

The Study of Word Stress and Accent

Theories, Methods and Data

Edited by

Rob Goedemans

Leiden University

Jeffrey Heinz

Stony Brook University

Harry van der Hulst

University of Connecticut



CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107164031

DOI: 10.1017/9781316683101

© Cambridge University Press 2019

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2019

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Names: Goedemans, Rob, editor. | Heinz, Jeffrey, 1974– editor. | Hulst, Harry van der, editor.

Title: The study of word stress and accent : theories, methods and data / edited by Rob Goedemans, Jeffrey Heinz, Harry van der Hulst.

Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2018. | Includes bibliographical references.

Identifiers: LCCN 2018011858 | ISBN 9781107164031 (alk. paper)

Subjects: LCSH: Accents and accentuation. | Grammar, Comparative and general – Phonology.

Classification: LCC P231 .S78 2018 | DDC 414/.6–dc23

LC record available at <https://lccn.loc.gov/2018011858>

ISBN 978-1-107-16403-1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contributors

ROB GOEDEMANS, Leiden University

JEFFREY HEINZ, Stony Brook University

LARRY M. HYMAN, University of California, Berkeley

RENÉ KAGER, Utrecht University

BJÖRN KÖHNLEIN, Ohio State University

HARUO KUBOZONO, National Institute for Japanese Languages and Linguistics

NATALIA KUZNETSOVA, Institute for Linguistic Studies, Russian Academy of Sciences

ANYA LUNDEN, College of William and Mary, Williamsburg, VA

VIOLETA MARTÍNEZ-PARICIO, University of Valencia

KEREN RICE, University of Toronto

IGGY ROCA, University of Essex

NICHOLAS ROLLE, University of California, Berkeley

HISAO TOKIZAKI, Sapporo University

VINCENT J. VAN HEUVEN, University of Pannonia, Veszprém

MARC VAN OOSTENDORP, Radboud University, Nijmegen

HARRY VAN DER HULST, University of Connecticut

ALEXANDRE VAXMAN, University of Connecticut

MARINE VUILLERMET, Laboratoire Dynamique Du Langage, CNRS & Université de Lyon

1 Acoustic Correlates and Perceptual Cues of Word and Sentence Stress Towards a Cross-Linguistic Perspective

Vincent J. van Heuven

1 Introduction

1.1 *Stress at the Word Level*

The languages in the world can be divided roughly into two types of word-prosodic systems. One type, probably a minority, has tone.¹ A tone language uses different pitches or melodies to differentiate between words in the lexicon, just as the vowels and the consonants do. The second type, which is the type that we address in the present chapter, has stress. When a language has stress, every word has one syllable which in some sense is more important, or more prominent, than any other syllable in the same word. This is also the crucial difference between tone and stress. In a tone language there is no difference in prominence attached to the syllables that make up the word, whereas stress is a culminative property: only one syllable can be the strongest (the prosodic head) within a constituent – such as a word.²

Which syllable is the prosodic head of the word is often predictable. For languages with fixed stress there is just one single rule that determines the position of the word stress for the entire lexicon. Hungarian words, for instance,

¹ The *World Atlas of Linguistic Structures* (WALS; Comrie et al. 2005) lists 220 tone languages versus 307 no-tone languages (Chapter 13); at the same time it lists 502 stress languages, divided in chapter 14 between 282 with fixed stress (281 in chapter 15) versus 220 with non-fixed stress (219 in chapter 15). Van Zanten and Goedemans (2007: 64) estimate that languages with stress-based word prosody, tone-based systems and languages without word prosody occur in 80, 16 and 4 per cent of the world's languages, respectively.

² It appears that the two types of word prosody are mutually exclusive. A language has either tone or stress but not both. To be true, there are so-called restricted tone languages in which only one syllable (i.e. the stressed syllable) in a word may carry different tones (e.g. Swedish, Norwegian) but languages that freely combine stress and tone are highly exceptional and seem to arise only as the result of accidental contact between a stress language and a tone language such as happened in the case of Samate Ma'ya (Remijsen 2002) and Papiamentu (Remijsen and van Heuven 2005).

always have stress on the initial syllable; in Weri (a Papuan language; Boxwell and Boxwell 1966) the stress is always on the last syllable of the word. Other languages may have more complex rule systems for assigning stress to words. In weight-sensitive languages such as English, German and Dutch, the complexity of (the rhyme portion of) the syllables determines where the stress goes, at least in monomorphemic words. For instance, stress in Dutch simplex words goes to the final syllable if it is superheavy (i.e. contains more than two morae in its rhyme); if not, stress goes to the pre-final syllable if this syllable is at least heavy (has two morae in the rhyme portion). It has been estimated that a relatively small portion of the monomorphemic lexicon is stressed by exception, that is, deviates from the weight-sensitive stress assignment (e.g. 15 per cent exceptions in Dutch; Langeweg 1988). The exceptions would be cases of unpredictable (or 'lexical') stress. In some languages, there are so many exceptions to any regularity one might want to formulate that stress rules do not make sense. Russian and Greek are often cited as examples of such lexical-stress languages.

Linguistically speaking, the inventory of stressed syllables in a language is richer (i.e. with a greater diversity of segmental structures) than that of unstressed syllables (see, for instance, the counts for Swedish (and four other languages) by Carlson et al. 1985 and for Dutch by van Heuven and Hagman 1988). Moreover, stressed syllables typically resist deleting or assimilating segments to neighbouring unstressed syllables, and whereas unstressed syllables tend to assimilate to adjacent stressed syllables, are susceptible to weakening processes and deletions. In this chapter we will not, however, be concerned with the linguistic properties of stressed syllables. The focus of interest will be on the phonetic realization of stress at the word and sentence level.

1.2 *Stress at the Sentence Level*

Prosody is hierarchically structured. Where one syllable is the prosodic head of the word domain, one word will be the prosodic head of the phrase or utterance it occurs in. Typically, when a word receives sentence stress, the marking of this stress will fall on the syllable within the word that carries the word stress. A syllable in a word with sentence stress has all the phonetic markers of a word stress plus some characteristics that mark it as a sentence stress. Which words in an utterance receive sentence stress and which ones do not depends on the syntax-prosody interface of the language. In Romance languages, for instance, the location of the sentence stresses is largely, if not fully, determined by the syntactic structure of the utterance. In Spanish, the sentence stress (indicated by capitals) will invariably be on the nouns in (1) even though the pragmatic contrast (indicated by square brackets) is in the prepositions (Ladd 1996):

- (1) ¿Quiere café [con] leche o café [sin] leche?
 'Require-you coffee [with] MILK or coffee [without] MILK?'

In other languages, such as those in the Germanic family, sentence stresses are assigned by default to specific words on the basis of the syntactic/prosodic structure of the utterance, but the default rules may be overridden by pragmatic considerations that delete or move sentence stresses so as to express the focus structure of the utterance. Typically only the prosodic head of a prosodic constituent that is in focus, that is, contributes new and contextually unpredictable information to the discourse, receives sentence stress, whereas sentence stresses are deleted (or moved away) from words and phrases that are out of focus, that is, contain relatively unimportant and contextually given information. Thus, in (2a) there is a contrast between two phrases: *the girl* and *the old man*. By default, sentence stress in the latter phrase goes to the noun, which is the prosodic head of the NP. In (2b), however, the pragmatic contrast is between the adjectives *young* and *old*. In this situation pragmatic rules delete the default sentence stress from the noun and reassign it to the adjective.

- (2a) Is Lesley [the GIRL] or [the old MAN]?
 (2b) Is Lesley the [YOUNG] man or the [OLD] man?

1.3 Acoustic Correlates and Perceptual Cues

The purpose of the present chapter is to present and discuss the way word and sentence stress are phonetically marked. It has been known since the 1950s that stress (whether at the word or sentence level) is never marked by a single acoustical property (for a survey see Lehiste 1970). To make the stressed syllable stand out from its neighbours, it is produced with greater physiological effort on the part of the speaker than its unstressed counterpart (e.g. Ladefoged 1967). The greater effort will be exerted at any stage in the speech production process, that is, by the subglottal mechanism (more air is pushed out of the lungs), by the glottal (laryngeal) system (contraction of laryngeal muscles, generating a change in pitch) and by the supraglottal organs (e.g. larger and faster displacement of lips, tongue and jaw, yielding more clearly articulated vowels and consonants). The greater effort is seen, first of all, in closer approximation of articulatory target configurations for segments in stressed syllables. More extreme articulatory movements require more time than small displacements of the vocal organs. The result of this is that segments in stressed syllables have longer durations – all else being equal – than unstressed segments.³ Moreover, in terms of the theory of articulatory phonology (e.g. Browman and Goldstein 1992), there is

³ This view on the relationship between expansion of the articulatory space and duration goes back to Lindblom (1963). Given that longer duration is not associated with clear articulation in

relatively little overlap between adjacent segments in a stressed syllable. In contradistinction to this, unstressed segments greatly overlap, which leads to considerable reduction of segmental contrast. This also accounts elegantly for the observation that segments at the edges of stressed syllables tend to maintain their identity (resist coarticulation with an adjacent segment in an unstressed syllable) whilst unstressed segments across the syllable boundary are disproportionately affected by coarticulation (e.g. Dogil and Williams 1999).⁴

Effort expended at the laryngeal level of speech production takes the form of contracting selected muscles that influence the speed with which the vocal folds vibrate during phonation. The result may be a rapid increase (through activation of cricothyroid and vocalis muscles) or decrease (through activation of the sternohyoid muscle) of the repetition rate of the glottal cycle, causing, respectively, a rise and fall of vocal pitch. A secondary effect of laryngeal effort may be a tightening of the vocal folds (*musculi vocales*), which will then snap together more forcefully than when in a less tightened state. Finally, increased effort at the subglottal level will push more air per unit of time through the glottis, causing, first of all, an increase in intensity of the sound produced by the glottal siren. Secondly, the greater volume-velocity of the airstream through the glottis boosts the Bernoulli suction effect. The increased suction and the tightening of the vocalis muscles conspire to shorten the closing phase of the glottal cycle, which causes the spectrum to become flatter (boosting the intensity of higher harmonics, thereby generating a louder sound – I will come back to this later).

In this chapter we will not deal any further with the physiological basis of stress (but see Erickson and Kawahara 2016 for a well-documented survey of current issues). We will concentrate on the acoustic consequences of increased versus decreased effort (as foreshadowed in the above) and ask (i) what acoustic correlates can be found for the difference between a stressed syllable and its unstressed counterpart, and (ii) what the relative importance is of each acoustic correlate in the marking of stress. At the same time we will consider the question of what acoustic properties are used by human listeners and to what extent these are used to decide whether or not a syllable is stressed. We will make a strict terminological distinction here between acoustic correlates of stress (which can be used, for instance, to identify a stressed syllable by some computer algorithm) and the perceptual cues used by the human listener. We will see that some acoustic correlates, notably the (peak) intensity of

phrase-final syllables (domain-final lengthening), it seems reasonable to assume that clarity of articulation is the primary goal which is subserved by lengthening in stressed syllables.

⁴ The coarticulation window of a stressed vowel may extend to an unstressed neutral vowel schwa in the preceding syllable, across an intervening consonant, and yield both acoustic and perceptual effects (e.g. Van Heuven and Dupuis 1991).

a syllable, allow excellent separation of stressed from unstressed tokens but are hardly used by the human listener.

There is no need, a priori, for the three subsystems of speech production to expend extra effort on the production of a stressed syllable in equal proportion. We may speculate, in fact, that languages differ in the way they exploit effort in each subsystem. For instance, Germanic languages seem to exploit the gradation of supralaryngeal effort more than Romance languages do. More generally, we will ask to what extent the acoustic correlates and perceptual cues of stress have the same ranking order across languages or are differently ordered from one language to the next. If the latter should be the case, then we may ask the supplementary question if the order of importance of correlates and cues can be predicted from the phonological structure of the language at issue.

2 Acoustic Correlates

2.1 *Some Methodological Considerations*

When trying to find acoustic correlates of stress, it is generally not a good idea to just compare acoustic properties of successive syllables in a word. If the segmental make-up of the syllables is different, the correlates of stress are obscured by the intrinsic and co-intrinsic properties of the segments. For instance, open vowels have inherently greater intensity (Lehiste and Peterson 1959) and longer duration than close vowels (Peterson and Lehiste 1960), so that an unstressed open vowel may, in fact, seem more stressed than a closed stressed vowel, as may happen in the English noun *impact*. Several tricks have been suggested to eliminate, or correct for, such inherent segmental properties. One is to run some extrinsic normalization procedure by which the intensity or duration of a segment is expressed in standard deviations away from the mean value of that segment (i.e. z-normalization) as produced by the individual speaker in a larger corpus of materials (e.g. Potisuk, Gandour and Harper 1996).⁵ Another way out would be to use so-called reiterant speech (Larkey 1982, Liberman and Streeter 1978, Nakatani and Shaffer 1978). In this speech mode the speaker replaces the syllables in a target word by repetitions of the same segmental structure; for example, repetitions of /ma/ or /lɪs/. For instance, the target utterance *please say import again* would be produced as *please say*

⁵ Extrinsic normalization expresses the location of an object relative to all other objects in a dataset. Z-normalization is a typical example of extrinsic normalization. This is in contrast to intrinsic normalization, which does not compare values across tokens in the dataset but computes relationships (ratios, differences) between variables obtained for a single token. The V_1/V_2 ratio computed by Fry (1955) would be an example of intrinsic normalization. For a discussion of normalization procedures see Nearey (1978).

mama again, or *please say lislis again*. The claim is that the speaker dubs all (and only) the prosodically relevant variations onto the reiterant version of the original utterance so that no normalization for intrinsic segmental differences is needed. A potential problem with these techniques is that stressed and unstressed syllables are compared syntagmatically, that is, in different linear positions in a larger structure, such as an initial stressed and a final unstressed syllable – so that, strictly speaking, the researcher does not know whether he measures correlates of stress or of sequential position. The safest precaution, therefore, would be to compare stressed and unstressed versions of the same syllables in a paradigmatic way; for instance, by comparing the stressed and unstressed realizations of the first and second syllables in a minimal stress pair such as *the import* versus *to import*. This solution, of course, can only be used if the language has at least one minimal stress pair – which means that it cannot be used in languages with fixed stress.⁶

It has also been found expedient to measure the correlates of stress separately for stress at the word level and at the sentence level. This is generally achieved by (paradigmatically) comparing tokens of stressed and unstressed syllables in a minimal stress pair which was produced in the same position in a surface-syntactically identical sentence with and without focus on the target. Focus on the target word, indicated in (3a–c) in square brackets, is often manipulated by having the speaker answer different questions that highlight one constituent or the other as in (3a–c):

- (3a) Q: who borrowed a chainsaw?
A: [OScar] **B**ORROWED a **CHAINS**AW
- (3b) Q: what did oscar borrow?
A: oscar **B**ORROWED [a **CHAINS**aw]
- (3c) Q: did oscar buy a chainsaw?
A: (no,) oscar [**B**ORrowed] a **CHAINS**AW

The recordings now contain tokens of the words *Oscar*, *borrow* and *chain-saw* produced with and without sentence stress, which can be directly compared: any difference between the readings must be the consequence of presence versus absence of sentence stress. The difference between stressed and unstressed syllables in the tokens that are produced without sentence stress (out of focus) will then be a matter of word stress only (indicated by bolded small capitals). Examining the effects of word and sentence stress in a single experimental setup using minimal stress pairs can only be achieved by using

⁶ Strictly speaking, correlates of stress can be investigated only in a language with non-fixed stress. In languages with fixed stress, such as Hungarian, where every word has stress on the first syllable, it is impossible to separate correlates of stress from word boundary effects.

highly contrived contexts, for instance, with target words used metalinguistically (as citation forms), as in (4a–d):⁷

- (4a) Q: did you read ‘the import’ or ‘the sale’ again?
A: i read [‘the IMport’] again
- (4b) Q: did you read ‘to import’ or ‘to sell’ again?
A: i read [‘to imPORT’] again
- (4c) Q: did you read ‘the import’ again or write it down?
A: i [READ] ‘the IMport’ again
- (4d) Q: did you read ‘to import’ again or write it down?
A: i [READ] ‘to imPORT’ again

We will now briefly review what has been reported in the literature on the acoustical marking of word and sentence stress. I will draw on publications on Dutch and English but occasionally digress to other languages. We will begin by discussing properties that are found equally in word and sentence stress and finish by zooming in on those properties that differentiate word from sentence stress (and are found, therefore, only when a syllable occurs in a word with sentence stress).

2.2 *Acoustic Properties of Word Stress*

2.2.1 *Temporal Organization* Since the work by Fry (1955) it has been clear that stressed syllables – all else being equal – are longer than their unstressed counterparts. Fry measured the duration of the first and second vowels (V_1 and V_2) in five English minimal stress pairs (noun-verb pairs *contract*, *digest*, *object*, *permit* and *subject*) spoken once by 12 American speakers in sentence-final position in a fixed carrier *Where is the accent in . . .*, which elicits sentence stress on the target words.⁸ With the duration of V_1 and V_2 as predictors, a Linear Discriminant Analysis (Klecka 1980), a classification algorithm often used for this purpose, yields correct classification of stress pattern in 83 per cent of the cases.^{9, 10} After z-normalizing V_1 and V_2 duration within minimal stress pairs, the percentage of correct classification

⁷ A very clever but elaborate way of obtaining minimal stress pairs in English with and without sentence stress was used by Huss (1978).

⁸ The information on the context sentence can be found in Fry (1958: 135).

⁹ This information is not found in the original paper but computed by me (VH). Fry (1955) provides the raw measurements of vowel durations and peak intensities in an appendix. This appendix contains one obvious error in that the duration values for the noun and verb reading of the target word *contract* have been switched (this error becomes apparent when the data are checked against the plot of V_1 against V_2 for *contract* in Fry’s Figure 2 (left panel), which shows the correct durations).

¹⁰ In meta-analyses of the type performed here, it is more customary to quantify the importance of a parameter in terms of effect size. Effect size can be interpreted as the degree of overlap between two samples, for example for words with initial stress versus final stress. The smaller

of stress pattern increases to 93. Using Fry's data, we may apply intrinsic normalization by computing the relative duration of the first vowel ($V_1\%$) as a percentage of the summed durations of V_1 and V_2 . Comparing the V_1 percent values for each of Fry's 60 minimal stress pairs, we find just one single case in which $V_1\%$ was the same for the noun and the verb reading of the pair; in all other 59 cases $V_1\%$ was larger for the noun (initial stress) than for the verb (final stress) reading (98% correct classification). The conclusion was that vowel duration (especially when expressed relatively within a token) is a very good correlate of stress. Fry (1955: 765), however, remarks that consonant duration ratios were 'not materially affected by the shift of stress'. Since word stress is generally believed to be a property of a syllable, this conclusion deserves further scrutiny. I turn to data on Dutch to examine effects of stress on subsyllabic units, that is, vowels, onset and coda consonants, separately.

An early study that examined the effect of stress on the durations of subsyllabic units in Dutch can be found in Nooteboom (1972: appendices 11–12). Target items were non-words /papapap/ and /pʌpʌpʌ/, with short/lax /a/ and long/tense /a/, respectively. These items were spoken with stress on the first, second and third syllable in turn in carrier sentences such that they were either 'accented' (with sentence stress) or 'unaccented' (word stress only). A large number of tokens were produced by each of two male Dutch speakers for each of the 3 (stress position) \times 2 (accentuation) \times 2 (vowel length) = 12 non-word types (between 17 and 26 tokens per type by speaker SG; between 12 and 24 by speaker IS). Duration of all plosives /p/ in positions C_1 to C_4 were measured physiologically (rather than acoustically) using electronic switches that were activated by lip contacts, as were the durations of the vowels in V_1 , V_2 and V_3 . A summary of the results is seen in Figure 1.1. This figure plots the segment durations, in milliseconds (ms), of C_1 , V_1 , C_2 , V_2 , C_3 , V_3 and C_4 , in this order, along the X-axis, with separate lines for items with initial, medial and final stress. The four panels are arranged by vowel length (rows) and by accentuation (columns).

The relative effects of stress on the temporal make-up of the non-words are very similar for accented and unaccented items – although durations are consistently longer overall under sentence stress. Hardly any effects of stress

the overlap, the fewer the number of classification errors. When the distribution of values for two (equally large) samples overlap completely, the percentage of classification errors will be 50; when there is zero overlap, there will be no classification errors. Cohen's d is commonly used as a measure of effect size. It is basically a z-score: if $d = 1$, then the mean of sample A is one time the common standard deviation away from the mean of sample B; there is considerable overlap between the distributions yielding classification errors (assuming equal variance in the samples) in 27.7% of the cases. For $d = 2$, the overlap is much smaller, yielding less than 5% classification errors (Cohen 1988). When the variance in the two samples is not uniform, more sophisticated procedures have to be applied, such as an LDA.

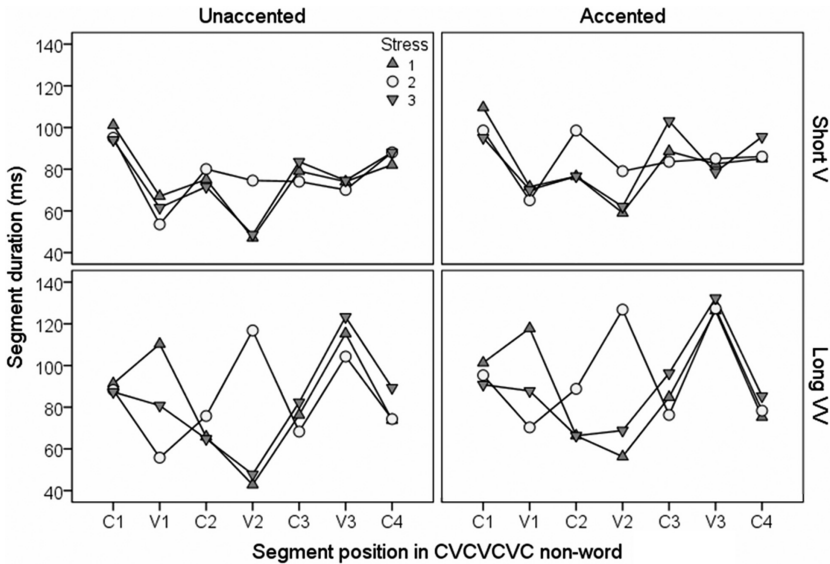


Figure 1.1 Duration (ms) of seven segments in the sequence /pVpVpVp/ as a function of stress position (initial, medial, final) in accented versus unaccented non-words with short (lax) and long (tense) vowels (data from Nootboom 1972, appendices 11–12).

can be seen in the final syllable.¹¹ There are very large differences in the durations of V_1 and V_2 depending on the stress position. When the item is spoken with initial stress, V_1 is very long and V_2 short (ratio $V_1/V_2 > 1$). With medial stress, this pattern reverses completely, with a very short V_1 and a very long V_2 (ratio < 1), while items with final stress have intermediate vowel durations for V_1 and V_2 (ratio ≈ 1).¹² The crucial observation, however, is that the effect of stress position on the duration of the consonant segments, though small in absolute terms, appears to be quite consistent as well: it is nearly always the case that a C, whether onset or coda, is somewhat longer on

¹¹ Segments in a word-final syllable in Dutch are affected by domain-final lengthening but will not be lengthened any further when stressed. Unlike what happens in English, the effects of stress and final lengthening in Dutch are therefore not additive (Cambier-Langeveld and Turk 1999).

¹² The difference between initial and medial stress is very clearly marked by the V_1/V_2 -ratio. The difference between initial and final stress is less clearly marked, especially when the vowels are lax/short. The appendices in Nootboom (1972) do not contain data on individual tokens (only means and number of tokens). Therefore no meaningful effect sizes can be computed.

average in the stressed version of the syllable than in the unstressed version (i.e. in a paradigmatic comparison).¹³

An experiment on a smaller scale involving both words and reiterant non-words in Dutch shows that the lengthening effect of stress is most clearly and consistently seen in the rhyme portions of the syllables (Sluijter and van Heuven 1995). The effect of stress on onset consonants is less systematic or absent.

2.2.2 Intensity The intensity of the sound pressure wave has long been considered as an acoustical correlate of stress. 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 inch (or dynes per cm²). 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; and when the ratio is 1:1000, the difference is 60 dB. So each time the absolute intensity difference is multiplied by 10, there is a 20 dB increase in 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 rise 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 stress. These differences are small but prove reliable correlates (i.e. with little variability) of sentence stress but are even smaller and less reliable when word stress is signalled

¹³ For each of the four consonant positions C₁, C₂, C₃ and C₄, 16 paradigmatic comparisons between stressed and unstressed conditions can be made in Nootboom's appendices. In each of these four positions the stressed version of C is longer than its unstressed counterpart in 15 conditions. A similar count for the vowels in positions V₁, V₂ and V₃ yields longer stressed than unstressed values in 13, 16 and 10 out of 16 comparisons. In this sense, consonant durations are at least as accurate as correlates of stress as are vowels.

(cf. Lea 1977, Beckman 1986 for English; van Katwijk 1974, Rietveld 1984, Sluijter 1995, Sluijter and van Heuven 1996a for Dutch). In all these (and other) studies, peak intensity was measured, 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 stressed vowel) or *average intensity* (as the preceding but normalized for vowel duration). The intensity integral proved a very stable correlate of stress, 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.

Since open vowels have more intrinsic intensity than close vowels (see Section 2.1), using raw peak intensity as a direct correlate of stress is rather pointless. In a paradigmatic comparison, that is, comparing the stressed and unstressed reading of the same vowel in the same position in minimal stress pairs (as in Fry 1955), the stressed version had more decibels than the unstressed counterpart in 52 out of 60 V_1 pairs and in 55 V_2 pairs. Note that the decibel is a logarithmic measure, so that the difference (obtained by subtraction) rather than a ratio (obtained by division) between the (peak) intensities of two vowels (e.g. in a stressed syllable and in an unstressed counterpart) is used here as the correlate of stress. Moreover, it is nearly always the case that the intensity difference between V_1 and V_2 was more positive in the noun reading (with stress on V_1) than in the corresponding verb reading (with stress on V_2). Out of 60 comparisons, 58 behaved as predicted, in one case the relationship was reversed and in one more the noun and the verb reading had the same intensity difference between V_1 and V_2 . This makes (peak) intensity, and especially the intensity difference between stressed and unstressed syllables, a very reliable acoustic correlate of stress in English. It should be pointed out in this context that Fry (1955) is often misquoted. It is not the case that his data show that intensity is a poor *acoustic* correlate of stress or that it is a poorer correlate than duration.

2.2.3 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 Section 1.3, 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 (Figure 1.2). Typically, the closing phase of the glottal period is

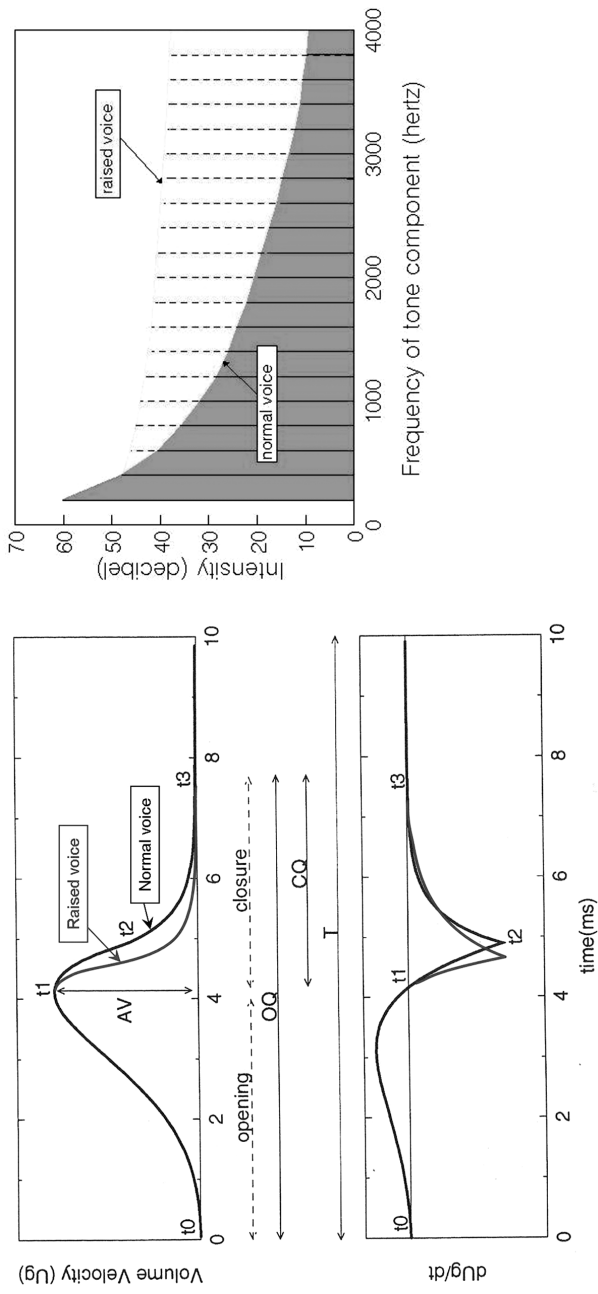


Figure 1.2 Effect of normal versus raised voice on volume-velocity of airflow through glottis (top left) and its first derivative (bottom left). The right-hand panel shows the effect of decreased Open Quotient (OQ) and Closure Quotient (CQ) due to raised voice on the spectral envelope (difference is exaggerated). t1: maximum flow during glottal cycle, t2 fastest decrease of glottal flow, t3 complete glottal closure (no flow). Graphs are based on Sluijter (1995) and van Heuven (2001).

shortened, yielding a smaller opening quotient (the duty cycle of the glottal pulse, that is, 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.¹⁴ 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 and Sundberg 1989). It has been shown that a similar phenomenon can be observed during the production of local vocal effort, that is, during the production of a stressed syllable (Sluijter and van Heuven 1996a for Dutch; Sluijter et al. 1995 for American English; Fant and Kruckenberg 1995, Heldner 2003 for Swedish; Campbell 1995 for Japanese; see also Campbell and Beckman 1995, 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. Undoing the resonance effects of the vocal tract is done by inverse filtering. Inverse filtering software is now readily available (e.g. Airas et al. 2005) but the routines are not included in more comprehensive speech-processing packages. In lieu of full-fledged inverse filtering, some fast-and-dirty approximations have been suggested by Stevens (1998) and were applied in earlier research (Sluijter 1995, Sluijter et al. 1995, Sluijter and van Heuven 1996b). 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 0–0.5 KHz, and three contiguous octave filters: 0.5–1 KHz, 1–2 KHz, 2–4 KHz, cf. Gauffin and 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

¹⁴ When vowel sounds are radiated from the mouth, some $+6$ dB/octave is added to the spectral slope, so that the spectral tilt of an average vowel equals $-12 + 6 = -6$ dB.

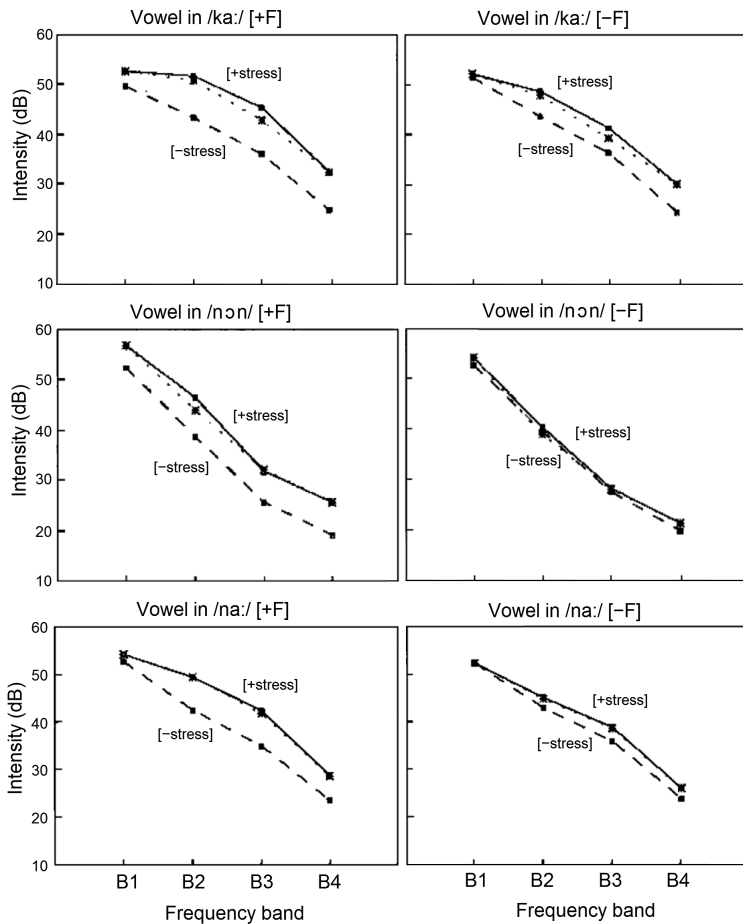


Figure 1.3 Effects of sentence (left-hand column) and word (right-hand column) stress on spectral tilt. Intensity (in dB) is plotted for four frequency bands (B1: <.5 KHz, B2: .5–1 KHz, B3: 1–2 KHz, B4: 2–4 KHz).

was obtained by just measuring mean vowel intensity (at the overall intensity peak) in the 0.5–2 KHz band (Sluijter 1995, Sluijter and van Heuven 1996a).

The effects of stress on spectral tilt at the sentence (left-hand column) and word level (right-hand column) can be seen in Figure 1.3 for a paradigmatic comparison of selected syllables in the Dutch minimal stress pair *canon* ~ *kanon* /'kanɔn ~ ka'nɔn/ 'round song ~ cannon' and reiterant mimicry by five male and five female speakers.

Figure 1.3 shows that generally no effects of stress can be observed in the base band (< .5 KHz). Effects are strong in the higher frequency bands, causing flatter spectral tilt, especially under sentence stress, and more clearly so in the initial syllable than in the final syllable.

2.2.4 Spectral Expansion Stressed vowels have often been described as ‘clear’ (or, spectrally expanded), reflecting greater articulatory effort and precision. These vowels lack the spectral reduction that is characteristic of unstressed vowels. The acoustic consequences of vowel expansion and reduction can be examined by measuring the centre frequencies of the lowest two resonances of the vocal tract, the first and second formants, where F_1 (the lowest resonance) reflects degree of openness of the vowel and F_2 (the second-lowest resonance) reflects vowel backness and lip protrusion (i.e. the length of the oral cavity). Degree of vowel expansion is best expressed in terms of the Euclidean distance of a vowel away from the centre of the (acoustical) vowel space, which is defined by the mean value of F_1 and F_2 found for the individual speaker, when the speaker has produced an equal number of all the vowels in his language (under identical circumstances). For an average male speaker this will be an F_1 at 500 Hz and an F_2 at 1500 Hz.¹⁵ Spectrally reduced vowel tokens will then be closer to the centre of the vowel space than their full or expanded counterparts.

An exemplary study of the effects of stress on vowel quality in Dutch was done by van Bergem (1993). In Dutch the acoustical effects of stress on vowel quality are particularly noticeable – maybe more so than in any other language. Figure 1.4 illustrates the effects of word and sentence stress 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 the centre of gravity of the vowel space. Spectral expansion is largest for vowels pronounced in isolation (‘isol’). Some reduction is 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

¹⁵ Since the F_1 and F_2 values differ considerably from one speaker to the next, especially when the speakers have different gender, normalization is called for when individuals are compared. Z-normalization (Lobanov 1971) is generally seen as the most adequate option. The centre of the vowel space is then by definition at $F_1 = F_2 = 0$. Comparisons across larger numbers of speakers can be safely done without normalization. Often, formant frequencies are psychophysically scaled (through Bark or Mel conversion) so as to reflect properties of the human auditory mechanism, which is more sensitive to differences between low frequencies than to (physically equal) differences between high frequencies (for details, see introductory textbooks such as Hayward 2000 or Johnson 2003).

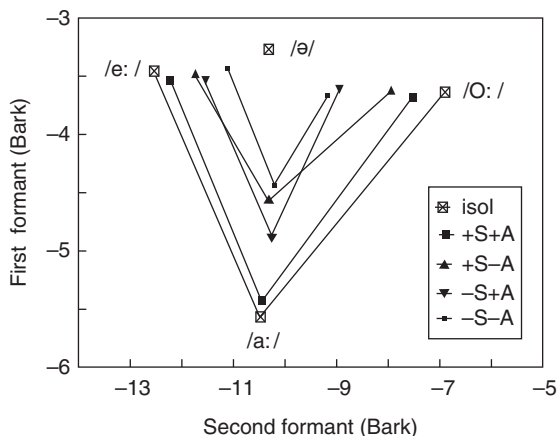


Figure 1.4 F_1 and F_2 (Bark) of three Dutch tense peripheral vowels produced by 15 male speakers in five stress conditions (after van Bergem 1993).

of unaccented words ('-S-A'): here the spectral distance to the centre of gravity /ə/ is minimal. Similar results were obtained for reiterant American-English non-words by Sluijter et al. 1995 (for details, see Sluijter 1995: 116–17, see also Section 2.3).

Automatic classification of stress by spectral expansion of Dutch vowels was done by Sluijter and van Heuven (1996a) in the minimal stress pair /'kanɔn ~ ka'nɔn/ (see Section 2.2.3) and their reiterant versions (/nana/) produced in a short carrier with and without word and sentence stress (four combinations). Predictors in the LDA were the F_1 and F_2 of V_1 and V_2 . Percentages of correct stress identification were 84 and 77 for words with and without sentence stress, respectively, and 68 and 71 for the reiterant non-words. These identification scores are better than chance (= 50%) but are poorer than what was observed for most other stress correlates (see following section).

2.2.5 Resistance to Coarticulation One characteristic of a spectrally expanded stressed syllable is that it shows minimal influence of coarticulation with abutting syllables, which in turn are strongly influenced by the adjacent stressed syllable. So properties of the stressed syllable are anticipated in the preceding syllable, and perseverate into the following syllable, but the stressed syllable itself is hardly influenced by the abutting unstressed syllables. Resistance to coarticulation was claimed to be the most important correlate of stress in Lithuanian by Dogil and Williams (1999; see also Pakerys 1982, 1987).

One way in which the mutual coarticulatory influence of abutting syllables can be quantified would be to locate the beginning and end of vowel-onto-vowel formant transitions (if the formants do not move in synchrony, study the behaviour of F_2 only) from the preceding syllable into the stressed syllable, and from the stressed into the following syllable (cf. Öhman 1967). Then determine the point along the time axis where half of the formant trajectory (i.e. half of the F_2 frequency difference between the consecutive vowels) from the stressed to the unstressed 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 per cent points divided by the duration of the stressed syllable. The larger the relative window size, the more resistant the syllable is to coarticulation. I am not familiar with published data on measurements of resistance to coarticulation.

2.3 *Acoustic Correlates of Sentence Stress*¹⁶

Theories have been proposed in which there is no principled difference between word and sentence stress. In such views, for example, in American structuralism (Bloch and Trager 1942) and early Generative Phonology (Chomsky and Halle 1968, Halle and Keyser 1971), sentence stresses were seen as merely stronger degrees of stress along a continuum, where degrees of stress differ along all stress-related acoustic parameters in proportion. More recently, phonetic research has brought to light, however, that sentence stresses – used to place constituents in focus – are marked in a principally differently way from mere word stresses. Typically, as long as there is no sentence stress on a word, the speaker makes no effort to change the vocal pitch. To be true, there may well be a small rise–fall contour on any vowel (with or without word stress) but this is due to an involuntary response of the glottal mechanism to the greater transglottal pressure that comes about when the oral tract opens during the articulation of the vowel sound; during the articulation of consonants the oral tract is fully or partially closed so that intraoral impedance yields a transglottal pressure drop causing the vocal folds to vibrate more slowly. It has been estimated that the involuntary effect of mouth opening on the rate of vocal fold vibration does not normally exceed a threshold of four semitones (a frequency rise and subsequent fall of less than 25%). Only when a word is produced with sentence stress does the speaker issue a voluntary

¹⁶ This section summarizes work done mainly on English and Dutch, with occasional excursions to other languages, concentrating on research methods and basic findings. In the last 25 years many more languages across the world have been studied using these or similar research methods. It is beyond the scope of this chapter to present a comprehensive overview of findings. The interested reader is referred to work by, for example, Hargus and Beavert (2005), Remijsen and Van Heuven (2006) and Gordon (2011).

command to the glottal muscles that brings about a change in vocal pitch that (greatly) exceeds the four-semitone threshold.¹⁷ Listeners intuitively know that small changes in vocal pitch require no planned action on the part of the speaker and therefore ignore these as a stress cue.

For a pitch change to impart sentence stress on a syllable, the change has to be strictly local, that is, has to take place within a time window that does not exceed the duration of a syllable. Gradual pitch movements (rises or falls that span a longer sequence of syllables) can never be prominence lending ('t Hart, Collier and Cohen 1990). Yet, not every large and fast change in vocal pitch is associated with sentence stress. Fast pitch changes may also be used to mark prosodic boundaries. The difference between prominence-lending and boundary-marking pitch changes is in their timing relative to the segmental structure of the syllable. In Dutch, for instance, an equally large and fast pitch rise located in the first half of a syllable imparts prominence (sentence stress) but it marks the syllable as domain-final (intonation domain boundary or question marker) rather than stressed when executed in the final portion of the syllable (end of rise aligned to end of voicing). The phonetic details of the segmental alignments are quite subtle. Pitch movements (or the component L and H targets) may be synchronized ('anchored') to segmental landmarks or with respect to each other (Caspers and van Heuven 1993, Ladd et al. 1999, Ladd, Mennen and Schepman 2000), the synchronization may be affected by phonological properties of the (stressed) syllable (Dilley, Ladd and Schepman 2005) and differ across languages (e.g. Arvaniti, Ladd and Mennen 1998 for Greek versus Ladd, Mennen and Schepman 2000 for Dutch) and even across dialects within a single language (e.g. van Leyden and van Heuven 2006).

Data collected by Sluijter et al. 1995 (see Sluijter 1995: 106–16 for a more extended report) illustrate the point. Three male and three female speakers of American English each recorded two tokens of four minimal stress pairs (the noun–verb pairs *export*, *uplift*, *digest* and *compact*) as well as their reiterant versions with syllables /bi/, /be/ and /ba/ medially in fixed carrier sentences such that targets received either sentence stress or not. The f_0 peak location was determined in each token as well as the excursion size of the f_0 movement (in semitones). The size of the f_0 change under sentence stress was two to three times larger (in semitones) than in items with word stress only. Most of the f_0

¹⁷ The four-semitone threshold arguably separates so-called micro-intonation from macro-intonation. Micro-intonation comprises all pitch perturbations that are involuntary, that is, caused by intrinsic or co-intrinsic (coarticulatory) interaction between segmental properties and vocal fold vibration. Macro-intonation presupposes voluntary, planned changes in vocal pitch (e.g. 't Hart et al. 1990: 14–15). Differences in intrinsic (i.e. vowel dependent) pitch do not exceed three semitones. F_0 drops in the vicinity of obstruents are limited to four semitones and so is the difference in onset pitch after voiced versus voiceless plosives (for a summary of findings see Di Cristo and Hirst 1986).

movements associated with word stress only were below four semitones. When the token was produced with sentence stress it was nearly always the case that the f_0 peak fell within the confines of the stressed syllable (in fact, without a single exception for the words) affording perfect identification of stress pattern in the four lexical pairs and near perfect stress identification in the reiterant versions (98% correct). However, when the tokens were produced with word stress only (and with a sentence stress on the phrase-final word), the location of the f_0 peak was distributed more evenly over the two syllables and was aligned with the stress in only 65 per cent of the cases (chance = 50%).

Secondary correlates of Dutch sentence stress can be found in temporal organization. It has been shown for Dutch that words with sentence stress are lengthened by some 10 to 15 per cent. Interestingly, all segments – whether stressed or not – in the word are lengthened to the same extent. The lengthening is restricted to only the word that carries the sentence stress; no lengthening spills over to adjacent words even if these are within the focus domain headed by the target – indicating that the lengthening is a correlate of sentence stress rather than of focus (Eefting 1991, van Heuven 1998).¹⁸ Languages appear to differ in the domain they use for lengthening under sentence stress. It has been found for English that this domain is the within-word foot (excluding pre-stress syllables from the lengthening domain) rather than the (morpho-syntactic) word (Turk and Sawush 1997). I am not familiar with attempts to evaluate the effectiveness of lengthening effects in automatic identification of sentence stress. My expectation would be that the contribution of accentual lengthening will be minor.

2.4 *Relative Strength of Stress Correlates*

The relative strength of acoustic correlates of stress (or of any other linguistic distinction) can be estimated by applying some technique to compute effect size (see Section 2.2.1). However, when the statistical distributions of acoustic correlates differ between samples, as they do in our case when correlates are measured for a sample of words with initial stress and a second sample of the stress partners with final stress (i.e. members of minimal stress pairs), more complex techniques are called for. I find it expedient to use the LDA automatic classification algorithm as an estimator of effect size. The number of (above chance) classification errors would then serve as a good approximation of the relative strength of an acoustic correlate of stress. Normally, LDA uses multiple

¹⁸ Van Heuven (1993) made a further attempt to ascertain whether the lengthening effect in Dutch would extend to compounds or would be confined by the word boundary inside the compound. The results were inconclusive. This is still an area for further research.

predictors to classify objects into categories. So, it seems tempting at first sight to have the algorithm make its classification with all measured acoustic correlates of stress in one run. This, however, defeats the purpose of the exercise. The acoustic properties of stress are generally correlated, some moderately, others more strongly. The LDA removes the shared variance from all but the most successful predictor, so that we will not get a true view of the effect sizes of the less successful predictors. Therefore, we routinely run the LDA with single predictors, repeating the procedure as many times as there are predictors. Only in this way can the percentages of successful classification be meaningfully compared. It is also necessary to instruct the LDA to assume equal probabilities for the two categories it has to predict (stressed, unstressed) rather than to compute a priori probabilities from the actual frequencies in the input data.

Sluijter and van Heuven (1996a) applied the LDA to the classification of initial and final stressed members of reiterant minimal stress pairs produced with and without sentence stress by six native speakers of American English.¹⁹ Analyses were run separately for word stress (targets outside focus) and sentence stress (targets in focus). Predictors were in both conditions: (i) the location of the f_0 peak (in first or second syllable), (ii) relative duration of the first syllable, (iii) difference in peak intensity between the syllables, (iv) the difference in Euclidean distance of the vowel from the centre of the formant space and (v) the difference between the two syllables in five glottal parameters (a) B_1 (an estimate of completeness of glottal closure), (b) estimated tilt of source spectrum based on fundamental and amplitude of F_2 , (c) tilt based on difference between fundamental and amplitude of F_3 , (d) difference in Open Quotient (OQ estimated by the difference in amplitude between fundamental and second harmonic) and (e) difference in amplitude of voicing (= amplitude of fundamental). The results are presented in Figure 1.5.

F_0 , duration and intensity afforded very good classification of stress pattern for sentence stress (above 95% correct), vowel quality yielded only 80% correct classification. The estimated glottal source parameters afforded between 69 and 79% correct classification (the latter for spectral tilt between fundamental and F_2), with an exception of amplitude of the fundamental, which yielded 97% correct and was in fact slightly better as a predictor than just overall peak intensity). Much poorer classification was obtained for word stress (in words out of focus). Location of the f_0 peak, intensity, OQ and amplitude of fundamental were all between 60 and 65 per cent correct (chance = 50%). B_1 and the two tilt measures were at 75 per cent correct. The best classification was given by duration and vowel quality (both at 80%).

¹⁹ See also Sluijter (1995: chapter 6) for a more detailed report of this study.

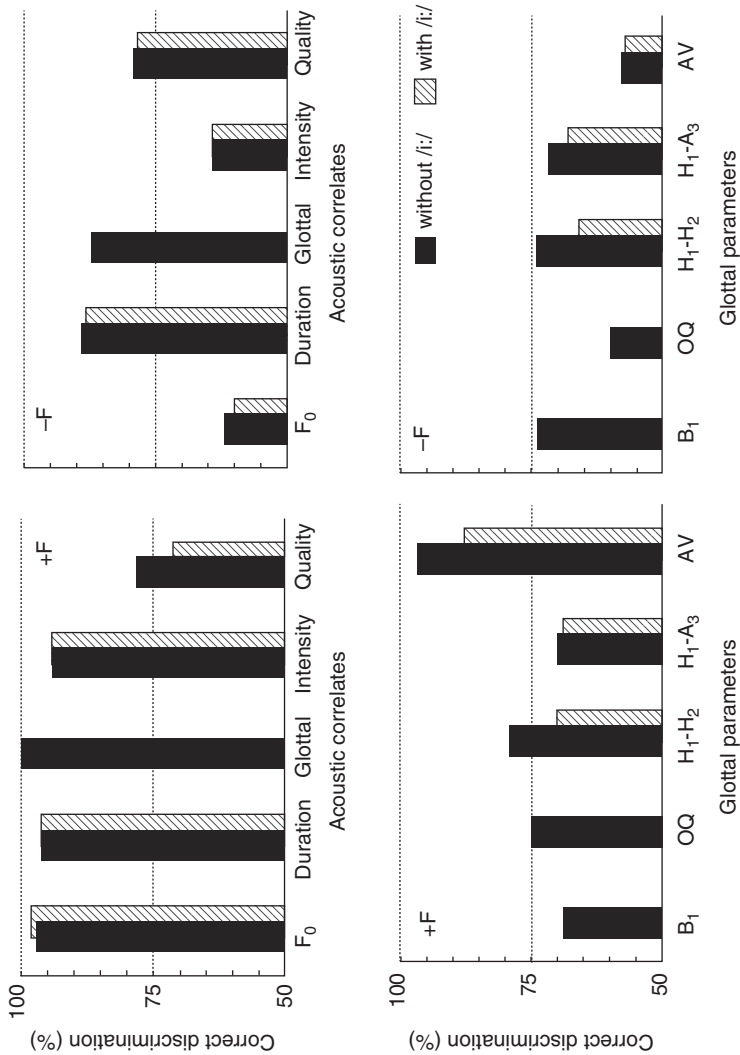


Figure 1.5 Correct classification (%) by Linear Discriminant Analysis of sentence stress (+F, left-hand panels) and word stress (-F, right-hand panels) in American-English reiterate disyllables /bibi/, /bebe/ and /baba/. Predictors are the traditional acoustic correlates of stress (top panels) or estimates of glottal parameters (bottom panels) (see Sluijter and van Heuven 1996b for details).

A provisional conclusion from this comparison of parameter strengths would be that the difference between initial and final stress is more clearly marked in English when it is a matter of sentence stress than when we are dealing with just word stress. The effect sizes of the parameters differ substantially between sentence stress and word stress. The location of the f_0 peak, peak intensity and amplitude of the fundamental are strong correlates in the sentence stress condition but not for word stress. Duration is a reliable correlate in both conditions, and so is spectral quality – be it less reliable than duration. The spectral tilt measures are only moderately successful correlates.

3 Perceptual Cues of Word and Sentence Stress

In the preceding sections we have seen that both word and sentence stress are acoustically marked by at least five different correlates. In an acoustic study it is quite possible to determine the relative strength of each variable as a successful correlate of stress, simply because the categorization of utterances is done by a computer algorithm and does not require the services of human listeners. When we want to establish the perceptual relevance simultaneously of all the acoustic correlates for a human listener, the problem arises that the experiments become unmanageably large and time consuming – a burden especially on the part of the human subjects. The practical solution is that the experiments are simplified in either of two ways (but hybrids between these two types also occur): (i) only two or three parameters at the most are included in the stimulus materials with maximally seven steps (values) along each parameter ($7 \times 7 \times 7 = 343$ stimulus types can be presented once in about half an hour) or (ii) more parameters are systematically varied but the number of steps for each parameter is severely limited, typically to two values – one realistic for stress and the other for no stress, often implemented as a straightforward exchange of values between the two extremes as found in natural tokens. Classical examples of type (i) studies are Fry (1955, 1958), Morton and Jassem (1965), and Mol and Uhlenbeck (1956). Type (ii) studies were done more recently by, for example, Beckman (1986) for English and Japanese, Ortega-Llebaria, Vanrell and Prieto (2010) and Ortega-Llebaria and Prieto (2011) for Spanish and Catalan.

Limiting the presentation, again, to just Dutch and English, we will now review the perceptual cue value for human listeners of the stress correlates discussed in Section 2. These studies typically compare the cue value of pairs of acoustic correlates in relatively small sets of stimuli. For instance, Fry published a series of three experiments comparing the perceptual strength of vowel duration (as a baseline condition) with that of three other parameters,

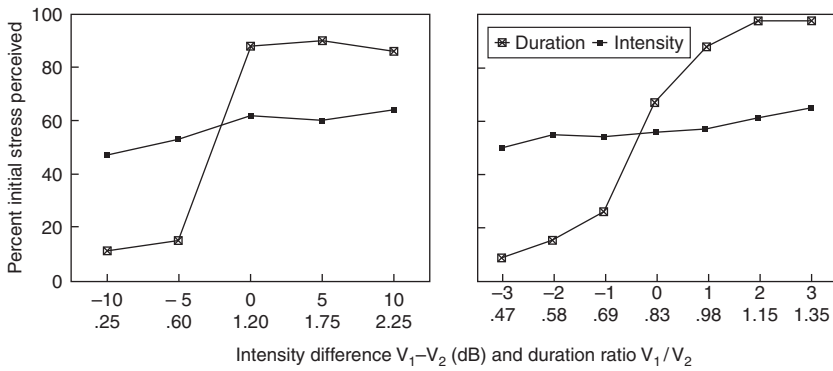


Figure 1.6a Initial stress perceived (%) as a function of intensity difference between V_1 and V_2 (in dB) and of duration ratio between V_1 and V_2 in minimal stress pairs in English (after Fry 1955).

Figure 1.6b As Figure 6a but for Dutch (after van Heuven and Sluijter 1996).

viz. peak intensity (1955), fundamental frequency (1958) and vowel quality (1965).²⁰ If done properly, the three experiments should yield a rank order of perceptual importance for the four correlates.

3.1 Duration versus Intensity

Figure 1.6a (left-hand panel) shows the main results of the perception study by Fry (1955). In the experiments, the durations of V_1 and V_2 in each of five minimal stress pairs (*object*, *subject*, *digest*, *compact*, *import*, see Section 2.2.1) were varied in five steps between (and including) values found (averaged over ten speakers) in natural tokens with initial and with final stress. These five duration steps were systematically combined with five intensity differences (by amplifying V_1 and at the same time attenuating V_2) such that the $V_1 - V_2$ difference varied between +10 and -10 dB. Listeners had to indicate for each of the 5 (word types) \times 5 (vowel duration ratios) \times 5 (intensity differences) = 125 stimulus types whether they perceived it as a noun (initial stress) or as a verb (final stress). Unfortunately, Fry did not present the results for the individual stimulus types. Instead, Figure 1.6 (after Fry's Figure 3) presents percent perceived initial stress for the five duration steps (averaged over words and intensity steps) and for the five intensity steps (averaged over words and duration ratios).

²⁰ Because the cue value of vowel quality was very weak, Fry (1965) limited the range of duration variation severely relative to the earlier two experiments.

The results show a cross-over from stress perceived on the first syllable to the second syllable. The cross-over takes place between duration steps two and three and is both steep (within one stimulus step) and convincing (75% agreement on either side of the boundary). In contrast to this, the intensity difference is inconsequential: although there is a gentle trend for more initial stress to be perceived as V_1 has more decibels than V_2 , the difference is limited to some 20 percentage points; the boundary width, which can only be estimated by extrapolation, would be some 15 times wider than for duration. This shows that duration outweighs intensity in Fry's experiment roughly by a factor of 15.

Figure 1.6b (right-hand panel) shows the results of a similar experiment run by Sluijter, van Heuven and Pacilly (1997) for a single Dutch minimal stress pair: the reiterant non-word *nana*. The results are the same as in English: a complete cross-over is obtained by varying the vowel durations, while the intensity difference does influence stress by a small amount only, certainly not enough to bring about a cross-over. There are more and smaller stimulus steps, which makes the cross-over appear somewhat more gradual. Also, the targets were presented medially in a sentence frame *wil je [target] ZEGgen* 'will you [target] SAY' with the sentence stress on the final verb; the stimulus variations were suggestive of word stress only – the range of intensity differences in the Dutch stimuli was much smaller (but reflected actual speech production) than that in Fry's English materials with sentence stress on the targets.

Figure 1.7a is a quasi three-dimensional plot of per cent initial stress (numbers in the circles at the X–Y coordinates) perceived as a joint function of the difference in vowel duration (seven steps along the X-axis) and of the difference in intensity (seven steps along the Y-axis). The boundary in the figure separates the white area with a majority of initial-stress decisions from the dark area with a majority of final stress responses. The boundary is defined as a straight line; it is the discriminant function that is computed by an LDA that optimally predicts stress responses from the X and Y predictors. The discriminant function defines all combinations of X and Y values for which the stress response would be undecided (50–50%) – it is a two-parameter category boundary.

If the boundary runs at a 45° angle, the X and Y parameters would be of equal strength. In the figure the boundary runs at an angle that is much steeper than 45°, though not completely vertical. The steep angle indicates that the duration parameter outweighs the intensity parameter as a stress cue. The figure also shows that intensity variations are largely inconsequential: they cannot swing the majority decision from initial to final stress for six out of seven duration steps; only when $V_1 = 170$ ms and $V_2 = 245$ ms does intensity yield a (shallow) cross-over from 43 to 60 per cent initial-stress responses.

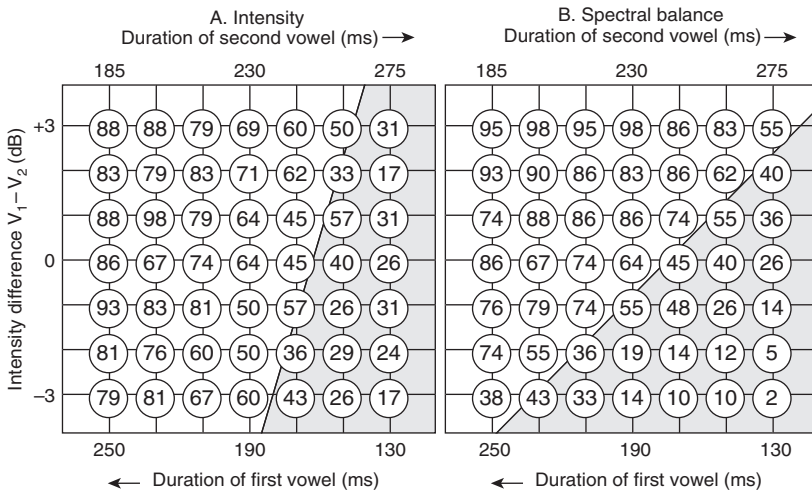


Figure 1.7a–b Percentage of initial stress perceived as a function of temporal structure (duration of V_1 and V_2 , horizontal) and of intensity difference (vertical). In panel A the intensity in V_1 and V_2 was varied uniformly (amplification/attenuation of gain factor); in panel B intensity variations were made selectively at frequencies above 500 Hz only (yielding differences in spectral tilt).

3.2 *Duration versus Selective Intensity (Affecting Spectral Slope)*

Sluijter et al. (1997) also included a set of stimuli in which the same intensity differences were generated on V_1 and V_2 but in such a way that no differences were made at frequencies below 500 Hz and all the changes were concentrated at frequencies above 500 Hz, thereby creating a change in spectral slope. This selective manipulation of intensity is a more realistic model of what a human speaker does when producing differences in loudness between vowels (see Sections 2.2.2–2.2.3). The results now show that (selective) intensity differences (affecting spectral tilt) are as strong a stress cue as are the duration differences: the boundary now runs at a 45° angle. In this experiment, the stimuli had been presented over headphones with artificial reverberation added. The reverb (which was realistic of room acoustics) obscures the temporal details in the stimulus. When the same materials were presented over headphones without reverb, the effects of selective intensity (affecting spectral tilt) were smaller than those of duration but still larger than those of uniform intensity differences (not affecting spectral tilt).

3.3 *Contribution of Consonant versus Vowel Duration*

Now that we have seen that duration generally outweighs other cues for word stress, let us examine the effects of the duration of subsyllabic units such as the onset consonant, the vocalic nucleus and the coda consonant. An experiment that addresses this issue was reported by van Heuven (2014). In reiterant stimuli, with short/lax vowels (/paʔpaʔ, tastaʔ/), and with long/tense vowels (/paʔpaʔ, tastaʔ/) the durations of onset, nucleus and coda were varied separately in steps of 50, 75, 100, 125 and 150 per cent of the original duration. The stimuli were synthesized from diphones which had been excerpted from stressed syllables produced in nonsense words with sentence stress, so that all original segments were equally suggestive of (strong) stress.

Figure 1.8 plots the percentage of perceived initial stresses as a function of the duration manipulation (shortening or lengthening by 0, 25 or 50% of the original segment duration) of the onset, nucleus or coda segment in first or second syllable with tense (long) versus lax (short) vowels.

Figure 1.8 shows that, overall, effects of changing the duration of the vocalic nucleus are large but changes in consonant durations, whether in the onset or in the coda, have little or no effect on stress perception. A complete cross-over from stress perceived on the first syllable to stress perceived on the second syllable is found for vowel duration change, except when the vowel is phonologically short (lax) and in the final syllable of the target non-word (top-right panel). Moreover, the effect of changing the (vowel) duration is weaker overall when the changes are implemented in the second (final) syllable than in the initial syllable.²¹ Changing the duration of a consonant only affects stress perception if the change takes place in a word-initial syllable with a short (lax) vowel (top-left panel) but even then the effect is still somewhat smaller for consonants than for the vowel. In this condition, it does not matter whether the consonant is in the onset or in the coda. So, it seems safe to conclude that the older literature was right in assuming that vowel duration by itself, rather than syllable duration or rhyme duration, is the relevant duration cue for stress perception.

3.4 *Duration versus Vowel Quality*

The only study on the effect of vowel quality on stress perception in English was done by Fry (1965). Fry manipulated the formants of vowels in four minimal

²¹ It is probably the case that varying the relative duration of a long vowel (such as /a/) is more noticeable, and influences stress perception more strongly, than applying the same relative duration changes to a short (lax) vowel. Also, if the changes are applied to the final syllable, the duration effects of stress are in competition with the effect of word-final lengthening, which may detract from the efficacy of duration as a stress cue in word-final syllables. The latter effect has been found on several other occasions in Dutch (e.g. Nootboom 1972, Cambier-Langeveld, Nespor and Van Heuven 1997).

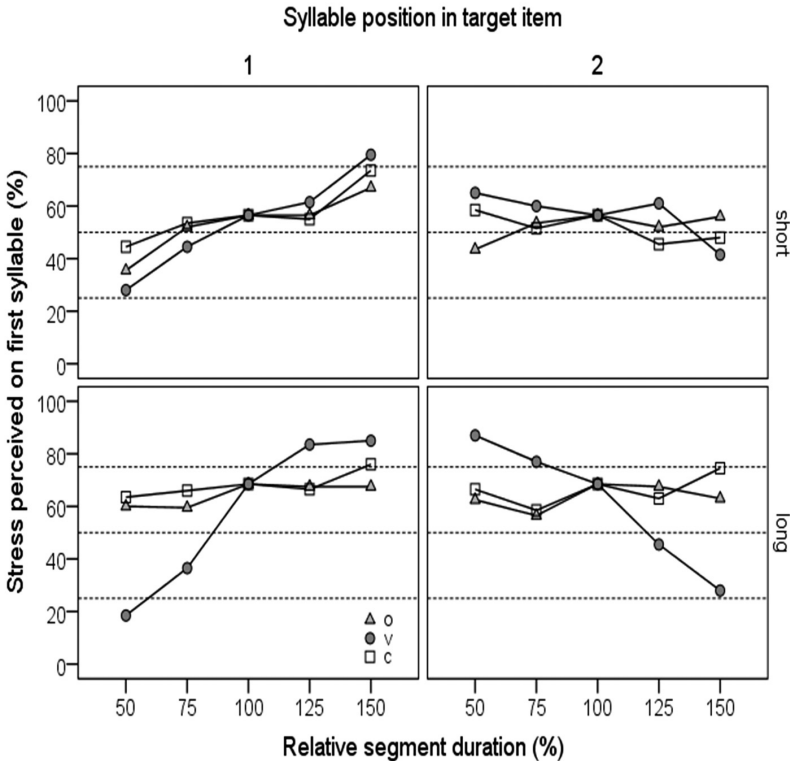


Figure 1.8 Per cent stress perceived on first syllable as a function of relative duration of manipulated segment (onset, vocalic nucleus, coda) in either first (left panels) or second (right panels) syllables with short/lax (upper panels) or long/tense (lower panels) vowel. Target segments are embedded in reiterant CVC.CVC non-words.

stress pairs (*contrast*, *digest*, *object*, *subject*). While keeping pitch and intensity differences constant, the duration ratio and formant structure of V_1 and V_2 were varied in three steps for each parameter, creating a $3 \times 3 = 9$ item stimulus space for each noun-verb pair, that is, 45 stimuli in all. Formants F_1 and F_2 in V_1 were manipulated for three words pairs (*contrast*, *digest*, *object*) while keeping V_2 constant; formants in V_2 were varied in *object* and *subject* while keeping V_1 constant. The formant manipulations were such that either F_1 or F_2 or both moved just one step closer towards the centre of the vowel space (suggesting vowel reduction). No attempt was made to systematically create multiple steps of equal magnitude along a spectral reduction/expansion continuum. Figure 1.9a shows Fry's results. In the figure, the duration and formant changes have been plotted

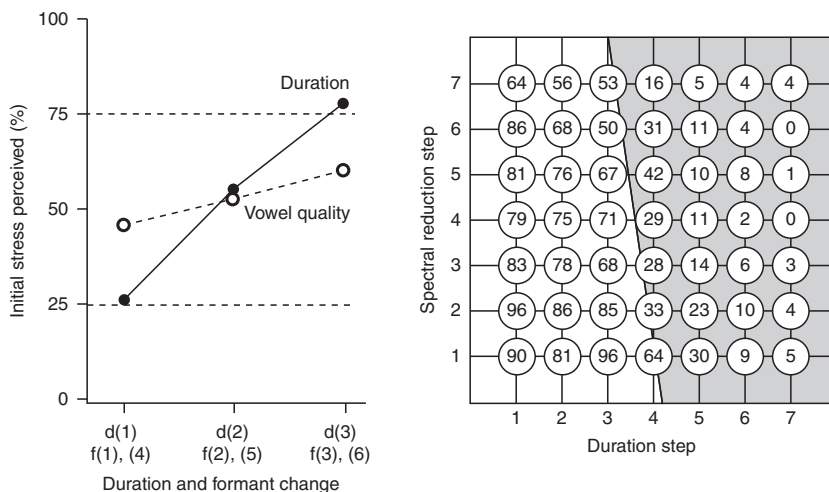


Figure 1.9a Percentage initial stress perceived in English as a function of V_1/V_2 duration ratio and of vowel reduction in F_1 (left), F_2 (middle) or both (right) in either V_1 (steps f1–f3) or V_2 (steps f4–f6) (after Fry 1965).

Figure 1.9b Percentage of initial stress perceived in Dutch as a function of V_1/V_2 duration ratio and spectral reduction in V_1 (from van Heuven and de Jonge 2011).

such that stimulus steps cue initial stress more strongly going from left to right. The results indicate that changing the vowel quality has a systematic but small effect such that stress is less likely to be perceived on the syllable with reduced vowel quality; the tendency is somewhat stronger when the vowel quality is reduced in the F_2 dimension (backness and rounding) than in the F_1 dimension (height) and is strongest when both quality dimensions are affected simultaneously. The effect of vowel quality is small, and does not yield a convincing cross-over: initial stress percentage changes from 45 to 60. The effect of vowel duration is clearly much stronger. Even with the smaller range of duration variation adopted in this experiment, there is a convincing cross-over spanning more than 50 percentage points.

Fry (1965) did not vary vowel quality in terms of an acoustic continuum. A more direct comparison of vowel duration (temporal expansion/reduction) and vowel quality (spectral expansion/reduction) was made for Dutch by van Heuven and de Jonge (2011), who varied the V_1/V_2 ratio and the vowel quality of V_1 in the Dutch minimal stress pair *canon* ~ *kanon* (see Section 2.2.3) in seven steps along each continuum. Targets were presented in post-focal position (no f_0 movement on the target) in a carrier *ik heb GISteren een canon (kanon) gehoord* /ɪk hɛp [ˈɣɪstərən]_{+F} ən ˈkanɔn (kaˈnɔn)

ˈɔhɔrt/ ‘I have yesterday a canon (cannon) heard’, that is, ‘I heard a canon (cannon) yesterday.’

The results are shown in Figure 1.9b, in quasi-3D format. Obviously, convincing cross-overs are obtained for the duration steps. Just one, very incomplete, change from perceived initial stress to final stress is obtained by changing vowel quality from clear to fully reduced to schwa; this change is obtained only when the duration cue is ambiguous, that is, at duration step four. Fry’s conclusion is confirmed here: vowel reduction is clearly a much weaker stress cue than vowel duration.

3.5 Duration versus Fundamental Frequency

Let us, finally, examine the perceptual effects of varying the size and segmental alignment of f_0 changes as a cue for stress. As I pointed out earlier, in natural human speech the f_0 change has to exceed a certain threshold (say ≥ 4 semitones) in order to function as a stress cue, and if it does it typically imparts sentence stress on the word that carries the f_0 change. Since sentence stress outranks word stress, this makes the f_0 change the strongest stress cue of all. Fry (1958) was among the first to examine the effect of f_0 change on stress perception, comparing its strength with that of varying the duration ratio of V_1 and V_2 in the English noun–verb pair *subject*. The duration ratio was varied as in Fry (1955) in five steps covering the natural range of duration variation found for this word pair. In one experiment, Fry synthesized the syllable *sub-* on a flat 97 Hz followed by stepwise f_0 rise to *-ject* of 5, 10, 15, 20, 30, 40, 60 and 90 Hz. This set of eight rises was supplemented with a similar set of eight falls, with the level higher f_0 on *sub-* and the low 97 Hz pitch on *-ject*.²² The total set of 5 (V_1/V_2 ratios) \times 8 (step sizes) \times 2 (directions) = 80 stimuli was judged for stress position (noun or verb) by a mixed group of 41 American and English native listeners. The results bear out that the frequency step-up generated stress on the second syllable (between 61 and 75% for the various f_0 changes but averaged over duration ratios) whilst a step down yields stress on the first syllable (between 48 and 80%), that is, the higher-pitch syllable is heard as stressed. The absolute size of the step, however, was inconsequential: a 5 Hz increment was as influential as a 90 Hz change. On average, however, the effect of changing f_0 turned out to be smaller than that of varying the duration ratio. Unfortunately, Fry does not give the full breakdown of results for each combination of duration ratio and f_0 change so that we cannot check to what extent the f_0 change can be counteracted by the vowel duration ratio.

²² Fry (1958) does not explain how the pitch change was implemented. Presumably, it was a step function yielding an (almost) instantaneous rise or fall. Although such abrupt changes cannot be produced by the human speaker, listeners do respond to them as if they are speech-like (see Isačenko and Schädlich 1966).

In a second experiment, Fry (1958) combined the five vowel duration ratios with 16 different f_0 contours. The contours were more realistic approximations of English intonation patterns. The f_0 change was not an instantaneous step up or down on the syllable boundary but a rise or fall that extended over a certain time span. Rises and falls were executed either over the entire vowel duration or started at the temporal midpoint of the vowel. F_0 movements were linear changes as a function of time, between 97 and 130 Hz. The choice of contours was rather arbitrary and has no systematic structure. Nevertheless, some regularities can be observed in the results. Two f_0 contours always yield initial stress, even if the duration ratio strongly suggests final stress. Three contours always yield a majority of final-stress judgements. In the remaining 11 contours there was always at least one duration ratio that could swing the stress from initial to final, thereby counteracting the effect of f_0 . In retrospect it seems reasonable to exclude a number of the f_0 contours on the grounds that they do not constitute legal intonation patterns on a single word in English. For instance, it seems impossible to have a fall on the first syllable followed by a rise on the second (or a rise followed by a fall). Other patterns have late rises in the final syllable, which suggest question intonation rather than sentence stress. Be this as it may, some patterns, however, seem perfectly plausible. Contour A is basically an H*L on the first syllable while contour B is the same H*L but synchronized with the second syllable. Pattern E is an H* followed by a level boundary. Patterns M and N have an H*L on the second syllable with a short plateau between the H* and the L. A and E always generate a majority of initial stress responses while B, M and N always have a majority of final stress judgements. So, our conclusion should probably be that an f_0 change involving a properly aligned H* target attracts a majority of stress responses that cannot be counteracted by any duration ratio. Even for the most extreme duration ratios, however, there is always an f_0 pattern that can swing the judgements from initial to final stress.

Bolinger (1958) reports four experiments in which f_0 and intensity were varied as cues for sentence stress in short English sentences made up of (mainly) monosyllabic words. The overall result is that f_0 changes attract stress judgements much more than intensity differences do. In the crucial experiment Bolinger systematically combined five temporal organizations of the minimal stress pair *undertaking* (initial stress: 'what a mortician does'; penultimate stress: 'enterprise') with 16 different f_0 patterns.²³

²³ The choice of duration variations is not explained by Bolinger. The ordering of the five duration patterns in Table 1.1 seems to have been done on the basis of the duration of the final syllable (which is not a candidate for stress). The f_0 patterns were probably obtained by systematically combining pitches of 80, 90, 100 and 110 Hz on the first and third syllables, yielding $4 \times 4 = 16$ combinations. The f_0 on the second and fourth syllables was probably kept constant at the 80 Hz flat baseline. Unfortunately, Bolinger specifies only three of the 16 f_0 patterns (see Table 1.1). The number of listeners in Bolinger's experiment was maximally eight.

Table 1.1 *Initial stress perceived (%) on undertaking for five temporal organizations (onset and rhyme durations in ms) systematically combined with three f_0 patterns (level f_0 on successive syllables, in Hz) (after Bolinger 1958: 126)*

Onset and rhyme durations (ms)							F ₀ on successive syllables (Hz)			Mean
un	d	er	t	a	k	ing	80–80– 80–80	90–80– 90–80	100–80– 90–80	
170	65	170	65	155	70	320	88	29	38	52
210	40	95	60	180	30	310	50	29	25	35
230	30	80	55	200	50	260	50	00	13	22
145	15	80	85	215	50	250	13	14	50	26
180	30	70	45	170	25	250	75	14	75	57
Mean							55	17	40	38

Bolinger does not present the results in full but summarizes as follows (1958: 125): ‘In all but 3 patterns the majority of the listeners reacted as the experimenter had predicted on the basis of pitch, and in only one of the 3 could the discrepancy be correlated with duration.’ The three pitch patterns that did not conform to the hypothesis (probably, the higher f_0 will attract stress, and later highs will be more prominent *ceteris paribus*) are the only results specified in full. Table 1.1, curiously enough, shows two pitch patterns for which duration yields a full cross-over (from 13 to 88% initial stress in the level f_0 pattern, and from 13 to 75% in the 100–80–90–80 pattern). The 90–80–90–80 pattern yields a preponderance of penultimate stress judgements, which is what we would predict from the hypothesis.

Van Katwijk (1974: 76–88) varied f_0 movements in a Dutch reiterant nonsense item /scæscæscæ/ in a rather realistic fashion. F_0 changes were implemented relative to a fixed declination of 5 st/s. Keeping intensity, quality and duration constant, f_0 rises and falls of 3 st during 100 ms were generated at 11 different time points in the stimulus, as indicated in Table 1.2. In this table the alignment is specified for the onset of the f_0 movement with respect to the duration of a segment. For instance, ‘V₁ 00’ means that the f_0 movement begins at 0% of the duration of the first vowel, that is, at the vowel onset. Van Katwijk also generated three stimuli with rise–fall contours, and two stimuli (one rise, one fall) with 6-st excursion sizes (during 200 ms). Another 15 stimuli, with multiple f_0 movements, will not be discussed here.

The results show that the location of the f_0 movement greatly influences the perception of stress. A simple pitch rise or a rise+fall combination located at the beginning of a syllable (preferably beginning before the vowel onset) suffices to attract a clear majority of stress responses to that syllable (indicated by

Table 1.2 Number of (sentence) stresses perceived on first, second and third syllable in the Dutch nonsense word /sæ:sæ:sæ:s/ by 45 Dutch listeners (free choice) for 27 f_0 configurations (rise, fall, rise–fall) with 3-st and 6-st excursion sizes (after van Katwijk 1974: 81–3)

Rise 3 st				Fall 3 st				Rise–fall 3 st			
align	S ₁	S ₂	S ₃	align	S ₁	S ₂	S ₃	align	S ₁	S ₂	S ₃
V ₁ 00	31	6	9	C ₁ 50	22	5	6	V ₁ 00	41	2	6
V ₁ 25	33	10	4	V ₁ 00	36	4	1	C ₂ 75	9	42	3
V ₁ 50	13	35	6	V ₁ 50	36	4	8	C ₃ 50	5	4	38
C ₂ 00	8	39	10	C ₂ 25	37	8	6	Rise 6st			
C ₂ 50	6	44	6	C ₂ 75	35	6	7	C ₁ 00	7	44	21
V ₂ 00	4	37	6	V ₂ 00	22	18	4	Fall 6st			
V ₂ 50	11	18	32	V ₂ 50	19	17	18	C ₂ 50	35	20	24
C ₃ 00	16	4	43	C ₃ 25	20	9	19				
C ₃ 50	14	2	45	C ₃ 75	17	4	32				
V ₃ 00	16	3	39	V ₃ 00	12	3	28				
V ₃ 50	22	4	6	V ₃ 50	17	2	4				

shading in Table 1.2). Simple falls tend to attract fewer stress judgements than rises do, especially when they are associated with the medial or final syllable.²⁴ For a simple f_0 fall to impart stress on a syllable it has to be aligned rather late in the syllable or even in the beginning of the next syllable. The complex rise–fall movement does not attract more stress judgements than a simple rise; long 6-st rises and falls do not attract more stress judgements than 3-st exemplars. Van Katwijk also generated stimuli with differences in vowel duration (lengthening either V₁, V₂ or V₃ by 30%) and in intensity (adding 5 dB to each vowel in turn), but never in combination with f_0 or with each other so that there are no stimuli in which stress parameters contradict one another.

3.6 Perceptual Cues: Conclusion

The above pairwise comparisons of stress cues in English and Dutch lead to the overall conclusion that the most important perceptual cue for stress is a change in fundamental frequency (if properly aligned with the segmental structure). The second-most influential cue is temporal organization, specifically the

²⁴ The responses indicate a strong bias favoring stress on the initial syllable. There may be two reasons for this initial stress bias: (i) listeners tend to perceive the onset f_0 as a rise from the speakers (inferred) lowest pitch; the higher the onset f_0 the more initial stress is perceived (Van Heuven and Menert 1996) and (ii) the stimuli with a simple fall begin with a high boundary tone %H, which listeners often mistake for a sentence stress. Unfortunately, Van Katwijk does not specify the onset f_0 of his stimuli.

duration ratio between the stressed and the unstressed version of the vowels (rather than of the consonants). Intensity would seem to rank third, but only if it is implemented such that the gain or loss of intensity is concentrated in frequency bands above 500 Hz, thereby affecting the slope of the spectrum (the flatter the spectrum, the greater the perceived loudness). Overall intensity and vowel quality are found to be the weakest stress cues. It is not clear from the experiments reviewed which would be the weaker of the two.

I should point out, however, that the older literature on perceptual cues of stress does not really substantiate the claims it makes. Both Fry and Bolinger (for English) as well as van Katwijk (for Dutch) insist that f_0 change is a stronger stress cue than duration but the claim seems tenuous, either because the experiment does not allow the conclusion to be drawn, or because the crucial data were not presented. Although Fry's (1958) results provide at least circumstantial evidence, it is not the case that an f_0 change can never be overridden by temporal cues in his materials.

The most important conclusion is that the strength of acoustic correlates of stress and the perceptual cue value of these correlates are not rank ordered in a one-to-one fashion. This has two reasons. First, the location of an f_0 change is a strong correlate of stress in speech production but only if the f_0 change exceeds a threshold of three to four semitones and if it is appropriately aligned with the segmental structure. When words do not receive sentence stress, the f_0 change is no longer a reliable correlate. For f_0 change to be a perceptual cue, no such threshold seems to be required. A change from 97 to 104 Hz is enough to evoke final-stress perception, while a fall of the same magnitude yields initial stress. Therefore, f_0 change may be perceptually the strongest cue but it is acoustically unreliable. Second, the human listener does not rely on uniform intensity differences between stressed and unstressed syllables. This probably makes intensity the weakest perceptual cue of all, even though it is acoustically quite reliable.²⁵ Differences in vowel duration are both perceptually strong and acoustically highly reliable, for both word stress and sentence stress.

4 Cross-linguistic Differences in Phonetic Marking of Stress

There has been some speculation on the question whether or not any language that uses the linguistic parameter of stress also uses the same correlates, with the same order of relative importance of these acoustic correlates and as cues to stress perception. The general feeling is that different correlates (and different perceptual cues) are employed depending on the structure of the language

²⁵ The relative unresponsiveness of the human hearing mechanism to differences in intensity has been known for over a century. The first to comment on this phenomenon was Saran (1907), see also Mol and Uhlenbeck (1956, 1957: 346) and Bolinger (1958: 114).

under analysis. We will discuss two sets of differences between languages, and their consequences for stress marking. The first set of differences concerns the type of stress system a language employs, whereas the second source of difference is located in the relative exploitation within a language of stress parameters for other linguistic contrasts.

4.1 *Contrastive versus Demarcative Stress*

It seems reasonable to assume 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. 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.²⁶

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.²⁷ Although hardly any research has been done to check these predictions, there is some evidence that the basic prediction is correct. Dogil and Williams (1999) present a comparative study of Polish (fixed penultimate stress) and German (quantity-sensitive plus lexical stress) stress marking, and conclude that stress position is less clearly marked in Polish. Similar results were found more recently in a strictly controlled cross-linguistic study of Spanish and Greek (with contrastive stress) versus Hungarian (fixed initial stress) and Turkish (fixed final stress) by Vogel, Athanasopoulou and Pincus (2016). Their results (Figure 1.10) show that the same set of acoustic stress parameters (applied in the same manner across the four languages) affords good to excellent automatic classification of stressed and unstressed syllables at the word level for the two contrastive-stress languages but not for the fixed-

²⁶ Although the case could be made that stress in fixed-stress languages 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.

²⁷ Note that this claim is limited to the marking of word stress only. Whether sentence stress will be strongly marked will additionally depend on the way the language expresses focus. If the location of the focus is fully determined by the syntax (or marked by lexical means), there is no need to also strongly mark it in the prosody. If, however, focus is independent of word order, then one should expect a phonetically prominent marking. See Rasier and Hilgsmann (2007) for more discussion on the division of work between syntax and prosody in the marking of focus.

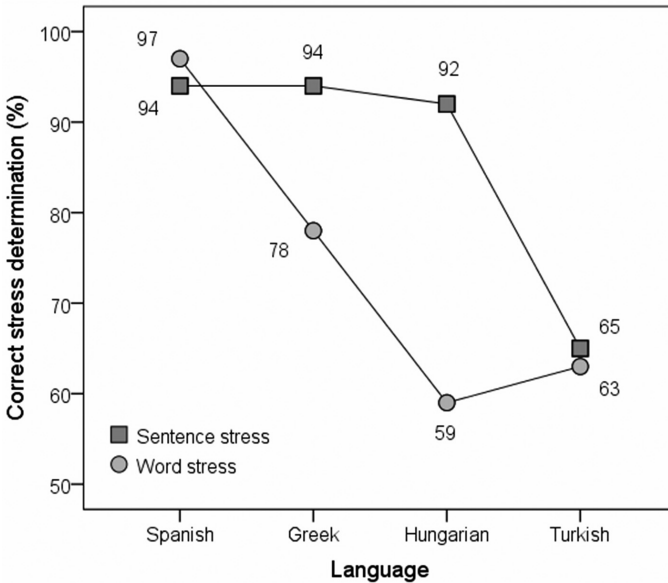


Figure 1.10 Automatic classification of syllables as stressed or unstressed (per cent correct) for two variable-stress languages (Spanish, Greek) and two fixed-stress languages (Hungarian, Turkish). Target words were either in focus (sentence stress) or out of focus (word stress). Based on Vogel, Athanasopoulou and Pincus (2016).

stress languages. Classification is much better for sentence stress – with the notable exception of Turkish (65% correct).

4.2 Functional Load Hypothesis

Berinstein (1979) was the first to formulate what would later come to be called the functional load hypothesis (FLH) of stress marking. The FLH predicts that stress parameters will drop to a lower rank in the hierarchy of stress cues when they are also employed in the segmental phonology of the language. For instance, if a language has a length contrast in the vowel system, vowel duration – which is normally a strong cue for stress – can no longer function effectively in the signalling of stress. Berinstein (1979) is often quoted in support of the FLH; for example, by Cutler (2005). It is not difficult to see, however, that Berinstein's claim is contradicted by her own results.

Berinstein presented sequences of four synthesized syllables /bɪ/ to native listeners of four different languages: English (36 listeners), Spanish (22),

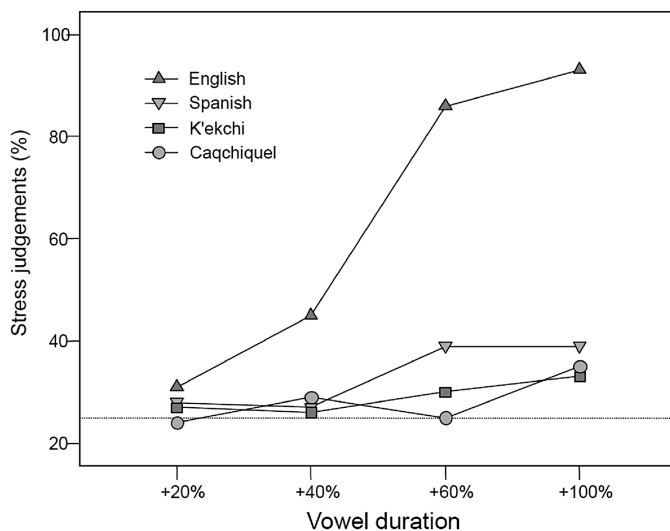


Figure 1.11 Stress perceived (%) by native listeners of four different languages on deviant syllable as a function of duration increment of deviant vowel in sequences /bɪbɪbɪ/ synthesized on a 100-Hz monotone (computed on data from Berinstein 1979).

K'ekchi (31) and Caqchiquel (46). The latter two are closely related Mayan Indian languages spoken in Guatemala.²⁸ English and Spanish have variable stress while K'ekchi and Cachiquel have fixed stress on the final syllable. Orthogonally to this, English and K'ekchi, unlike Spanish and Caqchiquel, have a length contrast in their vowel system. In the /bɪbɪbɪ/ sequences the standard vowel duration was 100 ms. The vowel duration in one syllable deviated from the standard and was adjusted to 70, 120, 140, 160 or 200 ms. This yielded a stimulus set of 1 (standard) + 4 (positions) × 5 (durations) = 21 types. Figure 1.11 shows the percentage of stresses perceived by the four groups of listeners on the deviant syllable as a function of the size of the duration increment.²⁹ The results show that English listeners are much more sensitive to the duration increment than the other three groups. Spanish listeners are rather insensitive, while the Guatemalan listeners were virtually insensitive: their responses are (marginally) above chance only for the largest duration increment.

²⁸ All Caqchiquel and most K'ekchi listeners were bilingual in their own language and Spanish.

²⁹ This summary figure was computed by me (VH) on the basis of tabular data in Berinstein (1979).

These results clearly run counter to the prediction of the FLH. The duration parameter is compromised in the vowel length contrast in English but not in Spanish and yet duration is a much more powerful stress cue in English. Similarly, given that K'ekchi has a vowel length contrast and Caqchiquel has none, the FLH predicts that duration should be a better stress cue in Caqchiquel, but there is no difference between these languages in the way they (fail to) use duration as a stress cue.

Native listeners are perfectly able to decompose segment duration into multiple sources of variation (e.g. Klatt 1976 for English, Nootboom 1981 for Dutch). Nootboom and Doodeman (1980), for instance, elegantly show that Dutch listeners adjust the category boundary between long and short vowels in the minimal pair /tak ~ ta:k/ 'branch' ~ 'task' with great precision depending on the depth of the prosodic break following the target word. By the same token, listeners appear to be able to decompose vowel duration into one part that is caused by a phonemic length opposition and another that is governed by stress. For this decomposition to work, the stress pattern on the target word should not be ambiguous, that is, the same segment string should not exist as a minimal stress pair in the language. So in a language with fixed stress, such as K'ekchi or Caqchiquel, there would be no problem. In English or Dutch minimal stress pairs, effortless decomposition of duration might be compromised.

Vogel, Athanasopoulou and Pincus (2016) compared the strength of correlates of word and sentence stress in Hungarian, Spanish, Greek and Turkish (see previous section). Of these languages only Hungarian has a phonemic length contrast in the vowel system. All four languages (Figure 1.12) show a slight elongation of vowels (whether stressed or unstressed) when the target word was in focus (i.e. produced with sentence stress). Duration did not vary as a function of word stress in Hungarian (as predicted by the FLH). However, duration did not vary as correlate of word stress in Spanish and Turkish either, which would run counter to the FLH.³⁰ Vowel duration turned out to be a strong correlate of word stress in Greek but even more so under sentence stress. If we add to this the knowledge that vowel duration is a highly reliable correlate of stress, both at the word and at the sentence level in English and Dutch (see Section 3.1), two languages that exploit vowel duration in a segmental contrast (tense versus lax vowels), we are left with little evidence supporting the

³⁰ Vogel, Athanasopoulou and Pincus's (2016) version of the FLH differs from the original formulation by Berinsein (1979). In the original formulation the default rank order of the stress cues would be $f_0 > \text{duration} > \text{intensity}$ (based on Fry 1955, 1958 and Bolinger 1958 for English, and generalized to all languages by Hyman 1977, see Berinsein 1979: 2). A higher cue that is implicated in a segmental contrast is predicted to assume a lower rank and is superseded by the other cues, which keep their relative position in the rank order. The newer version of the FLH makes no explicit prediction of the relative importance of the non-implicated stress cues.

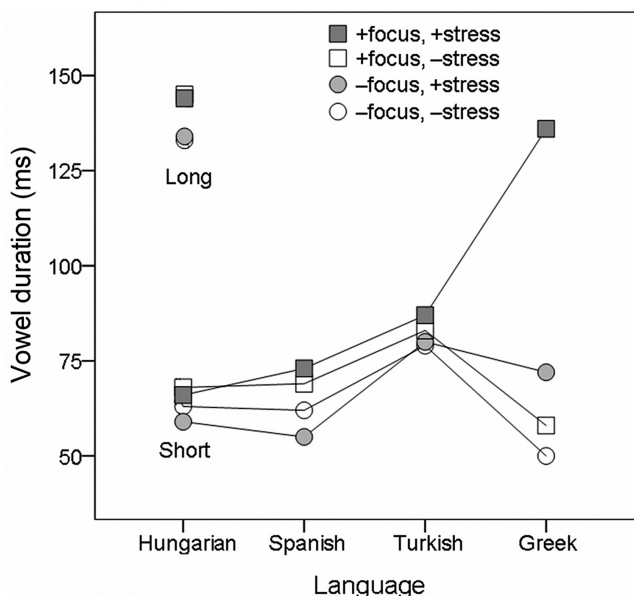


Figure 1.12 Vowel duration (ms) in four languages broken down by syllables with and without word stress and words with and without sentence stress (+focus and -focus, respectively). Of these only Hungarian has a short–long opposition in the vowel system. Data from Vogel, Athanasopoulou and Pincus (2016).

feasibility of the Functional Load Hypothesis. Note, finally, that Vogel et al. measured acoustical correlates; the issue of the perceptual strength of the various stress cues is not addressed directly in their chapter.

Differences in intensity and/or spectral tilt have never been advanced as primary cues of segmental contrasts. The only stress parameter that might also be exploited in segmental contrasts, other than duration, would be vowel quality, especially vowel bleaching, also known as spectral vowel reduction, that is, the tendency for unstressed vowels to assume a schwa-like quality. This parameter, however, has always been claimed to be the least effective stress parameter (see Section 3.4), so that it makes little sense to examine the position of this parameter in the rank order of stress cues: it will be the least important one, no matter if the language has spectral vowel reduction or not.

Although there seems little support for the FLH in the segmental domain, the situation may well be different when stress-related parameters are in competition with other prosodic contrasts. What, for instance, if a language has both

stress and lexical tone? In such cases it might be more difficult for the listener to disentangle the various cues for the competing contrasts.

Potisuk, Gandour and Harper (1996) investigated the acoustical correlates of stress in Thai, a language with five different lexical tones (high, mid, low, falling, rising) and with a length contrast in the vowels. The authors assumed that pitch would not be a high-ranking correlate of stress given that pitch is known to be the primary correlate of lexical tone. They also assumed that duration could not be an important stress cue since it was already implicated in the vowel length contrast. The results showed that the distinctive pitch pattern of the five tones remained basically the same between stressed and unstressed syllables, so that the lexical tone contrast would not be compromised much in unstressed syllables. However, in unstressed syllables, the pitch range was substantially reduced so that the tone distance between high, mid and low was smaller and the contour tones (rising, falling) would be flatter. This is very much the same mechanism that Mandarin Chinese relies on when signalling the difference between words in and out of focus (e.g. Shih 1988, Chen and Gussenhoven 2008). Automatic classification of syllables as stressed versus unstressed was largely unsuccessful when based on mean pitch and/or the standard deviation of the pitch contour on the syllable. Intensity was not significantly affected by stress for any of the five tones. Duration (of the rhyme portion of the syllable) proved the strongest stress correlate. An LDA with rhyme duration as the only predictor yielded 99 per cent correct stress decisions. These results, then, are in line with the idea that stress parameters can be used simultaneously in segmental and prosodic contrasts but not in simultaneous prosodic contrasts. It should be noted that stress in the study by Potisuk et al. (1996) is sentence stress only. The 25 target word pairs appeared in sentence-initial position and were contrasting strong-strong versus weak-strong syllable sequences in adjective-plus-noun phrases. Comparisons were made only of the strong versus weak initial syllables (paradigmatic comparison).

Remijsen (2002) investigated Samate Ma'ya, a language spoken in the Raja Ampat archipelago off the coast of Papua (Indonesia). The language has both lexical tone and stress but does not have a vowel length contrast. Acoustic correlates of stress were the pitch contour, intensity, loudness (i.e. intensity in selected frequency bands), vowel quality (expansion/reduction) and duration. Figure 1.13 presents the success of each of the four parameters in signalling the difference between a stressed versus an unstressed version of a syllable (left-hand panel) or between the high, rising and falling tone on the syllable. The results show a perfect inverse relationship between the parameters' positions in the rank order of word-prosodic cues.

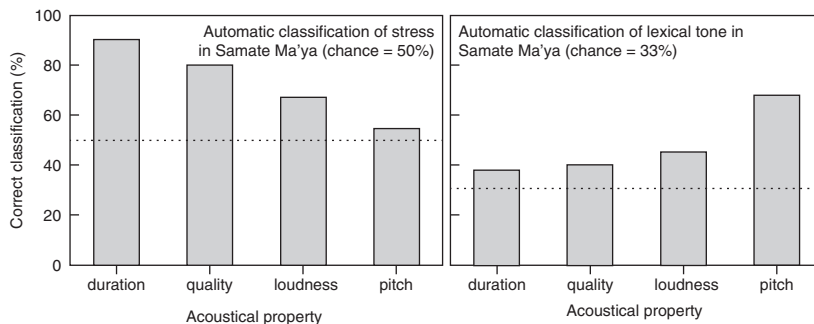


Figure 1.13 Per cent correct automatic classification of stress (left-hand panel) and lexical tone (right-hand panel) in Samate Ma'ya on the basis of the same four acoustic parameters (after Remijsen 2002: 64, 67).

So, the conclusion seems warranted that the Functional Load Hypothesis makes sense as long as parameters are in competition within the domain of prosody. There is no obvious competition when the same parameter is used separately in the segmental and prosodic domain. The original idea, as formulated by, for example, Berinsein (1979), Hayes (1995) and Posituk et al. (1996) was that the Functional Load Hypothesis should work when stress-related parameters were also involved in lexical contrasts. This, on second thoughts, seems to be too strong a generalization. What matters is that the FLH is valid only within the domain of prosody. Lexical distinctions that are non-prosodic in nature, such as segmental length contrasts, do not noticeably take part in the competition for resources. This provisional conclusion is based mainly on the results of acoustic studies, where the degree of success with which an automatic decision whether a syllable is stressed or unstressed is used to rank order the strength of the stress parameters. Only Berinsein tested the effects of duration directly in a perception experiment. In Section 3.6 we saw that the order of importance may deviate between acoustic correlates and perceptual cues. Before we take a definitive stance on the feasibility of the Functional Load Hypothesis, we need to compare cross-linguistic results in both the acoustic and the perceptual domain.

References

- Airas, M., H. Pulakka, T. Bäckström and P. Alku. (2005). A toolkit for voice inverse filtering and parametrisation. *Proceedings of Interspeech 2005*, 2145–8.
- Arvaniti, A., D. R. Ladd and I. Mennen. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25.

- Beckman, M. E. (1986). *Stress and Non-stress Accent*. Dordrecht: Foris.
- 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.
- Berinstein, A. E. (1979). A cross-linguistic study on the perception and production of stress, *UCLA Working Papers in Phonetics* 47: 1–59.
- Bloch, B. and G. L. Trager. (1942). *Outline of Linguistic Analysis*. Baltimore: Linguistic Society of America.
- Bolinger, D. L. (1958). A theory of pitch accent in English. *Word* 14: 109–49.
- Boxwell, H. and M. Boxwell. (1966). Weri phonemes. In S. A. Wurm (ed.) *Papers in New Guinea Linguistics, Volume 5*. Canberra: The Australian National University, 77–93.
- Browman, C. P. and L. Goldstein. (1992). Articulatory phonology: An overview. *Phonetica* 49: 155–80.
- Cambier-Langeveld, T. and A. Turk. (1999). A cross-linguistic study of accentual lengthening: Dutch versus English. *Journal of Phonetics* 27: 171–206.
- Cambier-Langeveld, T., M. Nespors and V. J. van Heuven. (1997). The domain of final lengthening in production and perception in Dutch. *Proceedings of Eurospeech '97, Rhodes*, 931–4.
- Campbell, W. N. (1995). Loudness, spectral tilt and perceived prominence in dialogues. *Proceedings of the Thirteenth Congress of Phonetic Sciences, Stockholm*, 676–9.
- Campbell, W. N. and M. E. Beckman. (1995). Stress, loudness, and spectral tilt. *Proceedings of the Acoustical Society of Japan, Spring Meeting*, 3–4–3.
- Carlson, R., K. Elenius, B. Granstrom and S. Hunnicutt (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. *Speech Transmission Laboratory-Quarterly Progress and Status Report* 26: 63–94.
- Caspers, J. and V. J. van Heuven. (1993). Effects of time pressure on the phonetic realisation of the Dutch accent lending pitch rise and fall. *Phonetica* 50: 161–71.
- Chen, Y. and C. Gussenhoven. (2008). Emphasis and tonal implementation in Mandarin Chinese. *Journal of Phonetics* 36: 724–46.
- Chomsky, N. A. and M. Halle. (1968). *The Sound Pattern of English*. New York: Harper and Row.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Mahwah NJ: Lawrence Erlbaum Associates.
- Comrie, B., M. S. Dryer, M. Haspelmath and D. Gil. (eds.) (2005). *World Atlas of Language Structures*. Oxford University Press.
- Cutler, A. (2005). Lexical stress. In D. B. Pisoni and R. E. Remez (eds.) *The Handbook of Speech Perception*. Oxford: Blackwell, 264–89.
- Di Cristo, A. and D. J. Hirst. (1986). Modelling French micromelody: Analysis and syntheses. *Phonetica* 43: 11–30.
- Dilley, L. C., D. R. Ladd and A. Schepman. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics* 33: 115–19.
- Dogil, G. and B. Williams. (1999). The phonetic manifestation of word stress in Lithuanian, Polish, German, and Spanish. In H. van der Hulst (ed.) *Word Prosodic Systems in the Languages of Europe*. Berlin: De Gruyter, 273–334.
- Eefting, W. Z. F. (1991). The effect of information value and accentuation on the duration of Dutch words, syllables and segments. *Journal of the Acoustical Society of America* 89: 412–14.

- Erickson, D. and S. Kawahara. (2016). Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistic Vanguard* 2: 1–16.
- Fant, G. and A. Kruckenberg. (1995). The voice source in prosody. *Proceedings of the Thirteenth Congress of Phonetic Sciences, Stockholm*, 622–5.
- Fry, D. B. (1955). Duration and Intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27: 765–8.
- (1958). Experiments in the perception of stress. *Language and Speech* 1: 126–52.
- (1965). The dependence of stress judgments on vowel formant structure. *Proceedings of the 5th International Congress of Phonetic Sciences*. Basel and New York: Karger, 306–11.
- Gauffin, J. and J. Sundberg. (1989). Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research* 32: 556–65.
- Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. In M. van Oostendorp, C. Ewen, E. Hume and K. Rice (eds.) *The Blackwell Companion to Phonology*. Malden, MA: Wiley-Blackwell, 924–48.
- Halle, M. and S. J. Keyser. (1971). *English Stress: Its Growth, and Its Role in Verse*. New York: Harper.
- Hargus, S. and V. Beavert. (2005). A note on the phonetic correlates of stress in Yakima Sahaptin. In D. J. Jinguji and S. Moran (eds.) *University of Washington Working Papers in Linguistics*, 24. Seattle: University of Washington, 64–95.
- Hart, J. 't, R. Collier and A. Cohen. (1990). *A Perceptual Study of Intonation*. Cambridge University Press.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago University Press.
- Hayward, K. (2000). *Experimental Phonetics*. Harlow: Pearson Education.
- Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31: 39–62.
- Heuven, V. J. van (1993). On the temporal domain of focal accent. In D. House and P. Touati (eds.) *Proceedings of an ESCA Workshop on Prosody*, Working papers 41. Department of Linguistics, Lund University, 132–5.
- (1998). Effects of focus distribution and accentuation on the temporal and melodic organisation of word groups in Dutch. In S. Barbiers, J. Rooryck and J. van de Weijer (eds.) *Small Words in the Big Picture. Squibs for Hans Bennis*. Leiden: Holland Institute of Generative Linguistics, 37–42.
- (2001). *Boven de klanken / Beyond the segments*. Amsterdam: Edita.
- (2014). Stress and segment duration in Dutch. In R. Kager, J. Grijzenhout and K. Sebregts (eds.) *Where the Principles Fail. A Festschrift for Wim Zonneveld on the Occasion of His 64th Birthday*. Utrecht Institute of Linguistics OTS, 217–28.
- Heuven, V. J. van and M. Ch. Dupuis. (1991). Perception of anticipatory VCV-coarticulation: Effects of vowel context and accent distribution. *Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence*, 78–81.
- Heuven, V. J. and P. J. van Hagman. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans and A. Hulk (eds.) *Linguistics in the Netherlands 1988*. Dordrecht: Foris, 59–68.
- Heuven V. J. van and M. de Jonge. (2011). Spectral and temporal reduction as stress cues in Dutch. *Phonetica* 68: 120–32.

- Heuven, V. J. van and L. Menert. (1996). Why stress position bias? *Journal of the Acoustical Society of America* 100: 2439–51.
- Heuven, V. J. van and A. M. C. Sluijter. (1996). Notes on the phonetics of word prosody. In R. Goedemans, H. van der Hulst and E. Visch (eds.) *Stress Patterns of the World, Part 1: Background*. The Hague: Holland Academic Graphics, 233–69.
- Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica* 35: 86–105.
- Hyman, L. (1977). On the nature of linguistic stress. In L. Hyman (ed.) *Studies in Stress and Accent*. Los Angeles: USC Linguistics Department, 37–82.
- Isačenko, A. V. and H. J. Schädlich. (1966). Untersuchungen über die deutsche Satzintonation. *Studia Grammatica* 7: 7–64.
- Johnson, K. (2003). *Acoustic and Auditory Phonetics*, 2nd edn. Malden, MA: Blackwell.
- Katwijk, A. van (1974). *Accentuation in Dutch: An Experimental Linguistic Study*. Amsterdam and Assen: Van Gorcum.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208–21.
- Klecka, W. R. 1980. *Discriminant Analysis*. Beverly Hills, CA and London: Sage.
- Ladd, R. (1996). *Intonational Phonology*. Cambridge University Press.
- Ladd, D. R., D. Faulkner, H. Faulkner and A. Schepman. (1999). Constant ‘segmental anchoring’ of F₀ movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–54.
- Ladd, D. R., I. Mennen and A. Schepman. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107: 2685–96.
- Ladefoged, P. (1967). Stress and respiratory activity. In P. Ladefoged (ed.) *Three Areas of Experimental Phonetics*. Oxford University Press, 1–49.
- Langeweg, S. J. (1988). The stress system of Dutch. Doctoral dissertation, Leiden University.
- Larkey, L. S. (1982). Reiterant speech: An acoustic and perceptual validation. *Journal of the Acoustical Society of America* 73: 1337–45.
- Lea, W. A. (1977). Acoustic correlates of stress and juncture. In L. Hyman (ed.) *Studies in Stress and Accent*. Los Angeles: USC Linguistics Department, 83–119.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, I. and G. E. Peterson. (1959). Vowel amplitude and phoneme stress in American English. *Journal of the Acoustical Society of America* 31: 428–35.
- Leyden, K. van and V. J. van Heuven. (2006). On the prosody of Orkney and Shetland dialects. *Phonetica* 63: 149–74.
- Lieberman, M. C. and L. A. Streeter. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America* 63: 231–3.
- Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773–81.
- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49: 606–8.
- Mol, H. and E. M. Uhlenbeck. (1956). The linguistic relevance of intensity in stress. *Lingua* 5: 205–13.
- (1957). The correlation between interpretation and production of speech sounds. *Lingua* 6: 333–53.

- Morton, J. and W. Jassem. (1965). Acoustic correlates of stress. *Language and Speech* 8: 148–58.
- Nakatani, L. H. and J. A. Shaffer. (1978). Hearing ‘words’ without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America* 63: 234–45.
- Nearey, T. M. (1978). Phonetic feature systems for vowels. PhD dissertation, Indiana University.
- Nooteboom, S. G. (1972). Production and perception of vowel duration: A study of durational properties of vowels in Dutch. Doctoral dissertation, Utrecht University.
- (1981). Speech rate and segmental perception or the role of words in phoneme identification. In T. Myers, J. Anderson and J. Laver (eds.) *The Cognitive Representation of Speech*. Amsterdam: North-Holland Publishing, 143–50.
- Nooteboom, S. G. and G. J. N. Doodeman. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America* 67: 276–87.
- Öhman, S. E. G. (1967). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151–68.
- Ortega-Llebaria, M. and P. Prieto. (2011). Acoustic correlates of stress in Central Catalan and Castillian Spanish. *Language and Speech* 54: 73–97.
- Ortega-Llebaria, M., M. M. Vanrell and P. Prieto. (2010). Catalan speakers’ perception of word stress in unaccented contexts. *Journal of the Acoustical Society of America* 127: 462–71.
- Pakerys, A. (1982). *Lietuviu Bendrines kalbos prozodija [The prosody of Standard Lithuanian]*. Vilnius: Mokslas.
- (1987). Relative importance of acoustic features for perception of Lithuanian stress. *Proceedings of the 11th International Congress of Phonetic Sciences*. Tallinn: Estonian Academy of Sciences, S.S.R., 319–20.
- Peterson, G. E. and I. Lehiste. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 693–703.
- Potisuk, S., J. Gandour and M. P. Harper. (1996). Acoustic correlates of stress in Thai. *Phonetica* 53: 200–20.
- Rasier, L. and P. Hiligsmann. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Cahiers de Linguistique Française* 28: 41–66.
- Remijsen, B. (2002). Lexically contrastive stress accent and lexical tone in Ma’ya. In C. Gussenhoven and N. Warner (eds.) *Laboratory Phonology VII*. Berlin and New York: Mouton de Gruyter, 585–614.
- Remijsen, B. and V. J. van Heuven. (2005). Stress, tone and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology* 22: 205–35.
- (eds.) (2006). Between stress and tone. *Phonology* 23: 121–333.
- 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.
- Saran, F. L. (1907). *Deutsche Verslehre*. München: C. H. Beck.
- Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83–109.
- Sluijter, A. M. C. (1995). *Phonetic Correlates of Stress and Accent*. The Hague: Holland Academic Graphics.

- Sluijter, A. M. C. and V. J. van Heuven. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organisation of syllables in Dutch. *Phonetica* 52: 71–89.
- (1996a). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100: 2471–85.
- (1996b). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSLP 96*. Philadelphia: Applied Science and Engineering Laboratories, Alfred I. duPont Institute, 630–3.
- Sluijter, A. M. C., V. J. van Heuven and J. J. A. Pacilly. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America* 101: 503–13.
- Sluijter, A. M., S. Shattuck-Hufnagel, K. N. Stevens and V. J. van Heuven. (1995). 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*, 630–3.
- Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Turk, A. E. and J. S. Sawush. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics* 25: 25–41.
- Vogel, I., A. Athanasopoulou and N. Pincus. (2016). Prominence, contrast and the Functional Load Hypothesis: An acoustic investigation. In R. Goedemans, J. Heinz and H. van der Hulst (eds.) *Accent and Stress*. Cambridge University Press, 123–67.
- Zanten, E. van and R. W. N. Goedemans. (2007). A functional typology of Austronesian and Papuan stress systems. In V. J. van Heuven and E. van Zanten (eds.) *Prosody in Indonesian Languages*. Utrecht: LOT, 63–88.