthesizers). The procedures were cumbersome and the results less reproducible and of poorer quality (cf. Cooper, Liberman & Borst, 1951; Flanagan, 1972).

while keeping all other measured parameters at their original human values[12].

7.3.3.1.2 Finding the relative importance of cues

A lot of scientific effort has been put into the determination of the relative importance of the various acoustic correlates outlined in § 7.3.2 in the perception of accents of various degrees of strength. The approach taken to this problem is straightforward. A stimulus continuum is generated by taking a word that can be stressed alternately on one syllable or an other, a so-called minimal stress pair such as English *import*, and creating multiple versions of this word by systematically changing the values of two competing parameters. For instance, if one is interested in finding the relative importance of vowel duration (one aspect of temporal structure) versus vowel peak intensity, as was the object of the classical study by Fry (1955), the parameter values are first determined in prototypical human exemplars of /'daɪdʒest/ (noun) and its stress counterpart /daɪ'dʒest/ (verb). In the study mentioned the typical vowel duration durations of V1 (/aɪ/) and V2 (/e/) were 191 versus 124 ms for the noun (initial stress, 12 tokens) and 143 versus 183 ms for the verb (final stress, 12 tokens). Similarly the peak intensities were 15 versus 16 dB for V1 and V2 of the noun and 17 versus 10 dB for the verb. Each of the two parameter ranges is then sampled in an equal number of steps, where each step size should at least be large enough to be heard[13]. In order to keep the experiment within manageable proportions the number of steps in this type of two-parameter study generally does not exceed 10 (10 × 10 steps = 100 systematically different tokens). The following figure exemplifies the stimulus range generated in Fry (1955). Note that each stimulus parameter is sampled in five steps with equal increments of 5 dB along the intensity parameter and with unequal (and rather arbitrary) steps for the vowel duration parameter. Next, the 5 × 5 = 25 systematically different tokens are embedded in a fixed carrier phrase and presented in random order for accent determination to a group of native listeners, who have to decide

[12] When only F0, duration or overall intensity needs to be manipulated in the resynthesis (rather than formants frequencies and bandwidths) better sound quality can be obtained by PSOLA (Pitch Synchronous OverLap and Add; Charpentier & Moulines, 1987; Verhelst & Moulines, 1995).

[13] A point that has been given little consideration in the literature is the nature of the sampling of the range: should the steps be of uniform linear magnitude or should the parameter ranges be scaled first so as to reflect the sensitivity of the human hearing mechanism. In other words, should the ranges be sampled in numerically equal steps or in perceptually equal steps? The practice followed in almost all the experiments in the literature is to opt for numerically equal steps. There is little justification for this easy-way-out solution.

for each token, with forced choice, whether they perceive a noun (initial stress) or a verb (final stress). The results are expressed in percentages as in figure 3).

It can be seen quite clearly that the effect of changing the vowel duration ratio on stress perception is stronger than that of manipulating the intensities of the vowels. When V1 is long (and V2 short) some 90 percent of the responses indicate stress on the first syllable; when V1 is short (and V2 long) only 10 percent stress judgments are obtained for the first syllable. Moreover, the complete cross-over is located between step 2 (V1 > V2) and 3 (V1 < V2). The effectivity of the duration cue is in sharp contrast with the intensity cue, which causes, at best, a partial and shallow cross-over from 46 to 61 percent stress perceived on the first syllable.

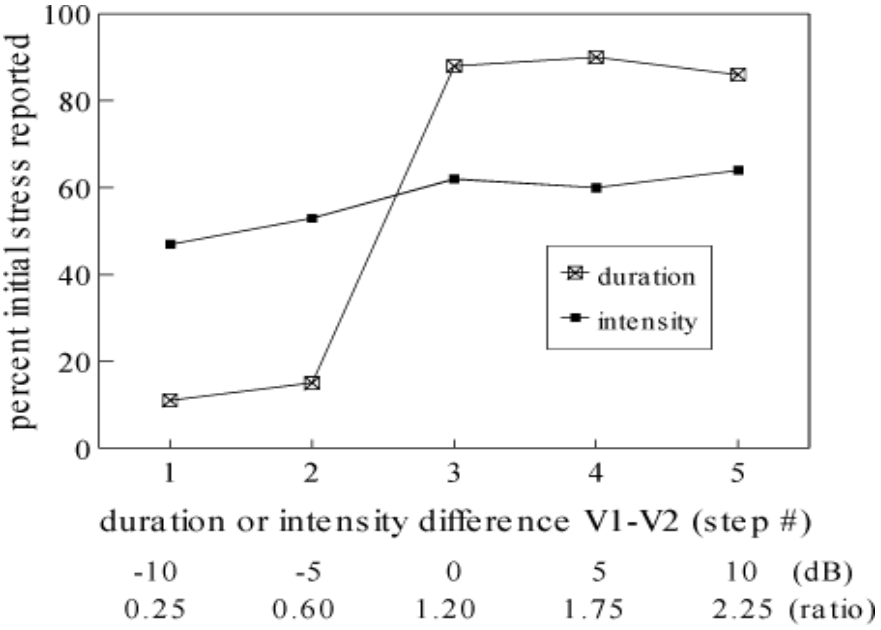


Figure 3: Percent stress reported on the first syllable as a function of vowel duration ratio and intensity difference (dB). Each measurement point is based on 500 responses (after Fry, 1955).

A somewhat more sophisticated analysis of the perceptual strength of stress parameters is afforded by the next figure, which is based on a rerun of Fry's experiment for Dutch (Sluijter, 1995), varying vowel durations in a nonsense word /na:na:/ embedded in a fixed carrier sentence. In one experimental condition overall intensities were manipulated in the same way as in Fry's experiment, in a second condition the intensity differences were concentrated

in the upper part of the spectrum only (i.e. above 500 Hz) so that spectral slope was more level (indicative of greater vocal effort) as the intensity increased. In figure 4 percent stress reported on the first syllable is plotted as a joint function of the vowel duration difference (horizontal axis) and of intensity difference (vertical axis), for overall intensity manipulation (left-and panel) and for intensity+spectral tilt (right-hand panel).

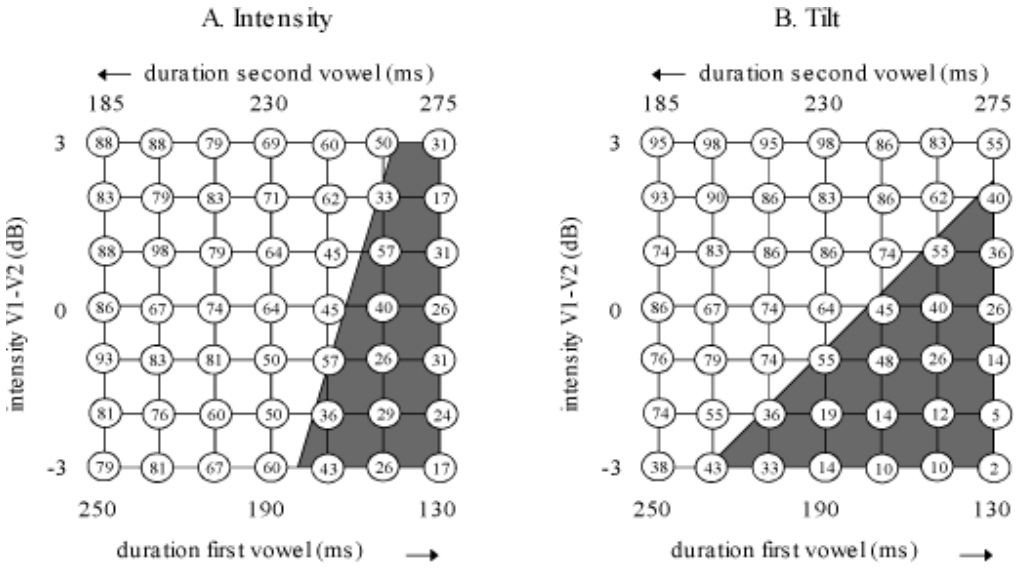


Figure 4. Percent stress reported on the first syllable as a function of vowel duration (ms) and intensity difference (dB). Shaded portions of the figures indicate the part of the stimulus space where final stress was perceived. Panel A: overall intensity manipulation; B: intensity increments above 500 Hz only.

A linear boundary line has been drawn so as to optimally run through the interpolated 50% cross-over points in the stimulus space. The slope of the boundary line in the figures expresses the relative importance of the two competing accent parameters: if the slope coefficient equals 1 (i.e. runs at a 45□ angle) both parameters are equally influential, if the slope is steeper than 1 the parameter plotted along the horizontal axis outweighs the parameter plotted vertically, and vice versa for slope coefficients < 1. The general outcome of this type is that one parameter is more influential than the other, and that there is a trade-off between the parameters. In the left-hand panel, where overall intensity has been manipulated as in Fry's classical study, the duration parameter by far outweighs the peak intensity parameter; one unit (step) along the duration parameter can be traded for (compensated

by) roughly 3 units along the intensity parameter, and even then treading is only possible in a restricted duration range where the duration cue by itself is ambiguous. In the right-hand panel, however, duration and intensity manipulations are about equally influential. Clearly, manipulating intensity in the upper part of the spectrum only (i.e., as is done in human speech) is a better accent cue than manipulating overall intensity.

7.3.3.2 Pitch change

Pitch change, at least in the Indo-European language family, is by far the strongest perceptual cue for the presence of an accent. This comes as no surprise since pitch change is the primary correlate of the highest degree(s) of accent at the sentence level, i.e. of focal accent (see § 7.3.2.1). Therefore a pitch movement cannot be used to mark a lower degree of accent on a word that is out of focus; it would simply raise the accent to a focal accent. Note, incidentally, that a consequence of this view is that any focal accent will have a higher degree of accent than any non-focal accent.

There has been some confusion in the literature on what was called the "all-or-nothing" character of the pitch cue. It was found in early studies (e.g. Fry, 1958, 1973; Morton & Jassem, 1965) that very small pitch changes went unnoticed, but a pitch change in excess of some threshold value would result in a massive effect on perceived accent position, crowding out the effects of any other competing parameter. This was more than likely the result of a ceiling effect, caused by the fact that a stimulus contained only one accentable word (in fact, in the studies mentioned the stimuli were always single-word utterances). When stimuli with multiple accentable words were used, it could easily be shown that the strength of an accent (perceived in terms of emphatic accent) increased with the size of the pitch change associated with it (e.g. Ladd, 1990).

Using stimuli with two pitch movements, each perceived as generating accent on a different word, has opened up the possibility to re-establish the perceptually optimal scaling of frequency intervals in terms of prominence. Until recently, it was widely held that the optimal scaling of pitch movements was in terms of a musical scale, i.e. in terms of parts of an octave, e.g. semitones (see § 7.3.2.1). Closer analysis of the performance of listeners who were instructed to adjust the size of an accent-lending pitch movement in one utterance so as to be perceptually equal to that of a similar movement in an utterance in a different musical register, showed that the optimal perceptual scale for pitch prominence is the ERB-scale (Equivalent Rectangular Bandwidth), which is roughly in between a linear Hertz scale and a logarithmic (i.e. percentual) musical scale (Hermes & van Gestel, 1991).

For a pitch change (or pitch movement) to generate the perception of a (focal) accent, the change in pitch has to be critically timed with respect to the segmental structure of the syllable. It has been found, for instance, that a Dutch accent-lending pitch rise has to start at the onset of the syllable that carries the accent (Caspers & van Heuven, 1993; Caspers, 1994); if the movement is delayed towards the end of the syllable it will not be perceived as an accent but as a boundary marker (‘t Hart, Collier & Cohen, 1990). Conversely, an accent-lending pitch fall in Dutch should start midway through the vocalic nucleus of the syllable that is being accented; if not, the fall is perceived as a (minor) phrase boundary marker.

Note that the more recent, sophisticated studies of the cue value of pitch movements for accent perception presuppose a formal theory of intonation, that specifies exactly what changes (in terms of excursion size, direction and steepness of movement, and segmental synchronisation) in the fundamental frequency are perceived as prominence-lending. When the pitch movements are generated in accordance with such theories (e.g. ‘t Hart et al., 1990 for Dutch) it is obvious that the crucial parameter differentiating initial focal accent from final accent on minimal stress pairs is the timing of the movement. This puts a different perspective on the findings of the older experimental literature. Fry (1958) reported that the high-pitched syllable in his minimal stress pair *ˈsubject* (noun) ~ *subˈject* (verb) is perceived as accented. High pitch on the first syllable followed by lower pitch on the second syllable is, in fact, a crude implementation of an accent-lending pitch fall, whereas low pitch on the first syllable followed by higher pitch on the second is a step-function approximation of an accent-lending pitch rise synchronised with the onset of the second syllable. Such stepwise pitch movements are impossible to produce in human speech.

It is unclear how large a pitch change should be in order to cue the perception of a focal accent. One may assume that the threshold value depends on the individual behaviour of the speaker. Some speakers generally produce relatively large pitch movements (lively voices); other speakers tend to reduce the pitch span of their voice (flat voices). Pitch span may also vary within speakers depending on mood, speaking rate, and communicative situation. Finally, given that some languages observe a wider pitch span (e.g. British English with a 12-semitone span) than others (e.g. Dutch with a 6 semitone span), it may well be the case that the criterion value for the size of an accent-lending pitch change varies as a function of the mean pitch span of the intonation system. Clearly, more research is needed in order to uncover the mechanism regulating the criterion value for pitch movements cuing focal accent.

7.3.3.3 Duration

Relative duration is a very reliable cue to not only focal but also non-focal (lower degrees of) accent. In most experiments that address the cue value of duration, temporal structure was implemented as the difference in duration of the vowels in the stressed versus unstressed syllable in minimal stress pairs (e.g. Fry, 1955). The general result is that manipulating the duration of just the vowels in a word effects a full cross-over from perceived stress in one position to an other. It follows from the previous section, however, that the effect of durational change favoring stress perception on one syllable can always be overruled by a pitch movement associated with an other.

Given that vowels are the nucleus of the syllable, carrying most of the syllable's weight (in terms of both loudness and duration), it makes sense that researchers intuitively limited their manipulations of syllable duration to changing the duration of the vocalic nuclei only; there are virtually no comparative studies on the differential effects of (onset and coda) consonant versus vowel duration manipulation on the perception of stress position. A study on Dutch showed, much to our surprise, that there were no systematic differences in the impact of consonant versus vowel duration changes (van Biezen, 1988). This raises the question how sensitive human listeners really are to subtle differences in the temporal structure of syllables; for instance, we would welcome experimental work on the perceptual importance of the differences that were described above (§ 7.3.1.3) for accent marking and for boundary marking.

7.3.3.4 Intensity and spectral tilt

The older literature persistently claimed that accented syllables in languages such as English and Dutch differ from their unaccented counterparts in terms of loudness. In early perception experiments this claim was tested by manipulating the overall intensity of syllables. The results were quite clear: intensity manipulations were by far the most ineffective cue for accent, and could easily be overruled by differences in duration and/or pitch (Fry, 1955; Mol & Uhlenbeck, 1956; Morton & Jassem, 1965; van Katwijk, 1974).

The problem with overall intensity is, of course, that this is not a realistic operationalisation of loudness. When a human speaker increases his vocal loudness, it is not so much overall intensity that is boosted but intensity in the higher harmonics (yielding a less negative spectral tilt), typically in the frequency range above 500 Hz (see above, § 7.3.2.4). Sluijter, van Heuven & Pacilly (1995), Sluijter & van Heuven (1995b) and Sluijter (1995) showed quite clearly that a shift of 6 dB overall intensity (a realistic copy of human speech) was completely ineffectual in a stress perception experiment;

concentrating the same intensity shift in the 500 - 2000 Hz frequency band (which amounted to an intensity shift of 18 dB within the limited frequency band) was equally effective on stress perception as a realistic change in temporal organisation. In fact, when the temporal organisation was obscured by introducing reverberation in the stimulus (which generally happens when we listen to speech in closed spaces such as rooms), the spectral tilt cue proved even stronger than the duration cue (see § 7.3.3.1.2: figure 4).

It seems to us that spectral tilt and duration can be traded as cues for accent on an equal basis. However, both cues are only important when we are listening to non-focal accents. In a focal accent there will always be a pitch change that overrules whatever effects might have been caused by spectral tilt and/or duration.

7.3.3.5. Spectral reduction

There are very few studies on the importance of spectral reduction/ expansion for accent perception. Fry (1965) did a small-scale study on the relative importance of vowel reduction and duration structure on the perception of stress in English, and found that vowel reduction was less important than duration as a cue to (absence of) accent on a syllable. Rietveld & Koopmans-van Beinum (1987) did a careful study of vowel reduction as an accent cue in Dutch, but since this was a one parameter study, it remains unclear just how important Dutch vowel reduction is relative to other accent cues.

We have no knowledge of any research studying the perceptual importance of reduction of consonants for accent perception.

7.4 Cross-linguistic differences among stress cues

There has been some speculation on the question whether or not any language that uses the linguistic parameter of accent, uses the same correlates of accent, with the same order of relative importance of these acoustic correlates as cues to accent perception. The general feeling is that different correlates (and different perceptual cues) are employed depending on the structure of the language under analysis. We will discuss two sets of differences between languages, and their consequences for accent marking. The first set of differences concerns the type of accent system a language employs, whereas the second source of difference is located in the relative exploitation within a language of accent parameters for other linguistic contrasts.

7.4.1 Correlates of different accentual systems

7.4.1.1 Fixed versus word-bound stress

Referring back to § 7.2.1, it seems reasonable to assume that that languages with fixed stress have a smaller need for strongly marked stress positions than languages in which the position of the stressed syllable varies from word to word (i.e. in word-bound systems). In the latter type of language the position of the stress within the word is a potentially contrastive property, whereas in the former type words are never distinguished from each other by the position of the stress because stress is invariably in the same position for all the words in the language[14]. Note that this claim is limited to the marking of non-focal accents only. We assume that focal accent will always be strongly marked, since it is important for the listener to know which word(s) are to be interpreted as in focus.

We would predict, therefore, that the size of the pitch movements does *not* vary as a function of the type of stress system of the language, but that the difference between stressed and unstressed syllables in non-focused words is less clearly marked along all the non-pitch parameters correlating with accent. Although hardly any research has been done to check these predictions, there is some evidence that the basic prediction is correct. Dogil (to appear) presents a comparative study of Polish (fixed penultimate stress) and German (quantity-sensitive plus lexical stress) accent marking, and concludes that stress position is less clearly marked in Polish. Similar results were found for Indonesian (fixed penultimate stress, or even free stress), showing that Indonesian stress is only weakly marked (Laksman, 1994; Odé, 1994), as had already been claimed almost a century ago (Gerth van Wijk, 1909).

[14] Although the case could be made that stress in languages with fixed stress has a demarcative function, which is not necessarily less important than a contrastive function, we have never come across any claims suggesting that demarcative stress should be the more strongly marked type.

7.4.1.2 Stress-accent versus pitch accent-languages

The following claims can be found in the literature on accent marking. Languages such as English and Dutch pronounce accented syllables with greater vocal effort, so that the accented syllables are not only pronounced with a pitch obtrusion and greater duration, but also by greater intensity and less negative spectral tilt (the cleanest correlate of vocal effort). Such languages, in which stress is marked by greater loudness, have traditionally been called 'dynamic stress languages' or 'stress accent systems' (cf. Beckman, 1986). Pitch-accent languages such as Japanese do not mark the stressed syllable by any other means than just a change in pitch. Beckman's results showed that, indeed, pitch changes were the strongest accent cue in both English and Japanese. However, in English duration and intensity proved secondary cues whereas in Japanese these were without perceptual effect. Spectral tilt was recently reported as a correlate of accent in Japanese (Campbell, 1995), but no perceptual follow-up study has been done yet. If the claim that pitch-accent languages do not use any other accent cues than just pitch is right, manipulating the spectral tilt should not be a (secondary) accent cue in Japanese.

7.4.1.3 Effects of dominant or fixed stress position

There is an old claim (as far as we know first formulated by Jakobson) that speakers of a language with fixed initial stress perceive the rhythm of a (perfectly regular) metronome different than speakers of a language with fixed final stress. The former group would parse the regular sequence of metronome ticks as a succession of trochees (Sw, Sw, ...) whilst the latter group would perceive a sequence of iambs (wS, wS, ...). In other words, listeners would superpose the default stress pattern of their language even on a non-rhythmical sequence of auditory events.

The only cross-linguistic study of rhythmical parsing (along the lines sketched in the previous paragraph) that we know, is Berinstein (1979). Her data show that speakers of Mayan languages with fixed word-final stress have a strong tendency to perceive stress on the last syllable in sequences of four repetitions of an identical 100 ms syllable [bi] synthesized on a 130 Hz monotone. Speakers of English, however, were strongly biased towards reporting the stress on the first syllable[15]. The position bias was also apparent

[15] The results for a group of Spanish listeners were uninterpretable. Berinstein's stimulus tape was also offered to a group of Dutch listeners, whose results were virtually identical to those of the English group.

from the subjects' behaviour to stimuli with one longer syllable duration within the sequence of four.

When the longer syllable coincided with the default syllable position for stress, its effect on stress perception was large; however, when the deviant syllable duration occurred in one of the non-preferred positions, its effect on stress perception was smaller.

In an unpublished study carried out in our laboratory (van den Bent, Buis & van Oudheusden, 1985) we noted similar bias effects within a single language, Dutch, when we systematically changed the duration of the syllables in Dutch polysyllabic words. Dutch listeners (and presumably English listeners as well) have a strong preference towards reporting the accent on the stressed syllable, even if some other syllable within the word is phonetically more strongly marked for accent than the stressed syllable[16].

The bias favouring perception of accent on the initial syllable in English (and Dutch) can be observed in all the studies mentioned earlier in this chapter. Yet it appears that the bias is only partly accounted for by the statistical properties of the stress system of the language (i.e. the vast majority of English and Dutch word tokens have initial stress, cf. Cutler & Carter, 1987 for English; Quené, 1992 for Dutch). The propensity towards hearing accent on the first syllable of a target word in Dutch increases and decreases as the stimulus word is resynthesized on a higher or lower fundamental, and disappears when the stimulus is whispered (resynthesized while replacing the periodic excitation signal by a noise source), or when it is embedded in a semantically neutral preceding context (van Heuven, 1987; van Heuven & Menert, 1996). We suggested that the initial stress bias in Dutch is largely due to a perceived "virtual" pitch rise from some hypothetical baseline (inferred low declination onset?) towards the

[16] In this type of experiment the stimulus is offered as an isolated word. It is an open question what will remain of the stress bias when words are presented with an accent on a non-stressed syllable when the word occurs in a context that calls for a contrastive accent on the non-stressed syllable.

physically higher pitch of the actual stimulus onset of the isolated stimulus word.

7.4.2 Effects of functional load of accent parameters

In § 7.3.3.1.2 we suggested that the unmarked order of importance for accent perception of the various prosodic cues discussed is a simple function of the acoustic range of each parameter and the sensitivity of the human hearing system to differences along each parameter. This does not mean that the predicted order will always be found in all the languages that have accent. Berinstein (1979) developed a hypothesis suggesting that when a prosodic parameter is already exploited to signal other (segmental or prosodic) contrasts than accent, the position of the parameter in the unmarked rank order of accent cues would fall. For instance, if a language has a phonological length contrast between vowels, the vowel duration parameter will be less effective as an accent cue than in a language − ceteris paribus − without the vowel length opposition. Berinstein (1979) showed that K'ekchi speakers were more sensitive to vowel duration changes as an accent cue than speakers of Caqchiquel; both are closely related Mayan languages spoken in Guatemala, but with one important difference: Caqchiquel has a vowel length contrast that is lacking in K'ekchi.

More or less along the same lines Potisuk, Gandour & Harper (1996) showed that pitch change is a relatively unimportant accent cue in Thai (less important than vowel duration), where the pitch parameter is already heavily exploited to signal a five-member lexical tone contrast.

References

Atal, B.S. & S.L. Hanauer (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America* 50: 637-655.

Beckman, M.E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, M.E. & J. Edwards (1990). Lengthenings and shortenings and the nature of prosodic constituency. In: J. Kingston & M.E. Beckman (eds.) *Papers in Laboratory Phonology: Between the grammar and pfysics of speech*, 152-178. Cambridge: Cambridge University Press.

Bent, H. van den, F. Buis, R. van Oudheusden (1985). The relative contribution of lexical knowledge and duration of segments to the perception of stress in single Dutch words, unpublished ms., Dept. Linguistics/Phonetics Laboratory, Leiden University.

Bergem, D. van (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class on the quality of vowels. *Speech Communication* 12: 1-23.

Berg, M.E. van den (1986). *Modern Standaard Chinees: Toon, accent en intonatie. Een handleiding voor het verstaan en spreken* [Modern Standard Chinese: Tone, accent and intonation, a manual]. Muiderberg: Coutinho.

Berinstein, A.E. (1979). A cross-linguistic study on the perception and production of stress,

UCLA Working Papers in Phonetics 47, 1-59.

Biezen, M. van (1988). Heeft medeklinkerduur invloed op klemtoonwaarneming? Onderzoek naar de effecten van klinker- en medeklinkerduur op klemtoonwaarneming [Does consonant duration influence stress perception? A search for the effects of vowel and consonant duration on stress perception]. Master's thesis, Dept. Linguistics/Phonetics Laboratory, Leiden University.

Bladon, R.A.W. & B.E.F. Lindblom (1981). Modeling the judgement of vowel quality differences, Journal of the Acoustical Society of America, 69, 1414-1422.

Bolinger, D.L. (1958). A theory of pitch accent in English. *Word* 14: 109-149.

Campbell, W.N. (1995). Loudness, spectral tilt and perceived prominence in dialogues. *Proceedings of the Thirteenth Congress of Phonetic Sciences* 3. Stockholm, 676-679.

Campbell, W.N. & M.E. Beckman (1995). Stress, loudness, and spectral tilt, *Proceedings of the Acoustical Society of Japan*, Spring Meeting, 3-4-3.

Caspers, J. (1994). *Pitch movements under time pressure. Effects of speech rate on the melodic marking of accents and boundaries in Dutch*. Holland Institute of Generative Linguistics Dissertations 10. The Hague: Holland Academic Graphics.

Caspers, J. & V.J. van Heuven (1993). Effects of time pressure on the phonetic realisation of the Dutch accent lending pitch rise and fall. *Phonetica* 50: 161-171.

Cooper, F.S., A.M. Liberman & J.M. Borst (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences* 37, 318-325; reprinted in J.L. Flanagan & L.R. Rabiner (eds.): *Speech synthesis*, 59-66. Stroudsburg PA: Dowden, Hutchinson & Ross.

Cutler, A., Carter, D.M. (1987). The predominance of strong initial syllables in English vocabulary, *Computer Speech and Language*, 2, 133-142.

Dauer, R.M. (1983). Stress-timing and syllable-timing realanyzed, *Journal of Phonetics*, 11, 51-62.

Dogil, G. (1995). The phonetic manifestation of word stress. To appear in: H. van der Hulst (ed.) *Word prosodic systems in the languages of Europe*. Berlin: Mouton de Gruyter.

Drullman, R. & R.C. Collier (1993). Speech synthesis with accented and unaccented diphones. In: V.J. van Heuven & L.C.W. Pols (eds.) *Analysis and synthesis of speech, strategic research towards high-quality text-to-speech generation*, 147-156. Berlin: Mouton de Gruyter.

Eefting, W.Z.F. & S.G. Nooteboom (1993). Accentuation, information value and word duration: effects on speech production, naturalness and sentence processing. In: V.J. van Heuven & L.C.W. Pols (eds.) *Analysis and synthesis of speech, strategic research towards high-quality text-to-speech generation*, 225-240. Berlin: Mouton de Gruyter.

Fant, G. (1973). *Speech sounds and features*. Cambridge MA: MIT Press.

Fant, G. & A. Kruckenberg (1995). The voice source in prosody. *Proceedings of the Thirteenth Congress of Phonetic Sciences* 2. Stockholm, 622-625.

Flanagan, J.F. (1972). *Speech analysis, synthesis, and perception*, Springer Verlag, Berlin.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27: 765-768.

Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech* 1: 126-152.

Fry, D.B. (1965). The dependence of stress judgments on vowel formant structure. In: E. Zwirner & W. Bethge (eds.) *Proceedings of the 6th International Congress of Phonetic Sciences*. Karger, Basel, 306-311.

Fry, D.B. (1973). Linguistic theory and experimental research, in W.E. Jones, J. Laver (eds.): *Phonetics in linguistics. A book of readings*, Longman, London, 66-93.

Fuchs, A. 1984. "Deaccenting" and "default accent". In: H. Richter & D. Gibbon (eds.) *Intonation, accent and rhythm*, 134-164. Berlin: Walter de Gruyter.

Gauffin, J. & J. Sundberg (1989). Spectral correlates of glottal voice source waveform

characteristics. *Journal of Speech and Hearing research* 32: 556-565.
Gerth van Wijk, D. (1909). *Spraakleer der Maleisische taal [phonetics of the Malay language]*. Batavia: Kolff.
Hart, J. 't, R. Collier & A. Cohen (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.
Hayes, B.P. (1980). *A Metrical Theory of Stress Rules*. Doctoral dissertation, MIT, Cambridge, Ma. (distributed in 1981 by the Indiana University Linguistics Club, Bloomington, Indiana).
Hess, W. (1983). *Pitch determination of speech signals*. Berlin: Springer.
Hermes, D.J. & J.C. Gestel (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America* 90: 97-102.
Heuven, V.J. van (1987). An unusual effect on the perception of stress. *Proceedings of the 11th International Congress of Phonetic Sciences* 5. Estonian Academy of Sciences, S.S.R., Tallinn, 306-308.
Heuven, V.J. van (1993). On the maximal phonetic scope of accent, In: D. House & P. Touati (eds) *Proceedings of an ESCA Workshop on Prosody*, Working papers 41, Department of Linguistics, Lund University, 132-135.
Heuven, V.J. van (1994a). What is the smallest prosodic domain? In: P. Keating (ed) *Papers in Laboratory Phonology III: phonological structure and phonetic form*, 76-98. London: Cambridge University Press.
Heuven, V.J. van (1994b). Introducing prosodic phonetics. In: C. Odé & V.J. van Heuven (eds) *Phonetic studies of Indonesian prosody*, Semaian 9, 1990, Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, RU Leiden, 1-26.
Heuven, V.J. van & L. Menert (1996). Why stress position bias? *Journal of the Acoustical Society of America* (under revision, 44 pp.)
Hulst, H. van der (to appear). Word accent. In: H. van der Hulst (ed.) *Word prosodic systems in the languages of Europe*. Berlin: Mouton de Gruyter.
Katwijk, A. van (1974). *Accentuation in Dutch, an experimental linguistic study*. Amsterdam/Assen: van Gorcum.
Klein, W.B., Paliwal, K.K. (1995). An introduction to speech coding, in W.B. Klein, K.K. Paliwal (eds.) *Speech coding and synthesis*, Elsevier Science, Amsterdam, 1-47.
Ladd, D.R. (1990). Metrical representation of pitch register, in J. Kingston, M.E. Beckman (eds.) *Papers in Laboratory Phonology: Between the grammar and physics of speech*, Cambridge University Press, Cambridge, 35-57.
Ladefoged, P. (1967). Stress and respiratory activity, in *Three areas of experimental phonetics*, Oxford University Press, London, 1-49.
Laksman, M. (1994). Location of stress in Indonesian words and sentences, in C. Odé, V.J. van Heuven (eds.) *Phonetic studies of Indonesian prosody*, Series Semaian No. 9, Leiden: Vakgroep TCZOAO, 108-139.
Langeweg, S.J. (1988). The stress system of Dutch, doctoral dissertation, Leiden University.
Lea, W.A. (1977). Acoustic correlates of stress and juncture, in L. Hyman (ed.) *Studies in stress and accent*, Southern California Occasional Papers in Linguistics, 4, 83-119.
Lehto, L. (1969). *English stress and its modification by intonation; an analytic and synthetic study of acoustic parameters*, Suomalainen Tiedeakatemia, Helsinki.
Liberman, M.Y., Streeter, L.A. (1978). The use of nonsense-syllable mimicry in the study of prosodic phenomena, Journal of the Acoustical Society of America, 63, 231-233.
B.E.F. Lindblom, B. Lyberg, K. Holmgren (1981). Durational patterns of Swedish phonology. Do they reflect short-term motor memory processes? Indiana University Linguistics Club, Bloomington, IN.
Mayer-Eppler, W. (1957). Realization of prosodic features in whispered speech, *Journal of the Acoustical Society of America*, 29, 104-106.

Miller, J.D. (1961). Word tone recognition in Vietnamese whispered speech, *Word*, 17, 11-15.
Mol, H., Uhlenbeck, E.M. (1956). The linguistic relevance of intensity in stress, *Lingua*, 5, 205-213.
Morton, J., Jassem, W. (1965). Acoustic correlates of stress, *Language and Speech*, 8, 148-158.
Moulines, E., Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones, *Speech Communication*, 9, 453-467.
Moulines, E., Verhelst, E. (1995). Time-domain and frequency-domain techniques for prosodic modification of speech, in W.B. Klein, K.K. Paliwal (eds.) *Speech coding and synthesis*, Elsevier Science, Amsterdam, 519-555.
Nakatani, L.H., Shaffer, J.A. (1978). Hearing "words" without words: posodic cues for word perception, *Journal of the Acoustical Society of America*, 63, 234-245.
Nooteboom, S.G. (1972). Production and perception of vowel duration, a study of durational properties of vowels in Dutch, doctoral dissertation, Utrecht University.
Nooteboom, S.G., Cohen, A. (1984). *Spreken en verstaan, een inleiding tot de experimentele fonetiek [Speaking and understanding, an introduction to experimental phonetics]*, van Gorcum, Assen.
Odé, C. (1994). On the perception of prominence in Indonesian, in C. Odé, V.J. van Heuven (eds.) *Phonetic studies of Indonesian prosody*, Series Semaian No. 9, Leiden: Vakgroep TCZOAO, 27-107.
Öhman, S.E.G. (1967). Coarticulation in VCV utterances: spectrographic measurements, *Journal of the Acoustical Society of America*, 39, 151-168.
Pakerys, A. (1982), *Lietuviu Bendrines kalbos prozodija [The prosody of Standard Lithuanian]*, Mokslas, Vilnius.
Pakerys, A. (1987). Relative importance of acoustic features for perception of Lithuanian stress, *Proceedings of the 11th International Congress of Phonetic Sciences*, Estonian Academy of Sciences, S.S.R., Tallinn, 1, 319-320.
Potisuk, S., Gandour, J., Harper, M.P. (1996). Acoustic correlates of stress in Thai, *Phonetica* (under revision).
Quené, H. (1992). Integration of acoustic-phonetic cues in word segmentation, in M.E. H. Schouten (ed.): *The auditory processing of speech*, Mouton de Gruyter, Berlin, 349-356.
Rietveld, A.C.M. (1984). *Syllaben, klemtoon en de automatische detectie van beklemtoonde lettergrepen in het Nederlands [Syllables, stress and the automatic detection of stressed syllables in Dutch]*, Doctoral dissertation, Catholic University of Nijmegen.
Rietveld, A.C.M., Koopmans-van Beinum. F.J. (1987). Vowel reduction and stress, *Speech Communication*, 6, 217-230.
Sluijter, A.C.M. (1995). *Phonetic correlates of stress and accent*, Holland Institute of Generative Linguistics Dissertations, 15, Holland Academic Graphics, The Hague.
Sluijter, A.M.C., V.J. van Heuven (1993). Perceptual cues of linguistic stress: intensity revisited. In: D. House, P. Touatin (eds.): *Proceedings of an ESCA Workshop on Prosody*, Working Papers, Dept. Linguistics and Phonetics, Lund University, 41, 246-249.
Sluijter, A.M.C., Heuven, V.J. van (1995a). Effects of focus distribution, pitch accent and lexical stress on the temporal organisation of syllables in Dutch, *Phonetica*, 52, 71-89.
Sluijter, A.M.C., Heuven, V.J. van (1995b). Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America* (under revision).
Sluijter. A.M.C., Heuven, V.J. van (1995c). Intensity and vocal effort as cues in the perception of linguistic stress, *Proceedings of Eurospeech 1995*, Madrid, 941-944.
Sluijter, A.M.C., Heuven, V.J. van, Pacilly, J.J.A. (1995). Spectral balance as a cue in the perception of linguistic stress, *Journal of the Acoustical Society of America* (under revision).
Sluijter, A.M., Shattuck-Hufnagel, S, Stevens, K.N., Heuven, V.J. van: Supralaryngeal

resonance and glottal pulse shape as correlates of prosodic stress and accent in American English, *Proceedings of the Thirteenth Congress of Phonetic Sciences*, Stockholm, 2, 630-633.

Stevens, K.N. (1994). "Source mechanisms" and "Basic acoustics of vocal-tract resonators", chapters 3 and 4 of an unpublished manuscript.

Trubetskoy, N.S. (1958). *Grundzüge der Phonologie*, VandenHoeck & Ruprecht, Göttingen.

Turk, A.E., Sawush, J. (1995). The domain of the durational effects of accent, MIT Speech Communication Group Working Papers, X (also under revision for *Journal of Phonetics*).

Chapter 7 in Rob Goedemans, Harry van der Hulst & Ellis Visch (eds) Stress Patterns of the World, Part 1: Background. Holland Institute of Generative Linguistics, Publications nr, 2. Holland Academic Graphics, The Hague, 1996: 233-269.