UNIVERSIDADE FEDERAL DE SANTA CATARINA

PÓS-GRADUAÇÃO EM LETRAS/INGLÊS E LITERATURA CORRESPONDENTE

ACOUSTIC CORRELATES OF WORD STRESS PRODUCTION IN THE CONNECTED SPEECH OF AMERICAN ENGLISH AND BRAZILIAN PORTUGUESE SPEAKERS

THAIS SADA RIBEIRO

Dissertação submetida à Universidade Federal de Santa Catarina em cumprimento parcial dos requisitos para a obtenção do grau de

MESTRE EM LETRAS

FLORIANÓPOLIS

Junho, 2006

Dedico este trabalho

a minha família

e a todos os amigos

que me acompanharam

nesta jornada

ACKNOWLEDGEMENTS

My warmest thanks to all of you who, one way or another, helped me accomplish this work : Ana Maria Cordeiro, Andréia Rauber, Adriana K. Dellagnelo, Anelise R. Corseuil, Ariane Girondi, Armando Silveiro, Barbara O. Baptista, Claudia B. de Fáveri, Cynthia Ann Freese, Daniel Moore, Dário F. Pagel, Denise Correa, Denise Klugue, Denize Nobre-Oliveira, Dóris Ribeiro Horne, Eliana Ávila, Eliete Marochi, Elizabeth A. R. da Silva, Fernanda Delatorre, Giles Istre, Izabel Seara, Jean Langdon, João Antônio de Moraes, Joice Lapa, José Luiz Meurer, Lilia Carioni, Luciany M. Silva, Mara Reis, Marcia Fernandes, Marco Rocha, Maria Aparecida Loss, Maria Carmen Sada Krane, Maria Clarisse S. Silva, Maria Cleonice S. Aita, Maria Climêne R. Pinto, Maria Clóris Sada Ribeiro, Maria do Socorro Baptista, Maria Lúcia B. Vasconcellos, Marilia A. Stein, Marta Zanatta, Melissa Bettoni-Techio, Nancy Muller, Oswaldo R. Pinto, Raul José Ribeiro Filho, Roque Estúdio, Rosana D. Koerich, Roziliane O. de Freitas, Sherry Blackwell, Sibele Corrêa, Susana Funck, Susana M. Fontes, Tanea Quintanilha, Vânia do Amaral e Viviane M. Heberle.

ABSTRACT

ACOUSTIC CORRELATES OF WORD STRESS PRODUCTION IN THE CONNECTED SPEECH OF AMERICAN ENGLISH AND BRASILIAN PORTUGUESE SPEAKERS

THAIS SADA RIBEIRO

UNIVERSIDADE FEDERAL DE SANTA CATARINA
2006

Supervising Professor: Barbara Oughton Baptista

This study set out to investigate the acoustic correlates of word stress in the connected speech of three Brazilian Portuguese (BP) and three American English (AE) speakers. After the subjects recorded an informative text in their native language, two bilinguals with formal training in phonetics pointed out the word stresses produced. Stressed vowels, as well as pretonic and posttonic vowels, were then measured as to their duration, F_0 peak and Intensity peak. Acoustic units were next transformed into perceptual units, so that statistically significant differences between stressed and pretonic or between stressed and posttonic vowels could be classified as perceptible or not. The effect of the co-occurrence of sentence stress upon the correlates of word stress was also examined. Due to the reduced number of subjects, significant and perceptible results were further classified as occasional, recurrent or consistent, according to their matching the results of one, two or three speakers in each group. While recurrent and consistent results were considered representative of group results, occasional results were considered suggestive of individual differences. By the criteria used, duration was the only significant and perceptible acoustic correlate of word stress for both groups of speakers. Word stressed vowels were longer than pretonic vowels for the AE group, and longer than posttonic vowels for the BP group. No significant differences in the duration of stressed, pretonic or posttonic vowels were found across both groups. Similarly, the co-occurrence of sentence stress did not result in any perceptibly significant change in the duration of word stressed, pretonic or posttonic vowels for these groups of speakers.

Number of pages: 107
Number of words: 26.621

RESUMO

CORRELATOS ACÙSTICOS DO ACENTO TÔNICO NA FALA CONECTADA PRODUZIDA POR FALANTES DO INGLÊS AMERICANO E DO PORTUGUÊS BRASILEIRO

THAIS SADA RIBEIRO

UNIVERSIDADE FEDERAL DE SANTA CATARINA
2006

Professora Orientadora: Barbara Oughton Baptista

Este estudo teve como objetivo investigar os correlatos acústicos do acento tônico na fala conectada de três falantes do Português Brasileiro (PB) e de três falantes do Inglês Americano (IA). Após os sujeitos gravarem um texto informativo redigido em sua língua nativa, dois falantes bilíngües com instrução formal em fonética delimitaram os acentos tônicos produzidos. As vogais tônicas, bem como as vogais pré-tônicas e pós-tônicas, foram então medidas quanto a sua duração, pico de F_0 e pico de intensidade. A seguir, unidades acústicas foram transformadas em unidades perceptuais, para que diferenças estatísticas significativas entre vogais tônicas e pré-tônicas, e entre vogais tônicas e pós-tônicas, pudessem ser classificadas como perceptíveis ou imperceptíveis. O efeito da co-ocorrência do acento frasal sobre os correlatos do acento tônico também foi examinado. Devido ao número reduzido de sujeitos, resultados significativos e perceptíveis foram posteriormente classificados como ocasionais, recorrentes ou consistentes, conforme sua coincidência com os resultados de um, dois ou três falantes por grupo. Enquanto resultados consistentes e recorrentes foram considerados representativos de cada grupo, resultados ocasionais foram considerados como indicativos de diferenças individuais. De acordo com os critérios adotados, a duração foi o único correlato acústico significativo e perceptível do acento tônico para os dois grupos de falantes. As vogais tônicas se mostraram mais longas que as pré-tônicas com relação ao grupo IA, e mais longas que as pós-tônicas com relação ao grupo PB. Nenhuma diferença significativa entre os dois grupos foi encontrada quanto à duração de suas vogais tônicas, pré-tônicas ou pós-tônicas. Da mesma forma, a co-ocorrência do acento frasal não resultou em mudança significativa e perceptível na duração das vogais tônicas, pré-tônicas ou pós-tônicas produzidas.

Número de páginas: 107
Número de palavras: 26.621

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Assuming that “o ato da fala não pode, em hipótese alguma, estar dissociado da situação de comunicação” (Istre, unpublished manuscript, p. 58), a speech event would minimally imply the production and the reception of a message. The production of this message would begin at the linguistic level “with the selection and ordering of suitable words and sentences” (Denes & Pinson, 1993, p. 4), it would then continue at the physiological level with muscular and peripheral neural activity, and end at the physical level with the propagation of sound waves. Its reception would naturally follow the opposite direction, starting at the physical, also called acoustic level, and from there moving on to the physiological and the linguistic levels, respectively.

Both the production and the reception of a message also share the common design of speech being “a purposeful human activity ... intended - under normal circumstances - to convey meaning” (Clark & Yallop, 1990, p. 1). As meanings derived from spoken words cannot help being “abstractions, reflections or translations of the actual physical energies” generated by a physiological source (Handel, 1993, p. 181), it seems reasonable that correspondences between the linguistic, the physical and the physiological levels should exist.

Thus, motivated by the notion of “a communication system in which ideas to be transmitted are represented by a code that undergoes transformations as speech events proceed from one level to another” (Denes & Pinson, 1993, p. 6), so that between the different forms of a message - that is, language units, articulatory movements, sound waves or nerve impulses - “systematic correspondences or correlations are maintained”

(Fry, 1979, p. 2), the present study sets out to investigate the acoustic correlates of word stress based on the analysis of three acoustic parameters: intensity peak, duration and fundamental frequency peak.

1.1. Objectives

The primary aim of this study is to describe and compare the word stresses produced by a small group of Brazilian Portuguese and American English native speakers so that the results and assumptions of previous studies can be discussed.

Its secondary aim, which was motivated by some controversy about the acoustic correlates shared by word and sentence stress, is to describe the acoustic correlates of sentence stress so that changes in word stress correlates due to the co-occurrence of sentence stress can be estimated for each group of speakers, and a decision can be made as to whether word stress studies in which sentence stress and word stress have co-occurred can be discussed in relation to this study.

1.2. Contents

Chapter II, containing the review of the literature, begins addressing sounds as physical and physiological events. Then, properties of sound propagation help describe the acoustic parameters used in stress measurements. After that, issues regarding the definition of stress are discussed, followed by the review of previous studies on stress production.

Chapter III, containing the method, begins addressing four research questions. Then, methodological issues related to stress conditioning factors and stress units of measurement are discussed. Next, the subjects and data collection are described, and the procedures for data measurements are reported.

In chapter IV, the criteria used for data analysis are reported and results are presented by acoustic parameter and group of speakers, in acoustic as well as in perceptual units. Significant and perceptible results are then compared and discussed in the light of the research questions, the literature reviewed, and the results of previous studies.

In chapter V, the findings are summed up and tentative suggestions for the teaching of American English pronunciation to Brazilian Portuguese native speakers are considered. Finally, the limitations of this study are discussed, and suggestions for future research are given.

CHAPTER II

REVIEW OF THE LITERATURE

The present review of the literature starts with a discussion of the issue of code transformation by means of describing sounds as physical and physiological events, after which properties of sound propagation will introduce the acoustic parameters employed in stress measurements. After that, moving to the linguistic level, some issues regarding conditioning factors and stress itself will be discussed. Finally, correspondences between the acoustic and the linguistic levels pointed out by previous studies on stress production will be reviewed.

2.1. Sounds as code transformation

Regarding their transmission, sounds are physical events. Whenever physical events take place, however, some supply of energy is necessary for "the work actually consists in converting this energy form into another" (Fry, 1979, p. 5). Accordingly, the air flowing out of the lungs "constitutes the force used in generating speech sounds", but its one-way flow of energy "has to be converted into to-and-fro movements or oscillations" in order that sounds come about (Fry, p. 62). This way, "the action of an air stream mechanism combined with the action of an interposed vibratile ... organ" can turn kinetic energy into acoustic energy (Denes & Pinson, 1993, p. 56).

The vocal cords, two small muscular folds inside the larynx, are the vibratile organ that produces most speech sounds. There is a gap between them, called the glottis, by which air passes freely while we are breathing. The moment we start speaking, however, the position of the folds changes in order to narrow this gap, so that the pressure built up by the reduction in the airflow makes them vibrate. As a result, air gets released as waves of discontinuous pulses or oscillations in air pressure.

These waves of pressure oscillations that flow from the glottis are then modified by articulatory configurations of the vocal tract which give them the shape they have when they come out of the speaker's mouth. From then on, the air pressure waves go on expanding into "the surrounding air, compressing neighboring particles of air", thus propagating acoustic energy until they end up reaching someone's ears (Denes & Pinson, 1993, p. 26).

Sound waves, or waves of oscillations in air pressure, make the eardrums move in and out and this way they mechanically transmit some of their feature patterns to adjacent hearing structures, which send them on to specialized receptors in the inner ear. There, those feature patterns are encoded into nerve impulses, which end up feeding brain wave patterns simultaneously produced for their linguistic processing.

Summing up, specialized sensory receptors inside the ear are mechanically set in motion by the oscillatory design that has first emerged from the speaker's mouth and then traveled along air pressure waves. This makes them produce electrochemical impulses that carry an equivalent design to the linguistic processing areas of the nervous system. From then on, the precise way the linguistic information will be interpreted is not so easily predictable, for it may rely on associations with past and present experiences and/or expectations as well.

Back to the acoustic medium, any such pattern of pressure variation under propagation can also displace the magnetic sensors of a microphone, and thus be transmitted as variations in electrical voltage that can be converted into graphic displays for speech analysis (Ladefoged, 1996, p. 11). In other words, when Ladefoged says that "a graph of a sound wave is very similar to a graph of the movements of the eardrum" (1982, p. 166), he is telling us that both of them are representations of the same mechanical force, and that this force has distinctive features.

2.2. Sound wave features and acoustic propagation

Air consists of molecules that move around at random while undisturbed, building up "a static pressure that is uniform in all directions" (Handel, 1993, p. 25). When sounds are transmitted, changes are imposed on this uniform pressure so that molecules of air are repeatedly pushed close together and far apart (Handel, p. 25). In this way, high pressure and low pressure waves can be alternately propagated outwards, up to the eventual damping of the sound transmission.

In fact, during the propagation of air pressure waves, the displaced molecules will shortly return to their initial positions, for it is actually a pattern of compression and rarefaction that is propagated, not the molecules themselves (Handel, 1993, p. 75). In other words, sound transmission depends both on the distance an air molecule moves to complete its vibratory cycle and on the distance that vibratory pattern travels along adjacent molecules of air.

2.2.1. Amplitude as a sound feature

The amplitude of a sound is the "maximum distance from equilibrium" that an air molecule reaches during its cycle of vibration (Handel, 1993, p. 10). That is, it refers to the size of each molecule's vibration as well as to the amount of air pressure displaced by that vibration. Amplitude also relates to wave propagation in that the distance reached in the displacement of each particle proportionally determines how far pressure waves can go - "the farther a sound body moves, the greater the displacement and compacting of air molecules", for instance (Handel, 1993, p. 27).

At the physiological level, sound amplitude is mostly the result of an increase in subglottal pressure caused by the action of the breathing muscles (Ladefoged, 1996, p. 66). But it can also be affected by "articulatory contributions such as the muscular tension of the vocal-tract walls and the amount of damping imposed by nasality" (Laver, 1994, p. 501).

The intensity or power of a sound differs from its amplitude in that both positive and negative variations in atmospheric pressure are mathematically transformed into positive values only, so that an average variation in air pressure can be provided. (Clark & Yallop, 1990). Different authors suggest alternative procedures for calculating the intensity of sounds[1] (Denes & Pinson, 1993 and Laver 1994, e.g.), but all of them agree that the result is proportional to their amplitude.

Another consensus is that intensity should be measured in decibels (dB), which express "a relation between two quantities" (Fry, 1979, p. 91), for "quando se fala de intensidade sonora, trata-se da intensidade relativa entre dois sons" (Istre, unpublished manuscript, p. 80). For the present purposes, *intensity* will refer to the propagation of vibratory patterns along sound waves, whereas *amplitude* will refer to the vibratory cycle of air particles.

2.2.2. Frequency as a sound feature

If we now turn from the amount of pressure increase to the rate at which these pressure increases occur, we'll be moving from the description of sound amplitude to the description of sound frequency. Accordingly, the frequency of a sound can be defined as "the number of complete repetitions (cycles) of variation in air pressure"

[1] One of the alternatives is to square amplitude values so that both positive and negative variations in atmospheric pressure are transformed into positive values only. After that, these positive values are averaged and square rooted.

(Ladefoged, 1982, p. 168) in a given unit of time[2]. As a consequence, what makes a sound frequency higher or lower is how long its cycles of vibration take, since shorter cycles yield more repetitions per time unit, and therefore produce higher frequencies, than longer cycles do.

At the physiological level, on the other hand, sound frequency corresponds to faster or slower sequences of pressure blows hitting the eardrums, that is, to a greater or lesser number of blows per unit of time. Sound amplitude, in its turn, makes these blows lighter or heavier, resulting in greater or smaller impacts on the eardrums.

Still in comparative terms, the amplitude and the frequency of sounds are two independent features in that they can vary separately (Ladefoged, 1996, p. 18). In order to build up a larger variation in air pressure without changing the frequency of vibration, for instance, air molecules will move faster (Ladefoged, 1996, p. 18). That is, as increases in amplitude imply larger distances to go, rates of vibration can stay steady if the speed of vibration compensates for them. Now let's turn to the description of frequency and amplitude as features of speech sound waves.

2.2.3. Frequency as a sound wave feature

Almost all speech sounds are complex tones, which means that different frequencies of vibration are simultaneously propagated as complex sound waves. In case these complex waves are periodic, as they are for vowel sounds, they will consist of sets of pure tones - that is, single frequencies of vibration - called harmonics.

Harmonics, by definition, have cycles of vibration whose durations are integer multiples of each other. As a result, the lowest harmonic in a periodic wave - called the fundamental frequency - has the longest cycles of all harmonics and therefore sets the timing of vibration of the periodic wave as a whole.

[2] For frequencies measured in Hertz , it will correspond to one second.

In other words, as multiples of the fundamental frequency (F_0, henceforth), all higher harmonics will have shorter cycles that will fit evenly into the time taken by an F_0 cycle. This is how the "the regularly repeating character of the wave-motion" (Fry, 1979, p. 82) in periodic waves is maintained, and also why the F_0 is called the frequency of vibration of periodic complex waves.

Aperiodic waves, on the other hand, are nonrepetitive complex waves since they "can have components at all frequencies, rather than only at multiples of a fundamental frequency" (Denes & Pinson, 1993, p. 35). One basic distinction between periodic and aperiodic waves is, therefore, the random wave movement in one case, and the 'regular, patterned character' of the wave movement in the other (Ladefoged, 1996, p. 83).

Going back to periodic waves, harmonic frequencies are generated at the vocal folds so that their F_0 results from the interaction between subglottal pressure and the tension of the laryngeal muscles (Denes & Pinson, 1993, p. 176). When the muscular tension is increased, "the folds are stretched tightly" and respond more readily to increases in subglottal pressure, producing a higher F_0. When the folds are "held together only loosely", they take somewhat longer to respond to increases in subglottal pressure, and a lower F_0 is produced (Clark & Yallop, 1990, p. 102).

The F_0 and the harmonics are not the only frequencies that periodic waves propagate, however. As these waves pass on to the vocal tract, resonances are excited which modify their pattern of vibration, so that resonant or formant frequencies also come into play.

Resonances arise because of the matching of identical frequencies, which leads to an increase in their amount of vibration. That is to say, when identical rates of vibration interact, the pushes they give each other arrive just in time to add to their amplitude of vibration. By the same token, potential rates of vibration of vocal tract configurations interact with the rates of vibration of waves generated at the vocal cords, and some harmonics end up with more intensity than others.

Resonant or formant frequencies are, therefore, those frequencies that stand out in relation to other frequencies in the periodic wave because of the extra intensity they gain when passing by the vocal tract. As all sounds have to pass through the vocal tract before they make their way into the atmosphere, all of them excite resonances - which means that all speech sounds have formants, but their actual localization in many sounds is difficult, except for speech sounds that are periodic.

In spite of the natural correspondences that exist between a wave generated at the vocal cords and its resonant frequencies, the F_0 and the formant frequencies can be varied separately and therefore be considered two independent features. While formants help distinguish the auditory quality of different speech sounds, the F_0 is considered the “responsável direto pelas características melódicas da fala” (Cagliari, 1981, p. 118).

As formant frequencies have greater intensity in relation to other harmonics in the periodic wave, it should follow that the F_0 was nearly as intense as the formant frequencies in order to be distinctively perceived. That’s not the case, however. In fact, even if ”the fundamental is removed by some form of electronic processing” a frequency known as the “phantom fundamental ” can still be perceived, since the main cue for F_0 perception is actually the relative timing of the harmonic frequencies (Clark & Yallop, 1990, p. 210).

2.2.4. Intensity as a sound wave feature

We know by now that “sons ... complexos possuem ... freqüências múltiplas simultâneas, cada uma dotada de uma certa amplitude.” (Istre, unpublished manuscript,

p. 85). The overall intensity of a complex sound wave will result, therefore, “da combinação da intensidade de cada um de seus componentes” (Istre, pp. 85-86), so that whenever the overall intensity of a complex wave is increased, the intensity of all wave components is “increased in the same proportion” (Ladefoged, 1996, p. 40).

The other way around is not necessarily true, however. In case the intensity of one or two of the lowest formants in a periodic wave is weakened, for instance, it will be mostly the relation of strength among formants - and therefore the quality of the sound - that will noticeably change[3] (O'Connor, 1973, p. 102). Only if the intensity of all lowest formants is proportionally weakened will the overall intensity of a periodic wave perceptibly change.

In other words, overall intensity relates to the total acoustic power of periodic waves, whereas the distribution of intensity along the wave's component frequencies - mainly in regard to the first three formants - relates to their auditory quality. Of these two features of sound wave intensity, overall intensity is the acoustic parameter used for stress measurements in this study.

2.2.5. Duration as a sound wave feature

The last acoustic parameter analyzed in this study seems to rely mostly on linguistic features in order to be meaningfully described. In fact, only a couple of written lines were found on duration throughout the literature reviewed on acoustics: the first saying that it is the acoustic correlate of speech articulatory gestures, and the econd suggesting its closer connection with frequency rather than with amplitude - probably due to the former being defined as movement repetitions within a given time unit, whereas the latter, as amounts of displacement in space.

[3] Since the lowest formants allow for the identification of periodic speech sounds

So now moving on to the linguistic level, duration, intensity and F_0 will be briefly described as linguistic features. After that, constraints and definitions regarding stress in connected speech will follow.

2.3. Duration, intensity and F_0 as linguistic features

2.3.1. Duration as a linguistic feature

Duration has several functions as a linguistic feature. It helps identify distinct phonemes - particularly vowels - and also their phonetic contexts, such as the voicing of postvocalic consonants, for instance. Duration can signal the number and the position of syllables in a word and, to some extent, the segmental structure of syllables as well[4]. In addition to that, duration signals overall sentence structure, prepausal contexts and suprasegmental features such as stress and accent. Its relative values vary according to speech style, reading materials and speech rate as well.

Most of the linguistic functions just mentioned are phonetic conditioning factors that might interfere in the analysis of stress in sentence context. As such, they will be addressed in detail when methodological issues are discussed. The specific role duration plays as an acoustic correlate of stress production, on the other hand, will be resumed later in this chapter.

2.3.2. Intensity as a linguistic feature

Intensity has three main linguistic functions: it helps identify distinct phonemes, vowels in particular, it signals overall sentence structure such as the beginning and end of sentences, and it is one of the correlates of stress. The role of intensity in the production of stress will be resumed later in this chapter, whereas its phonetic

[4] Particularly whether they are closed or open syllables.

conditioning effects at segmental and sentence level - particularly those that might interfere in the analysis of stress - will be discussed in section 3.2.2.3.

2.3.3. F_0 as a linguistic feature

Like intensity, F_0 has three main linguistic functions: it helps identify distinct voiced segments, vowels in particular, it signals overall sentence structure such as the beginning and end of sentences, and it is one of the acoustic correlates of suprasegmentals such as stress, accent and intonation. F_0 phonetic conditioning factors that might interfere in the analysis of stress will also be addressed in section 3.2.2.2. Likewise, the role F_0 plays in the production of stress will be resumed later in this chapter.

2.4. Suprasegmental features

Segments - that is, consonants and vowels - are defined by the presence of particular features that can be considered 'inherent' to their identification in relation to other segments in a given language (Laver, 1994, p. 452). Suprasegmentals, on the other hand, comprehend "all factors which can potentially be prolonged beyond the domain of the segment" (Laver, p. 152), thus spreading across syllables, words or sentences[5].

As a consequence, while segmental features "can be established either by inspection or ... comparison of an item with other items", suprasegmental features "are established by a comparison of items in sequence" (p. 2) and "must be described in

[5] Suprasegmentals have been variously described as a secondary function of the inherent features that identify segments (Lehiste, 1970), as overlaid or superimposed on segments (Laver, 1994 and Ladefoged, 1982, respectively), or else as spread across segments (Goldsmith, 1976).

relation to other items in the same utterance", for it is their relative values, not their absolute values, that matter (Ladefoged, 1982, p. 15).

Suprasegmentals also "tend, much more than consonants and vowels, to be directly related to higher levels of linguistic organization, such as the structuring of information" (Clark & Yallop, 1990, p. 278). Clark and Yallop point out, however, that "the segmental and suprasegmental dimensions of the speech signal do not function independently of each other", to the extent that "prosodic features can be just as discrete as consonants and vowels; and ... consonants and vowels are not always identifiable outside the context of speech in which they appear" (p. 282).

The main suprasegmental features are intonation, quantity and stress[6] (Ladefoged, 1982, p. 219; Lehiste, 1970, p. 1), which respectively correspond to F_0 "providing speech with recognizable melodical properties", to segments or syllables being shortened or lengthened, and to words made more prominent than others (Nooteboom, 1997, pp. 640-641).

Considering that "human perceptual processing of speech draws heavily on human short term memory", for listeners cannot retrieve the physical stimuli they are processing, "the less specified segmental structure is" - or might sound, in the case of foreign language learners - "the more support a listener needs from suprasegmental, prosodic cues" (Nooteboom, 1997, p. 668). In other words, suprasegmentals might prove useful in providing cues for the teaching of foreign languages as well.

2.5. Stress

It seems consensual that stress is "a conventional label for the overall prominence

[6] Which can be subdivided into word stress and sentence stress or accent. Other suprasegmental features are rhythm, speaking rate, voice quality and juncture (Laver, 1994, p. 450 ; Crutteden, 1986, p. 177).

of certain syllables relative to others within a linguistic system" (Clark & Yallop, 1990, p. 287) and that there must be a change in physiological and acoustic parameters so that these prominences occur. Stress is also considered a gradient phenomenon in that "a syllable is not stressed or unstressed in absolute terms, but is more stressed or less stressed than a neighboring syllable or some other point of reference" (Clark & Yallop, 1990, p. 94).

At the same time that "os modelos fonológicos mais recentes (não-lineares) têm definido 'acento' como uma relação de proeminência entre sílabas" (Massini-Cagliari, 1992, p. 9), texts on phonetics have associated the term 'prominence' to stress perception, rather than to stress production. Statements such as "from the perceptual point of view, all stressed syllables have one characteristic in common, and that is called prominence" (Roach, 1983, p. 73) and "syllables vary in their perceptual prominence in the chain of continuous speech" (Laver, 1994, p. 450) are frequent in the literature reviewed. For this reason, stress will be defined in this study as the overall salience - rather than prominence - of some syllables relative to others within a linguistic system (Behne, 1989, p. 19).

In fact, despite the widespread use of this concept, Coulthard, Brazil and Johns (1989), as well as Clark and Yallop point out that it is "misleading to suggest that there are standard definitions" for stress (Clark & Yallop, 1990, p. 289), to which Goldsmith adds that "the definition of stress is one of perennially debated and unsolved problems of phonetics" (1976, p. 5). Thus, three issues that help make the definition of stress seem uncertain and contradictory will be discussed next.

2.5.1. Issues regarding the definition of stress

The first issue regards stress perception and stress production being related but not identical processes (Roach, 1983, p. 72). Back in 1970, Lehiste had already noticed that "the points of view of the speaker and the hearer have often been confused in defining stress" (p. 106). Still to these days, several authors make statements about stress and studies on stress are published without any mention of to which of the two processes they refer.

Stress production is also often related to "what the speaker does in producing stressed syllables" whereas stress perception is related to "what characteristics of sound make a syllable seem to a listener to be stressed" (Roach, 1983, p. 72). In other words, acoustic parameters are assumed to concern stress perception whereas physiological processes are assumed to concern stress production. Research on the acoustic correlates of stress production and perception, however, may allow that correspondences between the two processes are mapped out, since the acoustic data provides a common code for what is produced and perceived in speech.

The second issue regarding the definition of stress refers to stress corresponding either to word stress or to sentence stress, each of them with specific linguistic functions and acoustic features. Very often, however, only the general concept of stress is addressed by authors and researchers in the area[7].

The third issue interweaves with the first two, making it more complex. It has to do with what the acoustic correlates of stress are. Next, focusing on stress production rather than perception, the second and the third issues will be discussed in more detail.

2.5.2. Word stress

Word stress is associated with lengthening and strengthening processes such as syllable lengthening and diphthongization of the syllable nuclei - both in American

---

[7] Sometimes due to their considering stress equivalent to word stress and accent, to sentence stress, and at other times, for no apparent reason at all.

English and in Brazilian Portuguese (Major, 1985, pp. 264-265). The absence of word stress, on the other hand, relates to shortening and weakening processes such as deletion of unstressed syllables and vowel reduction in American English, and vowel raising, monophthongization and desyllabification in Brazilian Portuguese (Major, pp. 266-270).

As to stress placement, on the other hand, word stress is considered 'a defining property of a word' (Laver, 1994, p. 511). In fact, word stress placement in American English (AE, henceforth) and in Brazilian Portuguese (BP, henceforth) can "eventualmente permitir oporem-se palavras" - at the syntactic and/or the semantic level - "que são idênticas quanto a sua composição segmental", such as sábia (adj.) / sabia (v.) / sabiá (n.) in Portuguese and áddress (n.) / addréss (v.) in English[8]. When it comes to sentence level, however, word stress placement may be subject to change in both languages.

2.5.2.1. Word stress at sentence level

The most frequent change in word stress placement at sentence level is "the dropping of some of the stresses" (Ladefoged, 1982, p. 109). Although "every word has at least one stress in its citation form", words that play certain syntactic functions - like articles, shorter prepositions, conjunctions and personal pronouns - are usually unstressed in connected speech[9] , both in AE and in BP (Cruttenden, 1986, p. 23; Fernandes, 1976, p. 44). Some authors even suggest renaming the grammatical word as a phonological word based on the fact that phrases such as 'the table' , 'leave it' or 'de

---

[8] The so called variable word stress, as opposed to fixed word stress in which word stresses are 'predominantly placed on a given syllabic location'; as in French (Laver, 1994, pp. 518- 519).

[9] Whereas most nouns, adjectives, adverbs and verbs are usually stressed.

nada' will normally be spoken like single words , i.e., with only one salient syllable[10] (Clark & Yallop, 1990, p. 295, Carioni, 1978, p. 25).

Carioni (1978), too, reports that "both English and Brazilian Portuguese authors agree on [sic] that some words, in connected speech, seem to lose some or all of their stress force, while others retain their inherent stress" (pp. 25-26). In relation to BP, for instance, Major (1985) reports evidence from whole words becoming destressed at sentence level, resulting in "a syllabicity shift for the normally tonic syllable"[11] (p. 271).

Such changes in word stress patterns at sentence level are also influenced by rhythmic constraints. The literature reviewed discusses two main language rhythms: stress timing and syllable timing. In stress timed languages, such as AE, stressed syllables tend to recur at regular intervals of time, whereas in syllable-timed languages such as BP, syllables tend to recur at regular intervals of time.

In this respect, it is worth mentioning that empirical studies have failed to provide physical evidence for stress timing in English[12], whereas some evidence has been found for stress timing in so called syllable-timed languages such as BP. Major (1981), for instance, analyzed the carioca dialect and found that "the language becomes increasingly stress-timed as style becomes more casual" (p. 280). Massini-Cagliari (1992), as well as Cagliari and Abaurre (1986), also found evidence "que classificariam o português, estudado do ponto de vista físico, tanto como língua de ritmo acentual, como de ritmo silábico" (Massini-Cagliari, 1992, p. 67).

---

[10] Fernandes (1976) calls it 'unidade acentual' (p. 46) , while other authors call it stress group (Fudge 1984, p. 1) or foot (Halliday 1970, quoted in Clark & Yallop, 1990, p. 295).

[11] As in ['tʃiɐ 'maχsiɐ] may become [tʃiɐ 'maχsyɐ] , for example.

[12] Its psychological reality has been claimed since then

Barbosa (2000), however, reasons that both kinds of rhythm actually co-exist in each language, except that one predominates over the other (p. 374). Thus, until it is evidenced otherwise, BP is predominantly a syllable-timed language whereas AE is predominantly a stressed-timed language. Accordingly, changes in word stress placement at sentence level will be constrained by the regular recurrence of stresses in AE and by the regular recurrence of syllables in BP.

2.5.2.2. Word stress and Sentence stress

At first, word stresses in sentences were considered “in a very real sense an abstract quality”, one that only pointed to the “capacity of a syllable within a word to receive sentence stress”. This assumption related to the premise that “what is realized phonetically is sentence-level stress rather than word-level stress” (Lehiste, 1970, p. 150)

Back in 1964, however, Bolinger had already suggested that one should “... reserve accent for the syllable which actually is highlighted in a sentence - to show the importance of its word - and apply stress to the particular syllable in the word that gets the accent if the word is important enough to get one“ (pp. 21-22). He also pointed out that “pitch and lexical stress collaborated in defining such prominence, neither alone being sufficient to make a syllable sound accented” (cited in Laver, 1994, p. 493).

Since then, the prevailing premise is that word stress and sentence stress actually co-exist at sentence level. That is, sentence stresses coincide with word stresses - hence accounting for the so called potential of word stress’ to be accented - without preventing other, ‘non-accented’ word stresses from being realized in the same sentence. Sentence stress will now be described in more detail so that further distinctions between the two kinds of stress can be drawn.

2.5.3. Sentence stress

Sentence stress - or accent - on the other hand, relates to "relative prominence within longer utterances" (Clark & Yallop, 1990, p. 288), and according to most authors reviewed, it could as well belong to the intonation system since it also relates to the placement of tones[13]. Ladefoged (1982), for instance, argues that the term 'sentence stress' should be actually changed to 'tonic stress', for "the sentence is a unit of grammar, while the location of tonic stress is a matter which concerns the tone-unit, a unit of phonology" (p. 143).

Such relative prominence within longer utterances can have syntactic or semantic purposes, whether in AE or in BP. As a syntactic marker, sentence stress helps the listener "parse the continuous flow of speech into syntactic structures" by making the last stressed syllable of each phrase salient - the so called phrase-final stress (Behne, 1989, p. 4) - or else, by signaling the beginning of sentences, as pointed out by Bolinger (1964, p. 23).

As a semantic marker, on the other hand, sentence stress provides "focus to the semantic information carried by lexical items". In other words, it highlights semantic content in a sentence (Behne, 1989, p. 6). Focal stress - as it is also called - is therefore "determined by the communicative intentions of the speaker" (Sluijter & Heuven, 1996, p. 2471) and is "usually impossible to predict" (Ladefoged, 1982, p. 100). Now let's turn to the discussion of the last issue regarding the definition of stress, namely, what the acoustic correlates of stress - that is, of word stress and sentence stress - are, as far as stress production is concerned.

---

[13] A tone may be defined as "the choice of a pitch pattern or melody" (Clark & Yallop, 1990, p. 289), within which there is a syllable - the sentence stressed syllable - that "stands out because it carries the major pitch change" (Ladefoged, 1982, p. 99).

2.5.4. The acoustic correlates of Word stress and Sentence stress

The literature reviewed is consensual in saying that the acoustic correlates of stress are duration, overall intensity, F_0 and vowel spectral composition[14]. But stress can refer either to word stress or to sentence stress and, adding to that, word stresses at sentence level may coincide with sentence stresses. Thus, the acoustic correlates of word stress and sentence stress will be reviewed here by way of discussing three alternatives.

First, word stress and sentence stress have distinct correlates that co-occur whenever both stresses converge. Massini-Cagliari (1992), for example, says that the "acento frasal poderia ser definido como uma sílaba com os correlatos físicos do acento lexical (a saber, duração, intensidade e qualidade vocálica), que ocorre num contexto em que esteja havendo variações significativas no valor de F_0" (p. 30). Bolinger (1964), quoted some paragraphs above, apparently has a similar point of view.

Second, word and sentence stress have F_0 as a common correlate, although with distinct patterns and/or magnitudes. Laver (1994), for instance, claim that the difference between the F_0 in word and in sentence stress would correspond to "an intonationally significant dynamic change in pitch" in sentence stresses - 'intonationaly significant' meaning that "the change in pitch must be more prominent than that produced by word-stress alone" (p. 492), as "the tone will reinforce and exaggerate that used for the realization of lexical stress" (p. 514). Ladefoged (1982), too, points out that F_0 changes that occur in speech often involve "small pitch increases on each stressed syllable", which Ladefoged describes as minor changes in relation to those that make up tonic stress (or sentence stress) (p. 99).

[14] Secondary correlates of stress would include glottal stops before word initial vowels and aspirated plosives in word initial vowels. Syllabic sonorants, on the other hand, would cue unstressed syllables. (Lea, 1997, p. 113)

In regard to BP, Moraes (1986) reports that the F_0 is an acoustic correlate of both word and sentence stress (p. 23), whereas Fernandes (1976) found that distinct F_0 movements are used to signal word stress and sentence stress[15]. Fernandes points out, however, that as the "altura tonal está comprometida com a estrutura intonativa do enunciado, torna-se muitas vezes difícil julgar a sua participação efetiva como correlato físico do acento" (p. 60).

Third, both stresses have the same acoustic correlates, only distinct in their magnitude. Denes and Pinson (1993), for example, suggest that "certain syllables of a word (already carrying word stress) receive additional stress because of the function of that word in the sentence" (p. 175). These three alternatives will be resumed when the results of this study are discussed.

2.5.5. Most consistent correlates of Word stress

In fact, there is not a single acoustic parameter or set of acoustic parameters that invariably signals word stress. Back in 1970, Lehiste wrote that there was "no one-to-one correspondence between stress and any acoustic parameter" (p. 110) , which seems to hold true to this day. Many authors reviewed (such as Laver, 1994; Clark & Yallop, 1990; Lieberman & Blumstein, 1988; Adams & Munro, 1978; Fernandes, 1976) claim, however, that the co-occurrence of parameters is rather frequent, and that some parameters are more consistently associated with word stress than others. In addition to that, languages may differ in the relative importance of these parameters, and in their relative independence (Lehiste, 1970, p 146).

Major (1985), for instance, claims that the most consistent word stress correlate in BP is duration (p. 261). Fernandes' study on word stress production at sentence level

[15] See pp. 113-114 for a summary on affirmative sentences.

(1976) also found that duration is the "parâmetro acentual mais constante" in BP (p. 65), mainly in its co-occurring with other parameters (p. 59)[16]. According to Fernandes' results, the second most consistent parameter in BP is the F_0 while intensity is the least influential one (p. 65).

In fact, Fernandes' results showed that all percentages for co-ocurring correlates are higher than those for single word stress correlates[17] (1976, p. 57). Likewise, Lea (1997) reports that the co-variance of intensity and duration - the so called 'energy integral'- made it possible to spot 84% of all word stresses in samples of connected speech in AE (p. 95). Morton and Jassem (1965) point out, however, that the combination of more than one parameter might occur not only in complementary distribution, but possibly also in free variation (pp. 161-162).

As to the most consistent correlate in AE, Lieberman's study (1960) on English word stress production found that the F_0 is "most strongly correlated with stress", whereas "intensity was found to be more closely associated with stress than vowel duration" (Behne, 1989, p. 22). Ladefoged, on the other hand, suggests that duration is the most consistent parameter in word stress production[18] , since "the most reliable thing for a listener to detect is that a stressed syllable frequently has a longer vowel" (1982, p. 14). Accordingly, the present study will analyze the production of word stress in relation to vowel duration, F_0 peak, and intensity peak, so that co-occurring as well as most consistent acoustic correlates for the AE and the BP groups can be pointed out as well as compared across both groups.

---

[16] Duration was a single correlate in 10% of her tokens, whereas in 64.5% it co-occurred with one or two more parameters.

[17] 16% for intensity and duration; 15% for the F_0 and duration; 33.5% for the F_0, duration and intensity, in contrast to 4,5% for intensity; 8.5% for the F_0; and 10% for duration

[18] As Ladefoged considers sentence stress to be a feature of the intonation system, it is assumed here that stress in his text refers to word stress.

Shattuck-Hufnagel S., Dilley L. and Ostendorf M. (1996) argue, however, that earlier studies on the acoustic correlates of stress “used stimulus configurations that placed the major phrasal prominence of the utterance on the main-lexical-stress syllable that they wished to study” so that the correlates of both sentence and word stresses were ‘confounded’ (p. 383). They also claim, after Beckman and Edwards (1994), that procedures with similar consequences have been carried out in many other studies since then. As a result, “it comes as no surprise when such different corpora yield conflicting results" (p. 383).

Carrier sentences, for instance, allow that environmental conditioning effects are best controlled for and segmentation procedures are simplified. For this reason, they have often been used in word stress studies such as Major's (1985) and Massini-Cagliari's (1992), which will be next reviewed. As target words within carrier sentences stand for new pieces of information among extensive repetitions of carrier words, however, they tend to be sentence stressed as well as word stressed.

Thus, in order that the results of this study can be discussed in relation to some of the studies that will be next reviewed, the acoustic correlates of word stress and sentence stress will also be compared by language group, so that the extent to which word stress correlates are modified by sentence stress can be estimated. Next, AE and BP studies on word stress production at sentence level will be reviewed in more detail.

2.6. Previous Studies on AE and BP Word stress Production

No studies comparing the acoustic correlates of word stress production in AE and BP were found. For this reason, the studies reviewed here will focus either on AE or on BP word stress production. Studies that did not provide enough cues as to whether word or sentence stress had been analyzed - such as Delattre's (1966) - will not be reviewed here, nor will those studies that did not analyze connected speech materials since Klatt (1975) found evidence for vowels being "much shorter in a connected discourse than in words spoken in isolation" which called into question "the practice of relying on published studies ... that are obtained from citation forms when one wishes to quantify sentence production performance" (p. 138). In all, six studies met the conditions mentioned above, four on BP and two on AE.

2.6.1. Studies on BP

Fernandes (1976)

This study on BP focused on intensity, F_0 and duration as acoustic correlates of word stress and intonation in BP connected utterances. Five subjects from São Paulo (4 women and 1 man), all of them with college degrees, recorded affirmative and interrogative sentences[19], in which stressed vowels were spotted according to acoustic changes arbitrarily set[20], above which acoustic parameters were considered significant for the production of word stress. Stressed vowels in different positions at word and sentence level were analyzed, as well as pretonic and posttonic vowels. The co-occurrence of stress correlates was also investigated.

According to Fernandes' results on word stress, the simultaneous occurrence of F_0, intensity and duration was most frequent (33.5%), whereas the occurrence of single parameters was least frequent in the production of word stress (4.5% for intensity; 8.5% for F_0 and 10% for duration). The co-occurrence of duration and F_0, on the other hand, accounted for 15% of the stressed tokens, while the co-occurrence of F_0 and intensity

---

[19] Interrogative sentences were meant for the analysis of intonation only.

[20] Based on previous studies such as Lehiste's (1970) and Gily and Gaya (1979). Acoustic changes within vowel sequences corresponded to 20% change in duration, 10% in intensity, and 3/4 of a tone in relation to F_0.

corresponded to 6% and that of intensity and duration, to 16%[21].

As to F_0, increases from pretonic to stressed vowels continued through posttonic syllables in non-final sentence positions, so that the highest F_0 measures often corresponded to posttonic vowels. In sentence final position, on the other hand, F_0 decreased from pretonic to stressed vowels, and it kept decreasing through posttonic vowels (pp. 60-61). Accordingly, F_0 movement was more significant in signaling word stress than F_0 height.

On the whole, Fernandes' results pointed to duration as the most regular correlate of BP word stress (74.5%), while intensity was its least regular correlate (59%) - either as a single or a co-occurring parameter (p. 65). F_0, in its turn, was shown to be a primary correlate of intonation[22] and only a secondary correlate of word stress (p. 63) Finally, BP intonation levels and contours were also described in Fernandes based on the analysis of *grupos fônicos*[23]. Due to length constraints, however, these results will not be reported here.

Major (198)

Major described not only the acoustic correlates of BP word stress production but also word stress degrees and their correlation to weakening and shortening processes in posttonic and pretonic positions. Syllables rather than vowels were analyzed and different speech styles[24] were also compared.

Three subjects from Bahia, Minas Gerais and Paraná recorded carrier

---

[21] None of the parameters were significantly present in stressed vowels in 6.5% of the tokens.

[22] Signaling tone group boundaries.

[23] Which probably correspond to tone groups.

[24] Formal, normal and casual speech styles.

sentences[25] whose target words had to be next replaced by logatomes - made-up words consisting of identical CV syllables (p. 260). Major’s analysis was based on his logatome data and his results pointed to F_0, intensity and duration as the correlates of BP word stress (p. 260) - duration being “the primary correlate of stress in BP” (p. 261) due to a considerable variance in F_0 and intensity whereas 'durational ratios remained fairly constant'. Only figures on duration were reported in this study, though.

Major also found that “the tonic syllable is the longest, the posttonic the shortest, and the pretonic intermediate“ in normal speech style (p. 267), which he related to BP making use of primary and secondary stress (pp. 261-281). Differences in BP phonotactic patterns in relation to pretonic, posttonic and tonic positions - as well as shortening processes related to those positions - were also reported as further phonological evidence for two degrees of stress in BP (p. 262). Finally, Major’s results might have been on sentence stress rather than word stress, since target words within carrier sentences stand for main content words and, as such, they tend to be sentence stressed as well.

Moraes (1986)

This study analyzed intensity, duration and F_0 as acoustic correlates of BP word and sentence stress production. Eight male and female subjects from Rio de Janeiro, all of them with college degrees, recorded 36 affirmative and interrogative sentences in which trisyllabic logatomes were inserted in different sentence positions. Stressed syllable position within logatomes as well as pretonic and posttonic syllables, were also taken into consideration. Due to length constraints, results on interrogative and one-word sentences will not be discussed here.

Sentence stresses in test sentences were arbitrarily identified based on previous perception studies, which showed that word stresses were usually perceived as sentence

[25] Carrier sentences were the ‘say ...... again’ type.

stressed at the end of *grupos prosódicos* (GPs, henceforth), while those inside GPs were not[26]. GPs, in their turn, were identified according to 'topic' and 'comment' information units, based on Haliday's concepts of given and new information, respectively (p. 28). Test sentences were then separated into 2 groups for analysis: those with nonsense words in ‘strong positions’ (i.e., at the end of GPs) and those with nonsense words in ‘weak positions’ (i.e., inside GPs). Thus, stressed syllables in strong positions corresponded to sentence stresses and stressed syllables in weak positions corresponded to word stresses. After that, sentence stresses were classified into two other groups: those in final sentence position and those in non-final sentence position.

In final sentence position, sentence stress showed an F_0 fall in relation to pretonic syllables which continued through posttonic syllables[27]. Moraes related this F_0 fall to a substantial fall in intensity which was also verified in posttonic syllables. As to duration, there was “um aumento significativo da duração da tônica em relação às átonas” in all syllable positions (p. 10). In non-final sentence position, on the other hand, sentence stress showed an F_0 increase in relation to pretonic syllables which went on through posttonic syllables. An intensity fall in posttonic syllables as well as greater duration in stressed than unstressed syllables were also verified in non-final sentence position[28] (p. 15).

Word stress, in its turn, was analyzed within topic GPs (given information) and comment GPs (new information). Word stress in topic GPs was signaled by F_0 and intensity peak falls in posttonic syllables. As to duration, stressed syllables were

---

[26] Except for proparoxytones and paroxytones, which are more likely to be perceived as sentence stressed within GPs than oxytones (p. 8).

[27] In oxytones, stress would be indicated by the absence of that fall.

[28] Only to a smaller degree than in final positions, probably due to prepausal lengthening.

significantly longer than unstressed syllables - except for oxytones[29]. Word stress correlates within comment GPs turned out to be the same as those within topic GPs[30].

Based on these results, Moraes suggested that the F_0 fall in posttonic syllables was a side effect of the intensity fall in those syllables, and did not signal word stress. It was also pointed out that intensity behaved in a similar way in word and sentence stress and that duration was less frequently associated to word stress than to sentence stress due to its lack of significance in oxytones (p. 20).

In Moraes' final remarks, on the other hand, it was argued that F_0 was not only the "correlato por excelência do acento frasal" but also one of the correlates of word stress, as in strong positions it "assinala a localização da sílaba tônica no âmbito da palavra" (pp. 22-23). It was also argued that duration did not signal sentence stress "por não haver aí incremento desse parâmetro" in relation to word stress (p. 25), and that the intensity fall in posttonic syllables was 'um índice obrigatório do acento', since it was shown to signal word and sentence stress.

Finally, Moraes pointed out that intensity and F_0 were word stress correlates in relation to posttonic syllables, while duration signaled word stress in relation to pretonic and posttonic syllables. He also pointed out that there was not "uma marca prosódica única, ou pelo menos mais importante, na exteriorização do acento" in BP, though distinctions could be drawn in relation to stronger and more regular parameters (p. 25). Accordingly, duration and intensity would be more regular stress correlates - mostly concerning word stress - whereas F_0 would be a stronger stress correlate - mostly concerning sentence stress (p. 27).

---

[29] Since stressed syllables in oxytones could not be compared with pretonic or posttonic syllables, they were compared with an average unstressed syllable.

[30] Except for a slight F_0 increase in stressed syllables followed by a mild fall in posttonic syllables in paroxytones (p. 19).

Massini-Clagiari (1992)

Massini-Cagliari set out to make a phonetic description of word stress and rhythm in BP. For this purpose, flexible carrier sentences were used in which the position of carrier clauses varied and two target words were presented at a time (p. 12). These carrier sentences were repeatedly recorded so that target words were replaced by logatomes[31] and faster speech rates were tested. Only one subject, from Campinas, São Paulo, took part in this study, which investigated the duration, intensity and vowel quality of BP stressed syllables[32].

Though not included among the acoustic parameters analyzed, some tendencies in relation to F_0 were also discussed. Changes in the direction of F_0 movement in stressed syllables, for instance, were reported to occur in most target words but not in carrier words - which might indicate that those target words were sentence stressed as well. In fact, Massini-Cagliari claimed that "acento frasal poderia ser definido como uma sílaba com os correlatos físicos do acento lexical ... que ocorre num contexto em que esteja havendo variações significativas no valor de F_0" (p. 30), which implies that sentence stress would not interfere with the acoustic correlates of word stress. According to the literature reviewed, however, there is some controversy about that.

Massini-Cagliari's study was designed to control for the number of syllables and stressed syllable position at word level, as well as for intrinsic segmental factors, stress position at sentence level, and speech rate (p. 15). According to the results, the acoustic correlates of word stress are greater duration, intensity decrease in posttonic syllables and vowel quality. As 90% of all stressed syllables were longer than unstressed

---

[31] Logatomes were used in the analysis of duration and intensity, whereas target words were used in the analysis of vowel quality. The analysis of rhythm apparently drew on both of them.

[32] Variations in F_0 movement were assumed to signal sentence stress only (p. 30).

syllables (p. 18), duration was claimed to be the main acoustic correlate of BP word stress.

Sharp intensity falls in posttonic syllables were also verified in 71% of the tokens[33], as well as a "damping conditioning effect" which caused stressed syllables in sentence-final position to be less intense than pretonic syllables (p. 21). Stresses in sentence-final position as well as syllables with higher intrinsic intensity consonants were later discarded, and intensity falls were reported to occur in 85% of the posttonic syllables (p. 22).

Intensity and duration were single stress correlates in 29% and 10% of the tokens, respectively, while they simultaneously signaled word stress in 61% of them. This was the highest percentage of co-occurrence for oxytones, paroxytones and proparoxytones, irrespective of sentence position[34] (p. 23). Based on that, Massini-Cagliari pointed out "a grande importância da interação entre esses dois parâmetros na atualização fonética do acento em português" (p. 23). The statistical significance of all percentages yielded was not reported, though, nor was the reference unit according to which duration was considered longer and intensity was considered higher in stressed than in unstressed syllables.

With respect to vowel quality, Massini-Cagliari reported that unstressed vowels in Brazilian Portuguese tend to be centralized and raised. A hierarchy in relation to raising and centralization processes, according to which "as pós-tônicas seriam as sílabas mais sujeitas a tais processos: depois viriam as pré-tônicas" (p. 28), was also referred to. After discussing current phonetic definitions for the syllable, Massini-Cagliari compared the duration of stressed syllables with that of stressed syllable nuclei and

---

[33] In middle and final sentence positions.

[34] Except for oxytones in sentence final position, due to prepausal lengthening.

concluded that "as evidências aqui apresentadas tendem muito mais para um modelo teórico que privilegie a sílaba para tratar da acentuação"[35] (p. 38).

Finally, phonological processes such as the compression and reduction of unstressed syllables on the one hand, and vowel deletion and diphthongation in posttonic syllables on the other, were also verified (p. 86). Accordingly, Massini-Cagliari's results in regard to BP rhythm pointed to "evidências que classificariam o português ... tanto como língua de ritmo acentual, como de ritmo silábico" (p. 67).

2.6.2. Studies on AE

Lieberman (1960)

This study investigated changes in duration, F_0 and envelope amplitude in word stressed syllables produced in sentence context by AE native speakers. Its objective was to verify the relevance of such changes for the mechanical recognition of stressed syllables.

Ten male and six female speakers, whose educational background varied from high school to graduate school, recorded 50 test sentences in which either the noun or the verb of noun-verb stress pairs appeared at the start, middle or end. The order of presentation of the sentences was randomized, and the subjects were asked to read every sentence in silence except for its test word, which they were to read aloud.

After that, two observers listened to the test words recorded and spotted their stressed syllables twice. Cases in which all four judgements did not coincide were discarded. The total number of tokens that showed higher F_0, greater intensity and longer duration in stressed than in unstressed syllables was then calculated in relation to unstressed syllables in test words as well as in corresponding stress pairs.

---

[35] See that discussion on pp. 39-40.

Lieberman's results showed that within the same utterance, 90% of the stressed syllables had higher F_0, 87% had greater amplitude and 66% had longer duration than unstressed syllables. When compared to unstressed syllables in corresponding stress pairs, 90% of the stressed syllables had greater amplitude, 72% had higher F_0 and 70% had longer duration. Within the same utterance, integrals of amplitude in relation to time were also shown to be greater in stressed than in unstressed syllables in 92% of cases. In addition to that, a "trading effect"[36] between amplitude and F_0 was reported to occur, whereas no cases of stressed syllables showing both lower amplitude and F_0 than unstressed syllables were verified (p. 453).

Based on these results, Lieberman claimed that F_0 and envelope amplitude were "the most relevant of the unidimensional acoustic correlates of stressed syllables", F_0 being the most relevant one (p. 453). Some of the test words in his study were main content words, however, and in 18 other instances they were sentence final - which means they would have taken phrase-final stress as well[37]. Finally, the statistical significance of the percentages yielded in his study was not considered, nor was any reference unit according to which the F_0, duration and intensity of stressed syllables would be considered higher, greater or longer than those of unstressed syllables.

Adams and Munro (1978)

Adams and Munro analyzed the acoustic correlates of stress production in the connected utterances of both native and non-native speakers of Australian English. They expected to find clues to why native speakers of syllable-timed languages are frequently unable to produce English rhythmic pulses. The duration, F_0 and amplitude of stressed

[36] His quote, on p. 453

[37] See Behne (1989).

and unstressed syllables were measured from texts[38] recorded by eight native and eight non-native speakers[39]. (p. 129).

The stressed syllables in the subjects' recordings were spotted by ten lay native English speakers. Only when there was agreement among 7 out of the 10 listeners, were syllables included in Adams and Munro's analysis. The kind of stress to be analyzed was not specified by the researchers, but considering their objective, it was assumed to be word stress. Due to length constraints, only the native speakers' results will be reported here.

In relation to amplitude, mean peak levels were quite the same for stressed and unstressed syllables: 32.6 and 32.8 dB, respectively (p. 147). The amount of fall from peak levels, on the other hand, was significantly greater in stressed than in unstressed syllables, and so were their amplitude end levels[40] (p. 141 and p. 148, respectively).

As to F_0, the researchers analyzed 5 different contours in terms of their amount of rise and/or fall, as well as their initial, peak and end levels. Fall and rise were the two most frequent contours in stressed syllables - 48% of fall and 25% of rise contours - whereas rise and fall were the most frequent contours in unstressed syllables - 44% of rise and 32% of fall contours (p. 139). Accordingly, the amount of fall in fall contours was shown to be significantly greater in stressed than in unstressed syllables. On the whole, rise-fall contours were more often used than fall-rise or level contours, and the

[38] The texts comprehended nursery rhymes, excerpts of verse, made-up equivalents and passages of colloquial and literary prose. They were claimed to 'feature reasonably basic structures and vocabulary' and be 'similar to the profile of natural spoken English' (p. 129).

[39] They were native speakers of various Asian languages who were graduate English teachers. They were described as 'proficient in written language as well as inexperienced in speaking English '(p. 129 e p. 153).

[40] Both were suggested to relate to the "disjuncture which occurs after stressed syllables", that is, to the longer 'intervals of relative silence' that follow stressed syllables (p. 148).

amount of rise and fall in rise-fall contours was significantly greater in stressed than in unstressed syllables as well.

Probably based on the use of rise contours in most unstressed syllables (44%) and fall contours in most stressed syllables (48%), Adams and Munro suggested that native speakers ‘frequently differentiated between unstressed and stressed syllables by changing from rise to fall contours’, especially at the end of utterances (p. 144). While the percentages for stressed syllables seem to intuitively support this assumption (48% fall; 25% rise), those for unstressed syllables do not seem so clearly distinct, though (44% rise; 32% fall). As for utterance final position, no figures were reported.

Finally, duration was analyzed in relation to F_0 and amplitude contours, so that the most significant differences between stressed and unstressed syllables were shown to occur in relation to F_0 fall contours. Concerning overall frequency of use, on the other hand, duration was the most frequent stress correlate for native speakers - either as a single correlate or not, and amplitude was used as often as F_0 (pp. 136-137)[41].

Summing up, stress was associated with 'greater duration, greater degree of F_0 change and greater fall in intensity peak', duration being the most predominant parameter (p. 141, p. 153). Tendencies towards the use of a single parameter, as well as individual differences accounting for variation in the parameters used, were also verified. Last, but not least, a more attentive reading of Adams and Munro's footnotes revealed that few test sentences had been reported and that sentence stress rather than word stress had been analyzed in their study. As a consequence, Adams and Munro's results will be discussed in relation to results in this study only in case sentence stress is not shown to interfere with the acoustic correlates of word stress.

[41] Amplitude and F_0. were only used as single cues to stress by one speaker (p. 153).

CHAPTER III

METHOD

3.1. Research questions

In relation to the main purpose of this study, which is to describe and compare the acoustic correlates of word stress produced by AE and BP native speakers in order that previous studies are discussed and suggestions for the teaching of AE to BP students might be given, two research questions are initially addressed:

1. What are the acoustic correlates of word stress for each group of speakers?
2. Are there perceptible and significant differences between coinciding acoustic correlates of word stress across both groups of speakers?

As to the secondary purpose of this study, which is to compare word stress and sentence stress acoustic correlates so that changes in word stress due to the co-occurrence of sentence stress are estimated, two more research questions are addressed:

3. Do word stress and sentence stress have coinciding acoustic correlates for the BP and/or for the AE group?
4. Are there perceptible and significant differences between coinciding correlates of word and sentence stress?

3.2. Methodological issues

3.2.1. Measurement unit: syllable vs. syllable nucleus

Both phoneticians and phonologists agree that “the smallest unit of speech which can be more stressed than another is the syllable” (Behne, 1989, p. 19)[1] . In spite of that, the concept of the syllable at the phonetic level apparently ‘enjoys no more general a consensus’ than that of the phonological syllable (Laver, 1994, p. 114).

In general terms, a syllable is described as ‘made up of a nuclear and marginal elements’ at the phonological level (Laver, p. 114). The 'nuclear element' - usually a vowel[2]- is the syllable minimal constituent and the ‘marginal elements’- those segments that may precede and/or follow the syllable nucleus - are necessarily consonants. Similarly, at the phonetic level syllables are generally described as “consisting of a centre which has little or no obstruction to airflow and which sounds comparatively loud” in relation to what might come after and/or before that center (Roach, 1983, p. 57).

There have been no phonetic approaches, however, “which can show any demonstrable, objective correlates on physically measurable parameters” for syllables (Laver, 1994, p. 113). Laver and other authors (see Roach, 1983, p. 57 ; Clark &Yallop, 1990, p. 97) argue that one of the main difficulties in this respect “lies in determining the possible boundaries of such a phonetic unit” (p. 113), that is, in the “desconhecimento, no sinal acústico, ... de pistas que levem a uma correta segmentação de unidades deste sinal que possam ser chamadas de sílabas” (Massini-Cagliari, 1992, p. 33). Istre (unpublished manuscript) also agrees that “alguns foneticistas instrumentais têm negado a realidade física da sílaba” because they cannot always determine “os limites da sílaba nos seus registros” (p. 53).

[1] See Massini-Cagliari (1992, p. 81) for an account of previous trends in phonology theory, which used to ascribe stresses to the syllable nuclei only.

[2] Nasal and lateral consonants may also be the nuclear elements in English, in case they are unstressed and word final - like in ‘eaten’ or ‘battle’. No examples in BP were reviewed in this respect.

At the phonological level, syllable boundaries are also an issue when (a) “medial consonants in words such as *falling* and *sugar* may be considered to belong to both to the preceding syllable and to the following“[3] (Clark & Yallop, 1990, p. 98), or (b) English speakers may disagree considerably with respect to the number of syllables in a sentence (Roach, 1983, p. 57). Definitions of the syllable based on the sonority and on the prominence of syllable elements[4] have also been attempted, though none of them ‘proved to be completely satisfactory’ (Ladefoged 1982, pp. 223-224).

Ladefoged sums up the peculiar status of syllables saying that “although nearly everybody can identify syllables, almost nobody can define them.” (1982, p. 220). The fact that most people can identify syllables, on the other hand, supports the assumption that “a aquisição e estruturação da língua na mente infantil é baseada na sílaba” (Câmara Júnior, 1977)[5], resulting in its awareness and acceptance (O'connor, 1973, p. 202). Whenever connected speech is produced, however, resyllabification rules may apply across word boundaries and even the intuition of native speakers will demand extra effort to correctly identify syllable boundaries.

Locating the boundaries of segments in an acoustic analysis of connected speech is in itself a considerable challenge, not to mention those cases in which surrounding contexts were not controlled for and all kinds of vowel-consonant transitions in two different languages were included, as in the present study. Mostly out of methodological feasibility, therefore, the syllable nucleus - rather than the syllable itself - was chosen for word stress measurements in this study. It is important to point out that this course of action does not challenge the fact that stress is a suprasegmental feature and, as such,

---

[3] Peaks of syllabicity would coincide with peaks of acoustic energy in the former and with a combination of sonority, length, stress and pitch in the latter.

[4] Which is addressed as ‘ambissilabicity’ in phonology theory . No information on BP was reviewed in this respect.

[5] Quoted in Souza, 1998, p. 13.

that it should relate to "those aspects of speech that involve more than single consonants or vowels" (Ladefoged 1982, p. 219). Regarding the acoustic parameters analyzed in this study, however, stressed syllable nuclei seem representative of stress production.

Syllable peaks of intensity, for instance, fall most of the time on vowels - probably due to their open articulation. Periodic waves in vowels also display, by definition, precise F_0 measures. Adding to that, vowels have repeatedly been reported to show significant changes in duration in relation to the production of stress (Ladefoged, 1982, Behne, 1989, Fernandes, 1976, Klatt, 1976). As to those instances where syllable peaks of intensity do not fall on syllable nuclei, where precise F_0 measures also comprehend, say, nearby sonorant consonants, or else, where unstressed vowel duration is similar to that of nearby stressed vowels[6], it is expected that they are not frequent enough to interfere with overall group results in this study.

There is empirical evidence reported in one of the studies reviewed on BP word stress, however, which argues for the fact that syllable duration provides unique, and therefore non proportional measures in relation to vowel duration. It is in Massini-Cagliari-Cagliari (1992), where the following hypotheses were tested: (a) vowel duration by itself signals stress placement, (b) vowel and consonant duration should be added in order to signal stress placement, (c) stressed syllables are shorter than or as long as unstressed syllables[7].

Massini-Cagliari's results showed that 82% of the tokens were assigned to (a), 8% to (b), and 10% to (c) . Next, she split the (a) category into (a1) where only vowel duration increased and (a2) where both vowel and consonant duration increased. Then

---

[6] Mostly in the case of unreduced vowels in unstressed positions, since unstressed vowels in pre-pausal positions will not be analyzed in this study.

[7] In Massini-Cagliari CV syllables were measured in logatomes within carrier sentences.

results showed that out of the initial 82%, 26% corresponded to (a1) and 56% to (a2). After that, Massini-Cagliari added the results in (a2) to the results in (b) and concluded that stress should be analyzed at the syllable level in BP[8].

Looking at these results from a different perspective, however, they not only give support to 'um modelo teórico que privilegie a sílaba' (p. 38) but also validate the premise that the syllable nucleus is representative of stress production in BP since, in all, vowel duration was shown to signal stress placement in 82% of the tokens[9].

3.2.2. Word stress conditioning factors

Lehiste (1970) says that "in a study of the linguistic function of suprasegmentals, it is ... necessary to start with the identification of all inherent constraints and conditioned variations" (p. 3) that might relate to them. Accordingly, conditioning factors for vowel duration, F_0 and intensity that can interfere with the analysis of the acoustic correlates of word stress will be next reviewed and discussed.

3.2.2.1. Vowel duration

Overall speech rate

According to Lea (1997), when speech rate in AE increases, the duration of unstressed syllables decreases more than that of stressed syllables. In other words, the relative ratios between stressed and unstressed syllables - and vowels as well - would be larger for faster speech rates (p. 114). Although no reports on BP were found in the literature reviewed, recordings with similar rates of speech across the two languages

---

[8] Massini-Cagliari also added to those results favorable to the syllable, the percentage corresponding to (c) on the pretext of being "melhor explicados considerando a sílaba como um todo" (p. 36). This way, her results favoring the syllable reached a total of 74% of her stressed tokens.

[9] As Massini-Cagliari might have analyzed sentence rather than word stress, this premise will be actually validated just after the acoustic correlates of sentence and word stress are compared in this study.

were selected for this study, so that differences in duration measurements from stressed to unstressed vowels due to different speech rates should not interfere with the results of this study.

Pre-pausal lengthening

It is consensual in the literature reviewed that "the syllable at the end of an utterance is longer than it would be within an utterance" (Klatt, 1975, p. 130) - a conditioning effect known as pre-pausal lengthening, also called phrase-final lengthening. An utterance, in its turn, is alternatively defined as limited by a phrase boundary, a following pause, or by the end of a breath group.

Klatt (1976) also claims that "most of the durational increment is restricted to the vowel, or else to any postvocalic sonorant or fricative consonants" (p. 1211), and Crystal and House (1988) add that in connected speech "a following pause lengthens the vowel in a word-final syllable seemingly by multiplicative factors which are independent of the effects of stress" (p. 1577). As to BP, the phrase-final lengthening of vowels in connected speech was also verified in Simões (1991) and in Major (1985).

Since utterances in this study will be defined as limited by breath groups, that is, by stretches of speech between two breath intakes, vowels at the end of breath groups will not be analyzed here. Pre-pausal vowels within breath groups, on the other hand, will be verified separately in order that lengthening effects in that position can be mapped out before a decision is made as to their inclusion in the analysis. Finally, pauses in this study will correspond to a minimum interruption of 200 ms in the speech flow (after Laver, 1994, p. 536 and Klatt, 1976, p. 1210).

Point and manner of articulation of postvocalic consonants

Vowel duration also depends on the articulatory characteristics of neighboring consonants[10]. With respect to their point of articulation, vowels are likely to be longer before alveolars or velars than before bilabials, whereas relative to their manner of articulation, vowels tend to be shorter before stops and longer before fricatives (Clark & Yallop, 1990, p. 72 and Klatt, 1975, p. 131).

Lehiste and Peterson (1960), for instance, found that in carrier sentences “the durations of all syllable nuclei in English are significantly affected by the nature of the consonants that follow the syllable nuclei”, whereas the influence of consonants preceding the syllable nuclei “appears to be negligible” (p. 200). In studies that analyzed other connected speech materials, however, this phonetic conditioning was not shown to be significant.

Klatt (1975) controlled the influence of postvocalic consonants according to syllable position at word level and found that “large durational differences conditioned by features of the following consonant were observed only in phrase-final syllables in the connected discourse” (p. 138). Umeda (1975), as well, found that such conditioning “holds true with most of our "prepausal" vowels, but the majority of the vowels in running speech do not fall into this situation”[11] (p. 435). Similar reports were not found in relation to BP, although it will be assumed that correspondences between AE and BP should exist with respect to this conditioning factor.

Voicing of postvocalic consonant

[10] Laver, 1994, p. 432; Clark & Yallop, 1990, p. 72; Lehiste, 1970, p. 24; Klatt, 1975, p. 131

[11] Except for unstressed vowels in word-final position (p. 442) . Umeda aimed at presenting “a large corpus of data for vowel duration in continuous text under as many conditions as possible“ (p. 434).

As reported in Laver (1994), Lehiste and Peterson (1960) found that AE vowels preceding voiceless consonants are shorter than those preceding voiced consonants by a ratio of 2 : 3. Lehiste and Peterson suggested that "the size of these contextually determined differences is language-specific" and that of other languages is probably different from "the rather extreme differences which typify English" (Laver, 1994, p. 446).

With respect to studies on AE using connected speech, however, this lengthening before voicing effect wasn't verified. Klatt (1976) claimed that "the influence of a postvocalic consonant on vowel duration is only large at phrase boundary" (p. 1220); Umeda (1975) found similar results (p. 435)[12]; and Crystal and House (1988), analyzing vowels in non-prepausal word-final position, found no effects of this phonetic conditioning either (p. 1561).

Concerning studies on BP, vowel lengthening before voiced consonants did not occur in Simões' connected speech data (1991), whereas de Faveri (1991) found that "o caráter surdo ou sonoro da consoante que segue a vogal tônica é um fator condicionador importante, exercendo influência sobre a duração da mesma"[13] (p. 35). Due to these conflicting findings, the influence of this conditioning effect will be considered when results for the BP group are discussed.

Intrinsic duration

Vowel articulation relates mostly to tongue height, frontness or backness, all of which influence intrinsic vowel duration[14]. Low back vowels, for instance, tend to be

---

[12] Except for vowels followed by voiceless stops.

[13] Also in respect to unstressed vowels (p. 56). De Faveri (1991) analyzed the production of BP vowels from Florianópolis in sentence context.

[14] English vowels are often classified by their tenseness as well.

intrinsically longer than high front vowels “because of the greater overall articulatory movement and biomechanical effort required” (Clark & Yallop, 1990, p. 72).

As to the comparative effect of stress and intrinsic vowel duration, Lea says that although stressed vowels tend to be longer than unstressed vowels in AE, the intrinsic identity of vowels also has ‘a significant effect’ on their duration, so that an unstressed / a / can be at times longer than a stressed / i / , for example (p. 99). As to BP, on the other hand, de Faveri’s results (1991) for oral vowels showed that "as durações inerentes às vogais átonas mostraram-se significativamente inferiores às durações inerentes às vogais tônicas“ (p. 67). Massini-Clagliari (1992) also reported that this conditioning effect does not seem to change the durational ratio between stressed and unstressed syllables and that "na grande maioria dos casos, a sílaba tônica é mais longa do que as átonas"[15] (p. 18). Thus, the influence of this conditioning effect will only be taken into consideration when results for the AE group are discussed.

Number of syllables in a word

Laver (1994) reports that “the number of syllables in a word is another factor which controls the relative duration of a given syllable and its individual segments”, each of them becoming “progressively shorter with the increasing number of syllables involved” (p. 448). Harris and Umeda (1974) point out, however, that most studies that found evidence for vowels being longer in monosyllabic words and progressively shorter in polysyllablic words have employed “nonsense words, isolated words, or short phrases, or words embedded in a carrier phrase“ as their reading materials (p. 1016).

After comparing the duration of stressed vowels in an ‘extended text reading’ and in carrier sentences, Harris and Umeda (1974) found that “the duration of the vowel in

[15] See the discussion on the proportional duration of stressed syllables and stressed vowels on p. ...

monosyllabic words in carrier phrases is significantly longer than for words having two, three, or four syllables" (p. 1017), whereas in other connected texts "the means, for the vowel in all non-prepausal situations, appear virtually identical" (p. 1018)[16]. Although no reports were found in relation to BP, correspondences between AE and BP will be assumed to exist in relation to this conditioning effect.

Syllable position at the word level

Klatt (1975) reports that some studies on English and Swedish claim that "segments in the medial syllables of polysyllabic words are shorter than the same segments would be in initial and especially in final syllables" (p. 130). Harris and Umeda (1974), however, did not find evidence for "this syllable-position durational effect in polysyllabic words of a connected discourse" (p. 1018).

In relation to BP, de Faveri (1991) found that syllable position at the word level tends to influence the duration of stressed vowels[17], whereas "não se constatou para as vogais átonas o mesmo que foi verificado para as vogais tônicas" (p. 71). Massini-Caglari (1992), on the other hand, reported that syllable position at the word level does not "parece alterar substancialmente a relação entre as medidas de duração das sílabas" (p. 18), whereas Moraes (1986) found that duration is a word stress correlate in paroxytones and proparoxytones but not in oxytones. Due to this relative lack of consensus, the influence of this conditioning effect will be considered when results for the BP and the AE group are discussed.

---

[16] Their study compared the compressibility of vowels due to the number of syllables in a word - both in connected texts and in carrier sentences.

[17] Increasing from word initial to word final position.

3.2.2.2. Vowel F_0

Intrinsic vowel F_0

In case segmental and suprasegmental environments are kept unchanged, there is a "connection between vowel quality and the relative height of the average fundamental frequency associated with it" according to which high vowels have higher F_0 whereas low vowels have lower F_0 (Lehiste, 1970, p. 68).

Clark and Yallop (1990) report that the difference between high and low vowels in English may be as great as 20-25 Hz in citation form, but they suggest that other phenomena associated to connected speech might override the effects produced by citation form alone (p. 284). Accordingly, Ladd and Silverman (1984) verified that the effect of intrinsic pitch was not as strong in connected speech as in citation form. As no information on BP was found, it will be assumed that correspondences between AE and BP should exist in this respect.

3.2.2.3. Vowel intensity

Intrinsic vowel intensity

Although overall intensity is primarily controlled by subglottal pressure, articulatory configurations of the vocal tract also influence the intensity of speech sounds. Given that suprasegmental and segmental environments are kept unchanged, open vowels - which are articulated with a more open vocal tract - will display higher overall intensity than closed vowels, for instance. All other things being equal, back vowels will be a little more intense than front vowels as well (Fry, 1979, p. 117). As intrinsic vowel intensity is mostly conditioned by physiological constraints, it is probably non-language specific and, therefore, should occur both in AE and BP.

Fry tells us that the difference between "the highest and the lowest intensity vowel sound on average is 7 dB" in relation to English (1979, p. 117)[18]. Although no reports were found regarding connected utterances, it is possible that the effects of intrinsic vowel intensities are overriden by phenomena associated to connected speech - as previously reported in relation to intrinsic vowel F_0 - or else, that stress itself may override these effects - as in the case of intrinsic duration. At any rate, the influence of this conditioning effect will be taken into consideration when the results of this study are discussed.

3.3. Subjects

Three female native speakers of AE and three female native speakers of BP, all of them with college degrees, participated in this study. The AE subjects - ages 34, 52 and 48[19] - had been living in Brazil for five to fourteen years at the time of data collection, although they continued speaking English rather frequently. The Brazilian subjects - ages 46, 51 and 48[20] - were university English teachers who used English regularly in their jobs. Although both groups of subjects were extensively exposed to a foreign language, no interference in their native accent was perceptible. The AE subjects were from Missouri, Colorado and California, whereas the BP subjects were from Santa Catarina, Rio Grande do Sul and Minas Gerais[20].

3.4. Data collection

All subjects were asked to read an informative text on the oral tradition of African peoples, written in their native language, as if telling its contents to someone else. The

[18] Probably based on citation form materials, which were the usual test materials in early analyses of that sort.

[19] AE and BP subjects 1, 2 and 3, respectively.

recordings took place in an insulated booth at a recording studio at the Federal University of Santa Catarina. After their first reading, subjects were asked whether they wanted to make an additional recording. As all of them did, the option of a third recording was given at the end of their second reading. In the end, all subjects read their texts at least three times. Of these readings, the second one was selected[20] and, of the six paragraphs in each text, the third one - where the subjects were already warmed-up in their readings - was used for the acoustic analysis.

Two bilingual speakers of BP and AE, one whose first language was AE and one whose first language was BP, both with formal training in phonetics, listened to the recordings in individual sessions in order to spot all prominences in them - irrespective of their degree of prominence or the perceptual parameter involved. Previous to that, they had been asked to spot prominences in shorter informative recordings in both languages, so that their rate of non-coinciding judgments could be estimated[21]. As to the recordings analyzed in this study, their non-coinciding judgments were discussed and solved in a follow-up joint listening session[22]. Later on, sentence stresses were spotted following the same procedure.

3.5. Data measurements

The F_0, intensity and duration of all vowels pointed out as prominent were measured in the CSL (Computerized Speech Lab) model 4300B, software version 5.X, as follows:

Each subject's recording was segmented into consecutive breath groups with overlapping areas of silence at their start and end, so that their continuity would be

---

[20] Except for one subject who asked for a fourth round and had this last recording chosen.

[21] It was fairly low: 10% to 13% for the recording in AE and 14% to 17% for the recording in BP

[22] They corresponded to 11.5% for the recording in AE and 15.5% for the recording in BP.

physically preserved to some extent. These overlapping areas of silence eventually proved to interfere with the analysis of the waveforms, though. Millisecond differences in their length were enough to change the parameters of waveform analysis, therefore resulting in distinct F_0 and intensity contours - both of them calculated from waveform data. As a consequence, each subject's breath groups had to be segmented again so that no stretch of silence was left either at their start or end.

Then, vowel duration was measured within 40 to 50 ms waveform frames. This frame size allowed wave periods to be distinctly displayed, so that changes in their shape and texture could be used for setting the start and end points of vowel-consonant (V-C) and vowel-vowel (V-V) transitions[23]. These short waveform stretches were alternatively double-gained once or twice, in order that more subtle changes in wave periods could be best displayed. The parallel inspection of spectrograms in two or three different gain adjustments, within two syllable-size frames, was also carried out. Movements and fade outs of the first three formants were indistinctly used as back up cues for transition boudaries in these cases.

An additional resource used for setting V-C and V-V transition boundaries was to double-length the waveforms two to four times, so that segments - or syllables - would be played at a much slower speed and more accurate listening judgments would result. For more difficult transitions such as V-V and sonorant-vowel transitions, all resources available were used. As a general procedure, however, only the inspection of waveform periods backed up by two spectrograms with different gains was employed. In relation to waveform periods, the criteria for vowel segmentation were the following:

a) For voiceless stop-vowel transitions as well as for voiceless and voiced fricative-vowel transitions, segmentation took place at the start of the first vowel period with

[23] See the appendix for samples of wave form segmentation

minimum intensity - which was arbitrarily set as corresponding to a four-dash high period within a 40ms frame. In relation to fricative-vowel transitions where friction had blurred the shape of vowel periods - particularly those including / ʃ / , / tʃ / , / s / , / dʒ/ and / ʒ / - a maximum of three vowel periods with 'friction' shape, irrespective of their intensity, was taken as reference for segmentation. For vowel-voiceless stop and vowel-voiced and voiceless fricative transitions, segmentation took place at the end of the last vowel period, following the same criteria.

b) For voiced stop-vowel transitions, segmentation took place either at the start of the first vowel-like period - double-gaining the waveform in order that subtle changes in the shape of the wave periods could be displayed - or, in case the consonant release was visible, at the start of the first period after the release. For vowel-voiced stop transitions, on the other hand, vowels were segmented either at the end of the last vowel-like period or, in cases where changes in the period shape were not clearly perceptible, at the point where vowels stopped being audible in relation to the following consonant[23]. Double-lengthening the waveform two or three times was the main resource used in such cases.

c) For nasal-vowel transitions, vowels were segmented either at the start of the first period after the consonant's release, or - whenever the consonant release was not visible - at the start of the first vowel-like period. For vowel-nasal transitions, on the other hand, segmentation took place either at the end of the last vowel-like period or - in cases where changes in the period shape were not distinctly perceptible - at the point where the vowel stopped being audible in relation to the following nasal[24]. Double-lengthening the waveform two or three times, and checking spectrograms with two or three different gains, were the resources used in such cases. Finally, when vowel nasalization periods preceded

---

[24] Also in relation to the following consonant and a subsequent vowel, at times. Some minimum stretch of another vowel was included so that the previous vowel could stand out in relation to it.

nasal consonant periods, vowels were segmented at the end of their nasalization periods. Nasal vowels in BP were segmented according to these criteria as well.

d) For glide-vowel , liquid-vowel and vowel-vowel transitions, the limits of the transition area were first set using all resources available. Then, segmentation took place at the start of the wave period that corresponded to the midpoint of the transition area. In relation to vowels preceded by liquids, on the other hand, either the vowel-like period criterion or the first period after the consonant release criterion were, at times, used instead. For vowel-glide and vowel-liquid transitions, the same criteria for segmentation was followed. In cases where vowels were rhoticized - due to an extensive co-articulation with a subsequent / ɹ / - segmentation took place at the end of the rhoticized section.

e) For flap-vowel transitions, segmentation took place at the start of the first vowel-like period. In case the consonant tap was clearly visible in the spectrogram, segmentation occurred at the first period after the release. For vowel-flap transitions, segmentation took place at the end of the last vowel-like period.

Once duration measures were set, intensity peak and F_0 peak measures were taken within each vowel duration. For this purpose, glottal stops and creaky periods - previously included in vowel duration[25]- were discarded. As to F_0 and intensity analysis range, frame length and frame advance, they were set for each subject according to specifications in the CSL manual. Finally, the last vowel in each breath group was not included in the analysis due to prepausal conditioning effects.

---

[25] Unless glottal stops were replacing consonants.