The Relative Weight of Statistical and Prosodic Cues in Speech Segmentation: A Matter of Language-(In)dependency and of Signal Quality

TÂNIA FERNANDES
PAULO VENTURA
RÉGINE KOLINSKY

Abstract

In an artificial language setting, we investigated the relative weight of statistical cues (transitional probabilities, TPs) in comparison to two prosodic cues, Intonational Phrases (IPs, a language-independent cue) and lexical stress (a language-dependent cue). The signal quality was also manipulated through white-noise superimposition.
Both IPs and TPs were highly resilient to physical degradation of the signal. An overall performance gain was found when these cues were congruent, but when they were incongruent IPs prevailed over TPs (Experiment 1). After ensuring that duration is treated by Portuguese listeners as a correlate of lexical stress (Experiment 2A), the role of lexical stress and TPs in segmentation was evaluated in Experiment 2B. Lexical stress effects only emerged with physically degraded signal, constraining the extraction of TP--words to the ones supported by both TPs and IPs.
Speech segmentation does not seem to be the product of one preponderant cue acting as a filter of the outputs of another, lower-weighted cue. Instead, it mainly depends on the listening conditions, and the weighting of the cues according to their role in a particular language.

Introduction

Speech is a continuous stream with few reliable cues to word-boundaries (e.g., Klatt, 1980; Liberman & Studdert-Kennedy, 1978). A vast bulk of research has demonstrated that many sources of information, both lexically driven (e.g., McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995) and signal derived (e.g., Cutler & Norris, 1988; McQueen, 1998; Saffran,

Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Vroomen, Tuomainen, & de Gelder, 1998), assist speech segmentation, with a compromise between them being established through linguistic development (e.g., Bortfeld, Morgan, Golinkoff, & Rathbum, 2005; Mattys, White & Melhorn, 2005; Mattys & Melhorn, 2007; Swingley, 2005).

The present study was aimed at evaluating the relative weighting of prosodic (suprasegmental[1]) and statistical (segmental) information in speech segmentation.

Listeners are sensitive to prosodic cues from the very beginning of language onset: newborns discriminate languages based on their rhythmic properties (Nazzi & Ramus, 2003; Ramus, 2002; Ramus, Hauser, Miller, Mones, & Mehler, 2000), two-month-olds are sensitive to Intonational Phrases[2] (henceforth, *IPs*; e.g., Dehaene-Lambertz & Houston, 1998), and 6-month-olds use pre-boundary length, pitch, and pause in clause segmentation (Seidl, 2007). After the age of 7.5-months, infants are able to use metrical prosody (Curtin, Mintz, & Christiansen, 2005; Jusczyk, Houston, & Newsome, 1999; Mattys & Jusczyk, 2001; Morgan & Saffran, 1995), lexical stress[3] (Houston, Santelman, & Jusczyk, 2004) and prosodic edges (Seidl & Johnson, 2006) to segment speech into words. Adult listeners are also sensitive to several types of prosodic cues (for a review, see Cutler, Dahan & van Donselaar, 1997), such as IPs edges (e.g., Shukla, Nespor, & Mehler, 2007), phonological phrases boundaries (e.g., Christophe, Gout, Peperkamp, & Morgan, 2003), metrical units (e.g., Cutler & Norris, 1988) and lexical stress (e.g., Mattys, 2000; see also electrophysiological evidence in Böcker, Bastiaansen, Vroomen, Brunia, & de Gelder, 1999;

[1] Recently Dilley and McAuley (2008) demonstrated that *distal* prosodic information (i.e., nonlocal, distant to the point where segmentation and lexical access occurs; e.g., the prosodic pattern of the first five syllables of a eight-syllable speech stream) modulates speech segmentation (e.g., of the last syllables of eight-syllable stream, as *bookworm*, or *book* and *worm*). Distal and proximal prosodic cues thus seem to have different roles in speech segmentation processing. The present study only regards the role of proximal prosodic information in speech segmentation, for the sake of clarity we will use the term prosody to refer to proximal prosody.

[2] A wide variety of terms are found in the literature, yet two levels above prosodic words are considered in the prosodic hierarchy: the Intonational Phrase (IP), which is the highest prosodic unit, mostly corresponding to whole clause or sentence and often marked by a pause at the end; and Phonological Phrases, below IPs, also referred as intermediate IPs or accentual groups (Grice, 2006; Werner & Keller, 1994). Both types of phrases are acoustically characterized by a final lengthening with a falling pitch contour (at their right edge; see Christophe, Pepperkamp, Pallier, Block, & Mehler, 2004; Shukla, Nespor, & Mehler, 2007).

[3] "Lexical" stress corresponds to word primary stress as a lexical property. In languages such as Portuguese, it also enables the distinction between two lexical items with the same phonological representation but that only differ in stress pattern (e.g., *pensão* /pẽˈsɐ̃w/, boarding house, and *pensam* /ˈpẽsɐ̃w /, they think). This prosodic information differs from *metrical prosody,* which is the rhythmic alternation between strong and weak syllables, the former including full vowels and the latter often reduced vowels in languages like English and Portuguese (see e.g., Cutler et al., 1997).

Cunillera; Toro, Sebastián-Gallés, & Rodríguez-Fornells, 2006; Friedrich, Kotz, Friederici, & Alter, 2004).

Statistical cues also assist speech segmentation processing since an early phase of language development. Many studies have focused on transitional probabilities (henceforth, *TPs*[4]) computation, more precisely on the ability to extract nonsense "words" based on high TPs between their adjacent syllables (henceforth, *TP-words*) from a nonsense continuous artificial language (*AL*) stream. The ability to track TPs appears to be precocious (e.g., Kirkham, Slemmer, & Johnson, 2002; Thiessen & Saffran, 2003), rapid (e.g., Saffran et al., 1996a), involuntary (e.g., Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; but see Toro, Sinett, & Soto-Faraco, 2005) and age-independent (e.g., Saffran et al., 1997). Since TPs computation does not require any, not even minimal, lexical knowledge, this could be a pivot mechanism in the acquisition of not only words but also other word-boundary cues (Thiessen & Saffran, 2003).

Recent studies have been devoted to the integrated study of statistical and suprasegmental cues in line with the proposal that the integration of multiple cues may provide evidence about linguistic aspects that cannot be derived from any single source, promoting the optimization of speech segmentation (see Christiansen, Allen, & Seidenberg, 1998; Christiansen & Curtin, 2005). These studies have used both *congruent cues* suggesting the same segmentation hypotheses and hence leading to the same parsing, and *incongruent cues* suggesting incompatible segmentation hypotheses.

With incongruent cues, in optimal listening conditions (i.e., with intact speech) adult listeners seem to underestimate prosodic information (e.g., strong syllables, syllable lengthening, lexical stress), favouring either high--level, lexical context (Mattys et al., 2005) or sublexical information such as phonotactic legality (McQueen, 1998), phonotactics (Mattys et al., 2005), and coarticulation (Mattys, 2004. Even artificial language learning (*ALL*) seems insensitive to incongruent prosodic cues (initial syllable lengthening: Saffran et al., 1996b; F0 peak: Vroomen et al., 1998).

When prosodic cues are congruent with other segmentation cues, the expected segmentation benefit or *redundancy gain* is far from being consistently observed in optimal listening conditions. In some ALL studies, adult listeners performed better when prosodic information (lengthening of and/or F0 peak on the "stressed" syllable of TP-words: Bagou, Fougeron, & Frauenfelder, 2002; Vroomen et al., 1998; final syllable lengthening of AL--words: Saffran et al., 1996b) was congruent with TPs than when only statistical information was available. However, in other studies the congruency of prosody with other segmentation cues did not have any impact. In ALL, Valian and Levitt (1996) only found a benefit promoted by "phrase prosody" (i.e., the rising pitch contour on the first two-word phrase of a

[4] TP is the conditional probability by which one syllable (x) predicts the following one (y): TP (y|x) = frequency (xy) / frequency (x).

sentence and a falling pitch on the other last two-word phrase) when listeners were unable to use other cues. Toro, Rodríguez-Fornells and Sebastián-Gallés (2007) reported that Spanish listeners were unable to learn TP-words stressed on their penultimate syllable, which is the default stress pattern of Spanish. Furthermore, they did not present better performance with TP-words stressed on either their first or last syllable (both stress patterns being legal – although not predominant – in Spanish) than when only TPs were available in the speech stream. Yet, these results should be interpreted with cautious: pitch variation (augmented by 20 Hz on the stressed syllable) was the acoustic correlate of stress used, but in Spanish duration seems to be the strongest acoustic correlate of lexical stress, regardless of the presence of a pitch accent (Ortega-Llebaria, 2006).

The impact of prosodic cues on speech segmentation is much clearer when the speech signal is degraded by noise superimposition. In this case, lexical stress is able to override any other incongruent segmentation cue, either lexically driven (e.g., the semantic context: Mattys et al., 2005) or signal derived (e.g., coarticulation and/or phonotactics: Mattys, 2004; Mattys et al., 2005).

As formalized by Mattys and colleagues (Mattys, 2004; Mattys et al., 2005), speech segmentation thus seems to be largely the product of the differential weighting of the types of information available in the signal and of the listening conditions. Their model is indeed the first integrated theoretical approach of the hierarchical organization of sublexical and lexical speech segmentation cues. It represents at three tiers several information types that are able to drive speech segmentation. The first, top, tier consists of lexical and post-lexical knowledge, which in adults is considered the most reliable information in optimal listening conditions. The second, middle, tier consists in the conjunction of segmental and subsegmental information, and the lowest tier corresponds to metrical prosody, which would act as a last-resource segmentation heuristic (see also Creel, Tanenhaus, & Aslin, 2006; Valian & Levitt, 1996), prevailing over information from the above tiers only when the signal is physically degraded.

However, this account may hold true only as regards language-dependent prosodic cues, namely those that vary across languages. This is the case of the location of lexical stress and whether it obeys to a fixed or varied pattern. For example, Finish has a fixed stress pattern, with stress always located at the first syllable of a word (Iivonen et al., 1998); Portuguese[5] like Spanish has a varied stress pattern (occurring in any one of the three last syllables of a polysyllabic word: d'Andrade & Laks, 1996; Mateus & d'Andrade, 2000), although it occurs by default on the penultimate syllable. Other prosodic cues

[5] European Portuguese (which is the native language of the participants of the present study) and Brazilian Portuguese share many properties, yet there are differences particularly regarding prosodic aspects (e.g., Frota & Vigário, 2001). For example, whereas vowel reduction is a prominent phenomenon in European Portuguese, it does not occur in Brazilian Portuguese (e.g., Abaurre & Galves, 1998). For the sake of simplicity, we will use the term "Portuguese" to refer to European Portuguese.

are used in any language from the onset of development (Dehaene-Lambertz & Houston, 1998) as well as by adults confronted with an unknown (or foreign) language (Shukla et al., 2007), probably because they are physiologically based (Grice, 2006; Shukla et al., 2007; Werner & Keller, 1994). This is the case of prosodic right-edges, which characterize IPs in many different languages (Werner & Keller, 1994). In fact, the acoustic marks of IP right-edge correspond to the slowing down of the articulators within a breath group, which is reflected in the signal as final lengthening and low pitch (Grice, 2006, see also Cutler et al., 1997).

Contrary to language-dependent prosodic cues, IPs prosodic contours also seem to prevail on statistical information in adults presented with intact speech. Indeed, with phonetically intact signal, Shukla et al. (2007) showed that IPs act as a preponderant segmentation cue over TPs between adjacent syllables. Using a variant of the ALL paradigm, when IPs were available (with left edge of IPs marked by a raising pitch and shorten duration of the beginning syllables, and right edge marked by a falling pitch and lengthen duration of final syllables), only TP-words within IPs (i.e., at middle positions) were correctly extracted from the stream. In contrast, TP-words straddling IPs (with TP-words' first syllables at the right-edge of one IP and the last syllable at the left-edge of the next IP) were not selected by listeners as "words" of the new language, probably because, although statistically cohesive, these TP-words straddled an important prosodic boundary. When some TP-words were aligned with prosodic edges (and thus were cohesive units on both statistical and prosodic grounds), while others were in IP's middle position (not prosodically marked and hence only cohesive on statistical grounds), only TP-words at IP edges, and hence supported by both types of information, were correctly extracted from the stream. Shukla et al. (2007) proposed that this type of prosodical information could act to filter out the output of statistical computation, with only TP-words compatible with it being selected.

These results are not necessarily incompatible with Mattys et al.'s (2005) model. These authors proposed that in the mature speech segmentation system the lowest weighted cues correspond to the ones earlier acquired. This is exactly the case of IPs, which are used much earlier in development than lexical stress. The role of the latter cue is modulated by its ability to predict word boundaries in a specific language (Dehaene-Lambertz & Houston, 1998; Houston et al., 2004). Under this view, these two types of prosodic cues may be differently weighted in adult segmentation.

In addition to the distinction between universal and language-dependent[6] prosodic cues, Fernandes, Ventura, and Kolinsky (2007) suggested that the

[6] Pitch, intensity and temporal variations associated to prosodic cues are also used in other domains (e.g., music) and by other species (e.g., birds). It is not our aim to discuss the (domain-)specificity of prosody, which is a highly debated issue (see e.g., Kolinsky, Cuvelier, Goetry, Peretz, & Morais, 2009; Peretz & Hyde, 2003; Moreno, Marques, Santos, Santos, Castro, & Besson, 2009). What seems clear, however, is that only some prosodic cues are used in all languages, while others vary across languages.

domain generality of the cues is also important. Indeed, the ability to track TPs, which also emerges very early in development (for speech segmentation, it is already observed at 6.5-month of age: Thiessen & Saffran, 2003; and for visual sequences, at 2-month of age: Kirkham et al., 2002), seems to involve a domain-general learning mechanism: TPs are also extracted in non-linguistic auditory materials (e.g., musical tones: Saffran, Johnson, & Aslin, 1999) and in visual sequences (Fiser & Aslin, 2001; Kirkham et al., 2002; but see Conway & Christiansen, 2005; 2006), with a possible phylogenetic origin (Hauser, Newport, & Aslin, 2001).

Both Mattys and colleagues' model (Mattys et al., 2005) and Fernandes et al.'s (2007) proposal would thus predict different interactions between domain-general segmentation cues like TPs and universal vs. language--dependent prosodic cues such as IPs and lexical stress, respectively. This is the general hypothesis we examined in the present study, particularly in what regards the impact of different listening conditions on the relative weighting of these cues.

Domain-general segmentation cues such as TPs are very resilient to physical degradation of the signal, being able to drive segmentation at similar levels in both intact and noisy listening conditions (degraded through white--noise superimposition, Fernandes et al., 2007). This is probably due to their fundamental role in speech segmentation. According to the same logic, IPs – another universal and fundamental cue – would be as resilient to signal degradation as TPs, or even more.

This prediction was tested in Experiment 1, using two (between--participants) signal quality conditions: *intact* (with no white noise superimposition) and *mildly degraded* (at 22dB SNR[7]). In order to assess the relative weighting of these cues in speech segmentation, within each signal quality condition the same AL was presented in three (between-participants) segmentation cue conditions differing by the number and congruence of the cues. In the *single-cue* condition, only TPs were available in the stream; in the *congruent-cues* condition, TPs suggested the same segmentation hypotheses as the IPs' right-edges; and in the *incongruent-cues* condition, TPs and IPs suggested different segmentation hypotheses, with TP-words spanning the prosodic boundary defined by IPs-edges.

In Experiment 2, we examined the extent to which a language-dependent prosodic cue such as lexical stress would act differently, in conjunction with TPs, than the universal prosodic cue examined in Experiment 1. In Experiment 2A we first ensured that duration is in fact an acoustic correlate of lexical stress in Portuguese (d'Andrade & Laks, 1996; Mateus & d'Andrade,

[7] The SNR ratio is measured as the noise intensity against the average signal intensity of the speech (i.e., here, the AL) signal. Here, since the signal intensity was 76 dB, the 22 dB SNR means that the noise intensity was set at 54 dB. This SNR was chosen on the basis of the identification in noise pretest used in Fernandes et al. (2007) for reducing intelligibility by approximately 50%.

2000). In Experiment 2B, the relative weighting of lexical stress (acoustically marked by syllable lengthening) and TPs in speech segmentation was evaluated using four (between-participants) conditions according to the stress pattern of TP-words. In one condition, AL stimuli were unstressed, and in the other three conditions TP-words were stressed on their first, second, or third (last) syllable; all these stressed patterns are permissible in Portuguese, although lexical stress occurs by default on the penultimate syllable. Based on previous findings (Mattys, 2004; Mattys et al., 2005), lexical stress would be treated as particularly reliable cue in impoverished listening conditions. Therefore, in Experiment 2B three (between-participants) signal quality conditions were adopted, i.e., intact, and mildly degraded conditions as in Experiment 1, and a 10 dB SNR[8], *strongly degraded* condition.

Experiment 1: Domain-general Statistical Cues vs. Universal Prosodic Cues.

In this experiment, the same AL was presented in three (between--participants) segmentation cue conditions differing by the number and congruence of the cues (see Table 1). In the two conditions in which the universal prosodic cue was available, IPs right-edges were acoustically marked by syllable lengthening and falling pitch, as happens in natural speech (e.g., Grice, 2006; Saffran et al., 1996b; Shukla et al., 2007).

After the AL familiarization phase, all participants performed the same ALL test, i.e., a two alternative forced-choice task. In this, they were asked to choose which, among two AL stimuli (always a TP-word vs. a *part-word*), was the "word" of the new language. Part-words are stimuli of the same length as TP-words, and that are constituted by AL syllables that have occurred adjacently in the AL stream during the familiarization phase (in some studies even with the same frequency of occurrence as TP-words: Aslin, Saffran, & Newport, 1998). The only difference between the two types of AL stimuli is their TP level: TP-words have higher TPs than part-words.

In the single-cue condition, this statistical difference between TP-words and part-words was the only available segmentation cue. In the congruent--cues condition, not only TP-words had higher TPs than part-words, but they were also acoustically marked by an IP right-edge (see Table 1), and hence TP-words segmentation was supported both by TPs and IPs. In the incongruent-cues condition the statistical segmentation outputs (i.e., the TP--words) were incongruent with the prosodic outputs since the first syllable of each TP-word was acoustically-marked with an IP right-edge. Therefore, whereas TP-words straddled an IP edge, part-words, and in particular the *part-*

[8] The 10 dB SNR corresponds to the superimposition of white noise with an intensity of 66 dB. This SNR was chosen on the basis of the identification in noise pretest used in Fernandes et al. (2007) for severely reducing signal intelligibility to a level of 16.6%

-*words 23#1* that are constituted by the two last syllables of a TP-word and the first syllable of the following TP-word, were plausible "words" according to IP-edges, but straddled TP-boundaries. Using TP-words and part-words, we could thus fully put statistical and prosodic information against each other. In particular, if listeners preferred prosodically plausible "words" even when these straddle a TP-boundary, they would discard the TP-words and consider part-words *23#1* as the correct units of the AL. However, this would not necessarily mean that listeners completely disregard the statistical information. To assess whether they were still sensitive to TPs (even if they considered it less reliable than IPs), the two-forced two alternative forced--choice task did not only include trials in which TP-words were confronted with part-words 23#1, but also trials in which TP-words were confronted with AL stimuli not supported by any one of the available segmentation cues, i.e., part-words 3#12 (see Table 1).

Table 1: Experiment 1: Orthographic translation of a sample of the stream heard in the familiarization phase in the four cue-conditions, and of the AL stimuli (TP-words; part-words 3#12 and part-words 23#1). The "#" defines word boundaries according to TPs, the "-" represents concatenation; the prosodic marked syllable is in bold capital letters and prosodic right-edges are marked by "]".

CUE-CONDITION (available cues)	SPEECH STREAM (familiarization phase)	PART-WORDS 3#12	PART-WORDS 23#1
Single-cue (TPs)	...-#bu-ka-la-#fu-fi-bu-#lu--fa-ba-#ki-la-bu-#...	ba-#ki-la	ka-la-#fu
Congruent-cues	...-#bu-ka-**LA**-]#fu-fi-**BU**-]#lu-fa-**BA**-]#ki-la-**BU**-]#...	**BA**-]#ki-la	ka-**LA**-]#fu
Incongruent-cues	...-#**BU**-]ka-la-#**FU**-]fi-bu-#**LU**-] fa-ba-#**KI**-]la-bu-#...	ba-#**KI**-]la	ka-la-#**FU**]

Method

Participants

Sixty-eight undergraduate students at the University of Lisbon participated in the experiment for a course credit. All were monolingual European--Portuguese speakers, with no reported history of speech or hearing disorders. Thirty-five were randomly assigned to the intact speech condition (12 in the single-cue condition, 11 in the congruent-cues condition, 12 in the incongruent-cues condition), and 33 to the physically degraded (22 dB SNR) condition (9 to the single-cue condition, 12 to the congruent-cues condition, 12 to incongruent-cues condition).

Material

All speech stimuli were synthesized using text-to-speech MBROLA software (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) with European-Portuguese female diphone database (http://tcts.fpms.ac.be/synthesis/mbrola.html) at 22.05 kHz and with a speech rate, close to conversational level, of about 270 syllables per minute.

The AL has already been tested and described by Fernandes et al. (2007). The selection of the vowels was particularly critical, since there is a close relation between vowel quality and lexical stress in Portuguese (Mateus & d'Andrade, 2000). The three vowels that constituted this AL phonological repertoire (i.e., /ɐ/, /i/, /u/) can occur in Portuguese in any position within a word (i.e., final and non-final, pre-stressed and post-stressed) and can also be either stressed or unstressed (Mateus & d'Andrade, 2000).

The AL included six TPs words (/lufɐbɐ/, /fufibu/, /kilɐbu/, /bukɐlɐ/, /bɐbuku/, and /kɐfubi/). TPs were always higher within TP-words than between TP-words, i.e., than within part-words, with average TPs[9] of 0.68 and 0.38, respectively.

All part-words were constituted by syllables of two different TP-words that occurred adjacently in the speech stream during the familiarization phase. Three of them (*part-words 3#12*; e.g., /bɐkilɐ/) consisted of the last (third) syllable of one TP-word (e.g., /bɐ/, the last syllable of /lufɐbɐ/) and the first two syllables of the next (e.g., /kilɐ/, the first two syllables of /kilɐbu/). The other three (*part-words 23#1*; e.g., /fibulu/) consisted of the last two syllables of a TP-word (e.g., /fibu/, the last syllables of /fufibu/) and the first syllable of the next (e.g., /lu/, the first syllable of /lufɐbɐ/).

Familiarization Phase. Three synthesized versions of the AL (defined according to the number and congruence of the available cues) were created. Each version included the same sequence of syllables, divided into three

[9] The average TPs were computed by averaging the two TPs associated to each trisyllabic AL stimulus (the TP between the first and second syllable, and the TP between the second and the third syllable).

listening 7-minutes blocks (rendering 21-minutes). Each block was created by concatenating 105 tokens of each TP-word (1890 syllables, 630 tokens of words) with the only criterion that two tokens of the same TP-word never occurred adjacently in the stream.

In the single-cue condition, only TPs between adjacent syllables could help listeners to locate word-boundaries since no acoustic cues were available: the speech stream presented a flat 220 Hz pitch and the average duration of all syllables was equivalent (i.e., 222 ms). In the other two conditions, both TPs and IPs were available. The prosodic information (IPs) was added by acoustically marking one syllable of each TP-word by 150 ms lengthening and linearly decreasing its pitch by 20 Hz, while the other syllables remained unchanged.

In the congruent-cues condition, TPs and IPs suggested the same word--boundaries, since the last syllable of each TP-word was acoustically marked, defining a prosodic right-edge. In the incongruent-cues condition the first syllable of each TP-word was acoustically marked defining the prosodic right--edge. Thus, TPs suggested that TP-words were the "lexical units" of the AL but these straddle a prosodic boundary. Conversely, IPs suggested that the part-words 23#1 (straddling TP-boundaries) that ended with the prosodically marked syllable were plausible words of the new language. Part-words 3#12 were neither supported by the statistical information (they had lower TPs than TP-words) nor by the prosodic information (they spanned an IP's edge).

For all cue-conditions, the AL signal was either acoustically intact or mildly degraded by white-noise superimposition at a 22 dB SNR, using *Adobe Audition 1.5*.

Forced-choice test phase. The three syllables that constituted each TP--word and each part-word were synthesized with the same average duration and 220 Hz flat pitch (with no prosodic acoustic correlates available), and concatenated (with no white-noise).

The forced-choice test included 36 trials corresponding to the exhaustive combination of all six TP-words and six part-words. In half of the trials, TP--words were confronted with part-words 3#12, and in the other half, TP-words were confronted with part-words 23#1 (for an orthographic illustration of the material, see Table 1).

Procedure

Participants were tested individually or in groups of two in a sound--attenuated room. Presentation of the AL familiarization phase was done with *Windows Media Player* through *Sennheiser HD 280 Professional Silver* headphones. For the forced-choice test phase, stimuli were also presented through headphones, with presentation, timing and data collection controlled by *E-Prime 1.1* (Schneider, Eschman, & Zuccolotto, 2002a, b).

All participants were instructed to listen to a new language. They were told that the language contained "words", but no meaning or grammar, and

were asked to try to find out what words constituted it. No information about structure, phonology, prosody, or length of the words was given. Participants in the degraded signal condition were warned about the poor quality of the signal.

After the 21-minute AL familiarization phase, all participants performed the same ALL test with phonetically intact stimuli. Each trial started with a warning tone, followed by two trisyllabic strings (a TP-word and a part-word) separated by 500 ms of silence. On each trial, participants had to choose which one of two strings – either the first or the second one – was a word of the language heard in the first phase, by pressing the "1" or "2" key on the keyboard, respectively. The next trial started immediately after participants gave their answer or if no answer was registered after a maximum of 10 seconds. Response accuracy was emphasized, although instructions required participants to answer to all trials, even if they were not totally sure about their decision.

Four practice trials were provided prior to the test in order to clarify the test structure and enable practice with key presses. On each practice trial participants had to decide which of two (an animal and an environmental) sounds was the animal sound. Feedback on the correctness of responses was only provided for these practice trials.

Order of presentation of test trials was randomized for each participant, and the order of AL stimuli within trials was counterbalanced within each group.

Results and Discussion

We evaluated whether TP-words choices were influenced by the type of part--words to which they were confronted with in the test phase (see Figure 1), through a mixed ANOVA with part-word type (part-words 3#12 vs. 23#1: within-participants), signal quality (intact, mildly degraded: between--participants) and cue-condition (single-cue; congruent-cues; incongruent--cues: between-participants) as factors.

All main effects were significant [cue-condition: $F(2, 62) = 56.65$, $p < .0001$; $MSe = 8.31$, $\eta_p^2 = .64$; signal quality: $F(1, 62) = 3.89$, $p = .05$; $MSe = 8.31$, $\eta_p^2 = .06$; part-word type: $F(1, 62) = 10.31$, $p < .005$; $MSe = 6.45$, $\eta_p^2 = .14$].

The interaction between signal quality and part-word type was not significant [$F < 1$], but the effect of cue-condition was modulated by part--word type [$F(2, 62) = 3.10$, $p < .05$; $MSe = 6.45$, $\eta_p^2 = .10$]. Since the three--way interaction was not significant [$F(2, 62) = 1.9$, $p = .15$], we further evaluated performance according to part-word type, separately in each cue--condition.

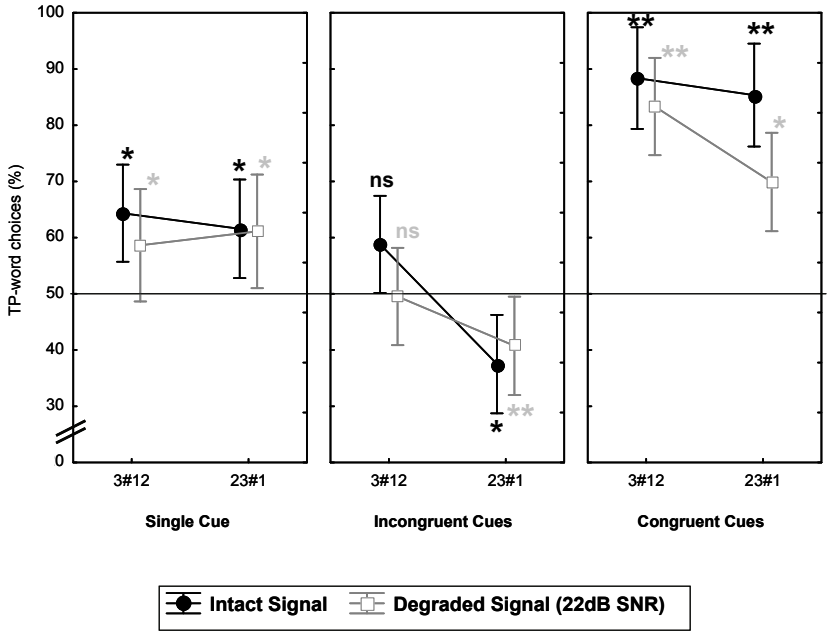

Figure 1: ALL performance pattern (proportion of TP-word choices, in percentage) broken--down by part-word type (3#12; 23#1), according to signal quality (intact; degraded) and cue-condition (single-cue; incongruent-cues; congruent-cues) in Experiment 1. Vertical bars denote standard error of the mean on each condition. Chance level corresponds to 50%. Level of significance for local one-sample *t-test* comparisons with chance-level is also indicated (i.e., ns: $p > .10$; *: $p < .05$; **: $p \leq .01$).

As illustrated in Figure 1, listeners' performance in the single-cue condition was not affected by signal quality [$F < 1$]. TPs were able to drive segmentation at a similar level with degraded and intact speech, confirming their resilience to physical noise (Fernandes et al., 2007). Furthermore, overall ALL performance, which reached 61.4%, on average, was above chance [$t(20) = 3.05$, $p < .01$] and was not modulated by the part-word type to which TP-words were confronted with [$F < 1$, $p \approx 1$].

Listeners exposed to incongruent cues presented the worst performance, in comparison to both the single-cue and congruent-cues conditions [$F(1, 62) = 18.8$ and $= 112.7$, respectively, both $p < .0001$]. As it was the case in the single-cue condition, performance was not modulated by signal quality [$F < 1$], but here it was strongly affected by part-word type [$F(1, 62) = 13.6$, $p < .001$]. Indeed, whereas their performance was at chance when TP-words were confronted with part-words 3#12, [$t(23) = 1.34$, $p > .10$], listeners systematically chose the part-words 23#1 rather than the TP-words as the

lexical units of the AL [$t(23)= -3.21$, $p <. 005$]. Thus, when TP-words were confronted with prosodically plausible "words", participants chose the latter.

The preference for part-words 23#1 over TP-words in the incongruent--cues condition was in sharp contrast to the response pattern found in the other two cues-conditions [vs. single cue: $F(1, 62) = 23.7$, $p < .0001$; vs. congruent cues: $F(1, 62) = 75.4$, $p < .0001$]. This suggests that the universal prosodic cue prevailed over the domain-general statistical cue: IPs were able to drive segmentation even in the presence of an incongruent statistical cue, at any signal quality condition at study.

In the congruent-cues condition, listeners were able to correctly choose TP-words as the lexical units of the new language, reaching on average 81.7%, an above-chance performance [$t(22)= 13.4$, $p < .0001$]. They had the best performance in comparison to both the incongruent-cues condition [$F(1, 62) = 112.6$, $p < .0001$] and the single-cue condition [$F(1, 62) = 39.93$, $p < .0001$], hence showing a redundancy gain in comparison to the latter. Their performance was however modulated by the type of part-words to which TP-words were confronted with, being slightly better with part-words 3#12 than 23#1 [$F(1, 62) = 3.90$, $p = .05$]. This was unexpected since in the congruent-cues conditions the two types of part-words were neither supported by TPs nor by IPs.

In sum, the present results cohere with Shukla et al.'s (2007) conclusions and add to previous findings. Not only universal prosody is more preponderant in speech segmentation than domain-general TPs in any listening condition, but it is also as resilient to physical noise as the domain--general statistical cue. Moreover, listeners exposed to incongruent-cues did not only consider universal prosody as more reliable than TPs but were also unable to choose between AL stimuli supported by TPs (i.e., TP-word) and AL stimuli not supported by any segmentation cue at all (part-words 3#12). This performance pattern suggests that universal prosody was filtering out the statistical segmentation outputs.

Most probably, this would not be the case of a language-specific prosodic cue like lexical stress. This issue was examined in Experiment 2.

Experiment 2

The aim of Experiment 2 was to contrast TPs with lexical stress. Yet, before examining this issue in Experiment 2B, in Experiment 2A we first ensured that syllable duration alone is considered as an acoustic correlate of "lexical" stress by Portuguese listeners presented with synthesized nonsense trisyllabic sequences.

Experiment 2A. Duration as an acoustic correlate of lexical stress in Portuguese.

Syllable lengthening (Mateus & d'Andrade, 2000), but not pitch (Grønnum & Viana, 1999), is one of the strongest correlate of lexical stress in Portuguese. Indeed, stress vowels are acoustically differentiated from unstressed ones by their longer duration, and Portuguese listeners seem to treat duration as the most important correlate of lexical stress (Delgado-Martins, 2002; d'Andrade & Laks, 1996). Moreover, vocalic reduction is a prominent phenomenon in European-Portuguese: while stressed syllables are lengthened, unstressed vowels are reduced in fluent speech, and in most cases even completely eliminated from the stream (Mateus & d'Andrade, 2000).

In the present experiment, we used a three-alternative forced-choice task to evaluate whether Portuguese listeners would consider as being stressed the lengthened syllable of synthesized nonsense trisyllabic sequences that were similar to the TP-words and part-words used in the test phase of Experiment 1.

Method

Participants

Thirty-eight fresh undergraduate psychology students (19 in the unstressed and 19 in the stressed conditions) at the University of Lisbon participated in the experiment for a course credit. All were monolingual European--Portuguese speakers, with no reported history of speech or hearing disorders.

Material

Four exemplars of each AL stimulus were created according to lexical stress location (i.e., unstressed; stressed on the 1st; 2nd or 3rd syllable). The unstressed exemplars corresponded to the TP-words and part-words used in the test phase of Experiment 1.

In the stressed materials, stress was acoustically marked in each TP-word and part-word on one of the three possible syllables. This was done by lengthening by 100 ms the stressed syllable and shortening by 50 ms each unstressed syllable, while pitch remained flat (220 Hz). The manipulation of duration of unstressed syllables was done because, besides lengthening of the stressed syllable, unstressed syllables are usually reduced in European--Portuguese (Mateus & d'Andrade, 2000).

One list with all 12 unstressed AL stimuli (6 TP-words and 6 part-words) was created, which was only presented to participants in the unstressed condition (i.e., the control group). The 36 stressed AL stimuli were distributed

across participants through three lists, each one including both TP-words and part-words stressed in one of the three possible syllables. The three stressed lists had the same proportion of TP-words and part-words in each lexical stress pattern condition, but each AL stimulus presented a different stress pattern in each one of the three lists.

Procedure

Participants were tested individually or in groups of two. Presentation, timing and data collection was controlled by *E-Prime 1.1* (Schneider et al. 2002a, b).

Participants in the unstressed condition were only presented with the unstressed items and participants in the stressed condition were randomly assigned to one of the three stressed lists (6 participants to list 1 and to list 2, and 7 to list 3). All were informed that on each trial they would hear through headphones a warning tone immediately followed by a trisyllabic pseudoword. Their task was to identify the stressed syllable of the stimulus presented on each trial by pressing the "1", "2", or "3" keyboard key, and corresponding to the first (antepenultimate), second (penultimate), or third (last) syllable, respectively. Participants were not informed about the presence/absence and the type of acoustic correlate used, and response accuracy was emphasized. Another trial began immediately after participants gave their answer or after a maximum of 10 seconds if no answer was registered. Three practice trials with real words with lexical stress located on the first (i.e., /ˈpɐniku/ *panic*), second (i.e., /bɐˈnɐnɐ/ *banana*) or third (i.e, /ʊɐvɐˈli/ *wild boar*) syllable were provided prior to the test in order to clarify its structure and enable practice with key presses.

Order of presentation of test trials was pseudo-randomized for each participant with the only criterion that two stimuli with the same stress pattern did not occur twice in a row.

Results and Discussion

Listeners' choices in each stress condition are presented in Figure 2. Although the material presented to the two groups of listeners differed on the presence/absence of an acoustic correlate of lexical stress (here, duration of the syllables), both groups performed the same task on the same phonological material and thus listeners' choices in the unstressed condition and listeners' correct responses in the stressed condition can be directly compared.

Performance in the unstressed condition was also compared with chance level (33.33%). Overall performance was at chance (on average, 33.3%, SE = 2.7), as expected when no acoustic correlate of stress is available. However, closer inspection of the data shows that listeners' choices of the stressed syllable on the second syllable, which is compatible with the default pattern

on their native language, was above chance [$t(18) = 4.2$, $p < .001$], in contrast to their choices of either the first [$t(18) = -1.3$, $p > .10$] or third [$t(18) = -1,7$, $p = .10$] syllable.

In the stressed condition, listeners were able to correctly locate the stressed syllable of each AL stimulus, performing above chance for all stress locations [$t(18) = 2.8$, $p = .01$, = 4.9, $p < .001$, and = 2.5, $p < .05$, for stimuli stressed on the 1st; 2nd or 3rd syllable, respectively].

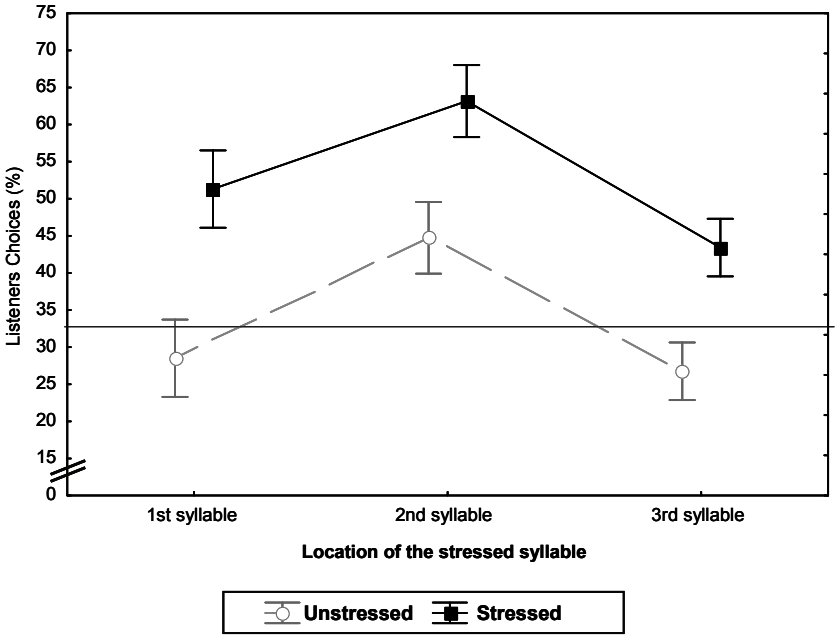

Figure 2: Listeners' proportion of responses on the three-alternative forced choice stress location task, broken-down by responses (1st syl; 2nd syl; 3rd syl) 23#1), in the two acoustic cue conditions (unstressed vs. stressed) of Experiment 2. Vertical bars denote standard error of the mean. Chance level corresponds to 33.3%.

The mixed ANOVA ran on the proportion of responses in the unstressed condition and (of correct responses) in the stressed condition, with material (unstressed vs. stressed: between-participants) and chosen stress location (1st-; 2nd; 3rd syllable: within-participants) revealed a significant main effect of group [$F(1, 36) = 26.2$, $p < .001$; $MSe = 0,04$; $\eta_p^2 = .42$]. Listeners exposed to the material with stress acoustically marked presented a higher proportion of (correct) stress localization than the proportion of responses found for listeners in the unstressed condition. Thus, Portuguese listeners are indeed sensitive to syllable duration as an acoustic correlate of lexical stress.

The main effect of chosen stress location was also significant [$F(2, 72) =$, $p < .001$; $MSe = 0{,}04$, $\eta_p^2 = .19$], with higher proportion of second syllable responses than of first or third-syllable responses [$F(1, 36) = 10.0$, $p < .005$, and $= 16.2$, $p < .001$, respectively]; the two latter response rates did not differ from each other [$F < 1$]. No interaction between the two factors was found [$F < 1$].

The present results corroborate the role of duration as an acoustic correlate of lexical stress in Portuguese (Delgado-Martins, 2002; Mateus & d'Andrade, 2000). Indeed, even with such an impoverished material as synthesized nonsense stimuli, listeners were quite able to correctly locate the stress syllable of trisyllabic stimuli when it was acoustically marked by duration (i.e., lengthened of the stressed syllable in comparison to the unstressed ones).

Interestingly, listeners who were presented with stimuli with no acoustic marker of lexical stress at all often considered that unstressed pseudowords were paroxytones, even though they obviously presented a lower proportion of such responses than listeners exposed to acoustically marked paroxytones. We cannot ensure whether this corresponds to a decisional bias or to a perceptual *stress illusion* (see Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999), but it clearly illustrates the fact that the suprasegmental characteristics of the native language affect processing of unfamiliar items (e.g., Dupoux, Pallier, Sebastian-Gallés, & Mehler, 1997).

In any case, the most important outcome of this experiment is that syllable lengthening alone was enough to enhance participants' performance in locating the stressed syllable of synthesized nonsense stimuli. This validates the manipulation of syllable length as an acoustic correlate of lexical stress, the language-specific prosodic cue which effects were examined in Experiment 2B.

Experiment 2B: Domain-general Statistical Cues vs. Language-dependent Prosodic Cues.

Previous results have suggested that lexical stress is a particularly preponderant segmentation cue in severely impoverished listening conditions. Indeed, when lexical stress was incongruent with an acoustic-phonetic or a phonotactic cue (Mattys et al., 2005; Experiments 1A and 2, respectively), lexical stress was only able to drive speech segmentation processing in a severely degraded condition (i.e., with a SNR of -5 dB). The same held true with incongruent higher-level information (i.e., lexical or semantic contexts: Mattys et al., 2005; Experiment 6A) but only with severely degraded signal (i.e., -5 dB) and not when the signal was either mildly (i.e., 5 dB SNR) or moderately (i.e., 0 dB SNR) degraded.

In Mattys et al.'s study, the material used corresponded to real words and in that case a SNR of 0 dB is known to reduce intelligibility of isolated words to about 50%, and hence a SNR of -5 dB corresponds to a severe reduction of

word intelligibility. Based on the tests ran by Fernandes et al. (2007), we already know that with AL material a SNR of 10 dB reduces the intelligibility of (nonsense) AL material to about 16.7%. This SNR thus corresponds to a strongly degraded condition in ALL. In order to evaluate whether stress pattern effects would only emerge in severely degraded conditions, in Experiment 2B, within an ALL setting three signal quality conditions were considered: an intact signal condition, a mildly degraded similar to the one of Experiment 1 (22 dB SNR), and a strongly degraded condition (10 dB SNR).

Within each signal-quality condition, four (between-participants) conditions of TP-words stress pattern were created. In the *TP–unstressed* condition, only TPs were available in the stream. This condition was thus identical to the single-cue condition of Experiment 1. In the *stressed* conditions, the stressed syllable of some exemplars of TP-words was lengthened by 100 ms, while the other (unstressed) syllables of these TP--words were reduced by 50 ms each. As illustrated in Table 2, in the *1st syllable-stressed condition*, stress was located on the first syllable; in the *2nd syllable-stressed* condition, stress was located on the penultimate syllable (i.e., the default lexical stress pattern in Portuguese); and in the *3rd syllable-stressed* condition, stress was located on the last syllable.

If a language-specific prosodic cue like lexical stress had the same status in speech segmentation as a universal prosodic cue like IPs, lexical stress would also filter out statistical units that are not compatible with it. In comparison to the other conditions, we would thus expect to find a performance gain when TP-words present the default stress pattern of the listeners' native language, i.e., in the 2nd-syllable stressed condition, since in the other stressed conditions listeners would consider the paroxytone AL stimuli (i.e., the 3#12 part-words in the 1st syllable-stressed condition and 23#1 part-words in the 3rd syllable-stressed condition, see Table 2) rather than the TP-words as the "words" of the AL.

However, with intact signal, lexical stress is not able to drive segmentation when other cues are available in the stream (Mattys et al., 2005; Valiant & Levitt, 1996). Moreover, in European Portuguese, lexical stress does not usually signal a word boundary, especially for polysyllabic words[10]. Stress

[10] Inspection of the Portuguese *Porlex* database (Gomes & Castro, 2003) reveals that from the 29,238 entries, about 98% (i.e., 28,859) are words with two or more syllables, with 59% of them being paroxytones. Considering only words with the same phonological structure as the AL stimuli used in the present study (i.e., CV.CV.CV; which correspond to about 14% of the entries), the predominance of paroxytone words is even clearer (i.e., 82%), while only 16% and 2% corresponds to proparoxytones and oxytones, respectively. Taking these statistical facts into account, stressed syllables in Portuguese likely correspond to the penultimate syllable of a word, which in the case of trisyllables correspond to the second syllable. Thus, contrary to the case of English (e.g., Cutler et al., 1997; Mattys, 2004), in Portuguese the stressed syllable does not define a possible word onset (nor a word ending, as in French), which would seem to reduce the role of lexical stress in speech segmentation.

may thus be considered by Portuguese listeners as an unreliable segmentation cue. If this were the case, as since we know that TPs are also very resilient to the physical degradation of the signal (Fernandes et al., 2007), it could be the case that in the strongly degraded condition listeners would take advantage of the joint assistance of both cues, being able to parse only the statistically cohesive units obeying to the default stress pattern of their native language, i.e., paroxytone TP-words.

Table 2: Experiment 2B: Orthographic translation of a sample of the stream heard in the familiarization phase in the four conditions of stress location. The "#" defines word boundaries according to TPs; the "-" represents concatenation; TP-words with stressed syllables are in bold and the stressed syllable is presented in capitalized letters.

STRESS PATTERN	SPEECH STREAM	PART-WORDS 3#12	PART-WORDS 23#1
TP–unstressed	...#lu-fa-ba-#ki-la-bu-#ka-fu-bi-#ba-bu-ku-#bu-ka-la-#fu-fi-bu-#…	ba-#ki-la	ka-la-#fu
1st syllable-stressed	...#lu-fa-ba-#**KI**-la-bu-#ka-fu-bi-#ba-bu-ku-#bu-ka-la-#**FU**-fi-bu-#…	ba-#**KI**-la	ka-la-#**FU**
2nd syllable-stressed	...#lu-fa-ba-#ki-**LA**-bu-#ka-fu-bi-#ba-bu-ku-#bu-**KA**-la-# fu-fi-bu-#…	ba-#ki-**LA**	**KA**-la-#fu
3rd syllable-stressed	...#lu-fa-**BA**-#ki-la-bu-#ka-fu-bi-#ba-bu-ku-#bu-ka-**LA**#fu-fi-bu-#…	**BA**-#ki-la	ka-**LA**#fu

Method

Participants

One hundred and twenty-eight psychology students at the University of Lisbon participated in the experiment for a course credit. Forty-three were

randomly assigned to the intact speech condition (12 to the TP–unstressed condition, 11 to the 1st syllable-stressed condition, 7 to 2nd syllable-stressed condition, and 13 to the 3rd syllable-stressed condition), 47 were assigned to the mildly degraded (22 dB SNR) condition (9 to TP–unstressed condition, 11 to 1st syllable-stressed condition, 14 to 2nd syllable-stressed condition, and 13 to 3rd syllable-stressed condition), and 38 to the strongly degraded (10 dB SNR) condition (9 to TP–unstressed condition, 10 to 1st syllable-stressed condition, 10 to 2nd syllable-stressed condition, and 9 to 3rd syllable-stressed condition).

Material and Procedure

All AL material was synthesized using the same method as in Experiment 1.

For the familiarization phase, four synthesized versions of the AL were created. Each version included the same sequence of syllables, divided into three 7-minutes listening blocks (rendering 21-minutes) as in Experiment 1.

The TP–unstressed version of the AL corresponded to the single-cue version of Experiment 1. In the stressed versions, the acoustic correlate of lexical stress was syllable duration. Stressed syllables were lengthened by 100 ms while unstressed ones were reduced each by 50 ms. Like Fernandes et al. (2007) did with their manipulation of coarticulation, only some (approximately one third of the exemplars) of the TP-words within the AL stream presented a stress pattern.

Since lexical stress – the last-resort segmentation heuristic (cf. Mattys et al. 2005; see also Valiant & Levitt, 1996) – is an abstract property often not acoustically realized, in Experiment 2B, only some exemplars of the TP--words presented a stress pattern, and hence any prosodic effect to be found would not be simply due to the unnatural isochronal repetition of an acoustical pattern, but instead to listeners' sensitivity to the available lexical stress information. Also, this procedure allowed direct comparison of the relative weighting of TPs and lexical stress vs. coarticulation. Indeed, we already know that, with this kind of manipulation, participants are perfectly able to use coarticulation in segmenting an intact signal (Fernandes et al., 2007).

The stressed TP-words were located as similarly as possible within each one of the three stressed conditions of the AL. The only between-conditions difference was the location of the stressed syllable: for the 1st syllable-stressed condition, it was the first syllable; for the 2nd syllable-stressed condition, it was the second; and for the 3rd syllable-stressed condition, it was the last syllable (see Table 2).

In order to create the two degraded conditions, white-noise was superimposed at 22dB SNR and 10dB SNR, using the same method as in Experiment 1.

The ALL forced-choice test and the procedure were identical to that of Experiment 1.

Results and Discussion

In the mixed ANOVA ran on participants' TP-word choices with part-word type (part-words 3#12 vs. 23#1) as within-participants factor and signal quality (intact, 22dB SNR, 10dB SNR) as well as lexical stress pattern (TP–unstressed, 1st syllable-stressed, 2nd syllable-stressed, 3rd syllable-stressed) as between-participants factors, neither the main effect of lexical stress nor its interaction with any of the other factors at study were significant [*F*s < 2, *p* >.10]. ALL performance was only affected by signal quality [$F(2, 116) = 5.9$, $MSe = 8.72$, $p < .005$; $\eta_p^2 = .09$], linearly declining with signal degradation [$F(1, 116) = 11.7$, $p < .001$: average scores of 62, 57, and 53% in the intact signal, mildly degraded and strongly degraded conditions, respectively].

Thus, contrary to what was found with IPs (Experiment 1), with lexical stress no overall impact of prosody was found on ALL. This coheres with previous findings on the unreliability of prosody in speech segmentation when other cues are also available, at least when the signal is intact (e.g., Vallian & Levitt, 1996).

Still, on the basis of previous results (e.g., Mattys, 2004; Mattys et al., 2005), we would have expected lexical stress to impact on speech segmentation with degraded signal. In order to evaluate this possibility more thoroughly, we examined performance separately for *low-TP words* and *high--TP-words* TP-words. Indeed, three TP-words presented higher TPs (ranging from 0.75 to 1.00) than the other three (ranging from 0.50 to 0.58). This distributional gradient is probably similar to what happens in natural languages (Saffran et al., 1996b) and might allow a fine-grained evaluation of any TPs effect. In particular, listeners are sensitive to the TP-gradient of statistical segmentation outputs (Saffran et al., 1996b). We thus evaluated whether the TP-level of TP-words had any impact on performance or interacted with lexical stress[11].

In the mixed ANOVA ran with TP-words type (high- vs. low-TP-words) as within-participants factor and signal quality as well as lexical stress pattern as between-participants factors, there was a significant TP-gradient [$F(1, 116) = 6.3$, $MSe = 5.34$, $p < .01$; $\eta_p^2 = .05$]: overall, listeners presented better performance for high- than for low-TP words (on the average, 59.5 and 55.1 %, respectively). The main effect of signal quality, already observed in the previous analysis, was also significant [$F(1, 116) = 6.1$, $MSe = 8.74$, $p < .005$; $\eta_p^2 = .09$] as well as the interaction between these two factors [$F(2, 116) = 3.5$, $MSe = 5.34$, $p < .05$; $\eta_p^2 = .06$]. Since the three-way interaction was also significant [$F(6, 116) = 3.0$, $MSe = 5.34$, $p < .01$, $\eta_p^2 = .14$], we next

[11] It is worth mentioning that this was note the case in Experiment 1, in which neither the main effect of TP-level [$F(1, 62) = 2.87$, $p = .09$; Mse = 4.36, ηp2 =.04] nor the interactions with other factors [all $p \geq .10$] were significant.

evaluated separately each signal quality condition (see Table 3). No other effect was significant [all $Fs < 1$].

With intact signal, no significant effect was found [all Fs < 1.5]. All groups performed above chance (see Table 3) and at similar levels, with no differences for low- and high-TP-words, independently of the absence/presence of an acoustic marker of stress and of its location in TP--words. Thus, with intact signal, listeners were able to use statistical information to correctly parse the TP-words from the speech stream, independently of the stress pattern of the stimuli.

With the mildly degraded signal (22dB SNR), the main effect of TP-level was significant [$F(1, 43) = 12.8$, $p < .001$, $Mse = 5{,}83$, $\eta_p^2 = .23$]. No main effect of stress pattern was found [$F(3, 43) = 1.3$]. The interaction between the two factors did not reached the conventional level of significance [$F(3, 43) = 1.9$, $p = .10$, $Mse = 5{,}42$, $\eta_p^2 = .12$] but the size of this effect was moderate (Cohen, 1988). In fact, in line with previous findings (i.e., Fernandes et al., 2007), and as indicated in Table 3, in the TP-unstressed condition, listeners were as able to choose both high- as low-TP-words [$F < 1$] as being the "words" of the new language, and they did so above chance in both cases (see Table 3). In sharp contrast, a significant TP-gradient effect was found for listeners exposed to acoustically marked stressed sequences [$F(1, 43) = 17.5$, $p < .005$], reflecting the fact that only high-TP-words were extracted from the stream. For these items, participants presented above-chance performance, which reached 63.2%, on the average [$t(37) = 6.0$, $p < .0005$]. Performance for low-TP-words did not differ from chance [$t < 1$], reaching only 52%, on average, and was obviously poorer than the one found in the TP–unstressed condition [$F(1, 46) = 4.0$, $p = .05$].

This result pattern suggests that, although listeners' performance is not affected overall by acoustically marked lexical stress patterns (since no main effect of stress pattern was found), lexical stress does play some role in speech segmentation in mildly degraded listening conditions. In particular, lexical stress seems to have restricted the extraction of statistical segmentation outputs to those with the highest support. Indeed, when lexical stress was available in the stream, only high-TP-words were correctly selected above chance as the plausible words of the AL, and this effect was independent of the precise stress pattern of the stimuli. This is not surprising: any one of the three stress patterns we used here is in fact legal in Portuguese, even though by default Portuguese words are paroxytones.

The modulator impact of lexical stress in the extraction of statistical outputs is also observed with the strongly degraded signal (10 dB SNR). In this condition, only the interaction between stress pattern and TP-level was significant [$F(3, 34) = 5.7$, $p < .005$, $Mse = 4{,}069$; $\eta_p^2 = .33$; all other $Fs < 1.2$]. As already happened in the mildly degraded condition, only listeners exposed to TP-unstressed sequences were able to correctly extract above chance both high- and low-TP-words (see Table 3), and they did so at similar

levels for both types of TP-words [$F = 1$]. When stress cues were available, listeners were not able to extract the low-TP-words at all, presenting performances at chance [$t < 1$]. Contrary to what had been observed with the mildly degraded signal, here performance on high-TP-words varied as a function of the stress pattern. In fact, listeners exposed to a stress pattern diverging from the default one in their native language (i.e., in the 1st and 3rd syllable-stressed conditions) presented overall performances at chance level, reaching 48.6 and 53.7% on average, respectively [ts < 1]. For participants exposed to the AL with the default lexical stress pattern in Portuguese (2nd stressed syllable condition), there was a significant TP-gradient [$F(1, 34) = 12.6$, $p = .001$]. This TP-gradient corresponded to the fact that, although these listeners were not able to extract low-TP-words from the stream (see Table 3), they were quite able to decide that high-TP-words were the "words" of the new language: they did so above chance (see Table 3), and at a level similar to the one of listeners exposed to statistical information only [$F = 1$].

Table 3: Experiment 2B: Mean TP-words choices (in percentage) for both low-TP--words and high-TP-words for each condition considering signal-quality and lexical stress pattern. Standard errors of the mean in each condition are presented in parenthesis.

	SIGNAL-QUALITY					
	Intact		Mildly degraded		Strongly degraded	
Lexical Stress pattern	High--TP--words	Low--TP--words	High--TP--words	Low--TP--words	High--TP--words	Low--TP--words
TP-unstressed	67.6 (4.2) **	58.4 (4.3) *	60.5 (4.8) *	59.2 (5.0) *	56.8 (4.8) *	61.2 (5.0) **
1st syllable stressed	61.1 (4.4) **	57.1 (4.5) *	59.6 (4.4) *	43.9 (4.5) ♦	45.5 (4.6) ♦	51.7 (4.7) ♦
2nd syllable stressed	59.5 (5.5) *	66.7 (5.6) **	66.3 (3.9) **	51.6 (4.0) ♦	59.4 (7.6) *	41.7 (4.7) ♦
3rd syllable stressed	60.7 (4.0) **	63.2 (4.1) **	63.7 (4.0) **	55.1 (4.1) ♦	54.9 (4.8) ♦	52.5 (5.0) ♦

One-way *t*-tests (in comparison to chance level):
* $p < .05$
** at least $p < .01$
♦ scores at chance level, $t < 1$.

Thus, in physically degraded listening conditions lexical stress seems to impact speech segmentation, even when another resilient cue (i.e., TPs) is also

available in the stream. Since statistical information is as able to drive speech segmentation with intact as with physically degraded signal (Fernandes et al., 2007), the ALL impairment observed in the stressed conditions of the present experiment can only be attributed to an impact of language-dependent prosodic information on segmentation.

In sum, with phonetically intact speech, no impact of lexical stress pattern was found on ALL performance: both high- and low-TP-words were correctly extracted from the stream at similar levels, in both stressed and unstressed conditions. With mildly degraded signal, a lexical stress effect emerged, narrowing the extraction of AL units: in stressed conditions, low-TP-words that with intact speech were correctly parsed from the stream were now no longer selected as possible "words" of the new language. This led to a much poorer performance with mildly degraded signal than with intact speech [$F(1, 116) = 10.7$, $p = .001$]. In line with Mattys et al.'s (2005) results demonstrating the reliability of lexical stress cues in speech segmentation with strong noise, the 10 dB SNR condition maximized the strength of the lexical stress effects: only statistical outputs that obeyed to the default stress pattern in Portuguese were selected as plausible words of the new language. Statistical learning was inhibited in the stressed conditions diverging from the default one in Portuguese. For both 1st and 3rd syllable-stressed conditions, with intact signal both low- and high-TP-words were correctly chosen as the units of the AL, but with mildly degraded signal only those with high-TPs were still extracted, and with strongly degraded signal no statistical outputs were extracted at all. This was also demonstrated by the significant linear trend of signal quality in ALL performances in these stressed conditions [$F(1, 116) = 6.9$, $p < .01$].

This result pattern does not seem to be compatible with the suggestion that the role of any prosodic cue in speech segmentation would be one of filtering out statistical byproducts. Had this been the case, we would have found an overall above chance performance with no TP-gradient effect in the 2nd syllable-stressed condition, since in that condition all statistical segmentation outputs (both high- and low-TP-words) were congruent with the listeners' native language default stress pattern. Instead, in the strongly degraded condition, only statistical outputs supported by the strongest evidence (i.e., high-TP-words obeying to the default stress pattern) were extracted from the stream.

General Discussion

Recent studies regarding different segmentation cues (e.g., Fernandes et al., 2007; Mattys et al., 2005; Shukla et al., 2007) cohere with computational simulations (e.g., Christiansen & Curtin, 2005) in suggesting that speech segmentation does not correspond to the sum of independent byproducts of different segmentation cues. Instead, as proposed by Mattys and colleagues

(Mattys, 2004; Mattys et al., 2005; Mattys & Melhorn, 2007), speech segmentation is largely the product of both the differentially weighting of the available cues and of the listening conditions.

Mattys and colleagues proposed a hierarchical model of speech segmentation in which the cues lowest weighted in the mature speech segmentation system would correspond to the ones earlier acquired. Fernandes et al. (2007) proposed that the weighting of these segmentation cues might also depend on their domain-generality and/or universality (vs. language-dependency). Both domain-general cues, like TPs, and universal prosodic cues, like Ips, have a fundamental role in segmentation since the very beginning of language acquisition (e.g., Dehaene-Lambertz & Houston, 1998; Thiessen & Saffran, 2003). On the contrary, language-dependent cues such as lexical stress depend on the particular language, and hence intervene later in speech segmentation. Therefore, while universal prosodic cues could maintain an important role in speech segmentation throughout linguistic development (and even in adulthood), the relative weighting of language--dependent prosodic cues probably depends on their word-boundary predictability in a specific language (cf. Mattys et al., 2005).

This proposal coheres with infants' studies. When TPs and lexical stress are incongruent, while 6.5-months-olds consider TPs as the most reliable cue (Thiessen & Saffran, 2003), at 9 months the weighting of these cues is already reversed (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). At this age, infants are also sensitive to the statistical information of their native language (i.e., phonotactics: cf. Mattys & Jusczyk, 2001), and by the age of 10.5 months, they are already able to integrate multiple sources of information, while language-dependent prosodic cues seem to start loosing their previous importance (Jusczyk et al., 1999). In other words, language-dependent cues, lower weighted in adulthood (Mattys et al., 2005), seem to have a predominant role only during a transitory phase in infancy. After that period, they gradually loose reliance and/or are supplanted by other cues with higher word-boundaries predictability.

In the present study, we used two ALL settings with adult listeners to investigate the relative weighting of TPs and either one of these two types of prosodic segmentation cues in several conditions of signal quality. In Experiment 1, we investigated the weighting of TPs and Ips by acoustically marking prosodic right-edges through syllable lengthening and falling pitch. Ips were not only as resilient to physical degradation of the signal as TPs, but seem also able to drive segmentation even when TPs were incongruent with them. This held true in any listening condition. In addition, when TPs and prosodic right-edges suggested the same segmentation hypotheses, a performance gain was observed.

This result adds to Shukla et al.'s (2007) proposal on the predominance of universal prosodic information over the statistical learning mechanism. Both domain-general and universal prosodic cues (here, Ips) are resilient to

physical noise, possibly on the grounds of their fundamental and precocious role in segmentation (Dehaene-Lambertz & Houston, 1998; Saffran et al., 1996a). Our data are thus partly in agreement with this proposal, but also show that the function of prosody is not only to allow or to suppress statistical segmentation outputs. Indeed, a "higher" (IP-based) filter would not explain the redundancy gain observed when TPs and universal prosodic cues suggested the same parsing in the congruent cues condition of Experiment 1.

In any case, the role of prosody in speech segmentation is rather different when, instead of Ips, the available prosodic cue is a language-dependent one, namely lexical stress, as it was the case in Experiment 2B. In this case, the availability of congruent stress cues with TPs did not lead to a redundancy gain. Had that been the case, we would have found the best performance in the condition in which TP-words were acoustically marked as paroxytones, which corresponds to the default stress pattern in Portuguese. In fact, three findings suggest that language-dependent prosody does not filter out statistical segmentation outputs (at least in Portuguese). First, with intact speech no impact of the stress pattern was found on ALL. Second, in line with Mattys and colleagues' proposal (Mattys, 2004; Mattys et al., 2005), stress pattern effects only emerged with degraded signal, and were maximized in the strongly degraded (10 dB SNR) condition. Third, when statistical and stress cues were both available in the stream and the strong degradation of the signal enabled stress cues to operate efficiently, the only units extracted from the stream were those highly compatible with both cues – paroxytone high-TP--words. Neither all statistical byproducts (high- and low-TP-words) nor all stress byproducts (TP-words for listeners in the 2nd syllable-stressed condition and part-words paroxytones for listeners in the 1st and 3rd syllable-stressed conditions) were treated as potential words of the AL, probably because these were not jointly supported by the two available cues.

It could be argued that the differential pattern of results found in Experiments 1 and 2B could be merely due to the between-experiments difference in the number and implementation of the acoustic correlates of prosody. In Experiment 1, two acoustic correlates were used (duration and pitch), and they were available in all TP-words embedded in the AL stream. In contrast, in Experiment 2B only one acoustic parameter was used (duration) and only some exemplars of TP-words were acoustically marked (as done with coarticulation cues by Fernandes et al., 2007).

It seems however improbable that the number and implementation of the acoustic cues can account by itself for the observed between-experiments difference. First, the manipulation of the acoustic parameter available in Experiment 2B was similar to that used by Fernandes et al. (2007) with coarticulation. Therefore, if the acoustic manipulation per se were responsible for the results found, the same (weak) impact found with lexical stress in Experiment 2B should have also been found with coarticulation by Fernandes et al. (2007). Quite on the opposite, in that study coarticulation was able to

drive speech segmentation with intact signal and enabled a redundancy gain when TPs suggested the same segmentation hypotheses. Second, although prosodic information tends to be multiparametric, the availability of different acoustic parameters does not necessarily produce cumulative effects in prosody-driven speech segmentation. For example, Diley and McAuley (2008) used duration only, F0 only, as well as both duration and F0 to signal distal prosodic boundaries. Although listeners' performance was less compatible with prosodic information when duration only was used, their performance was equivalent when prosody was marked either by F0 only or by F0 and duration. Similarly, Bagou et al. (2002) found no difference in ALL performance of French listeners exposed to an AL acoustically marked by either syllable lengthening, or F0 rise, or both acoustic cues. Moreover, if pitch and duration (in Experiment 1) and duration alone (in Experiment 2B) had been treated as correlates of the same prosodic information, a qualitatively similar ALL pattern should have been found in the two experiments. In other words, participants exposed to the 3rd stressed syllable condition in Experiment 2B (in which the third syllable of TP-words was marked, and hence corresponded in acoustic terms to the congruent-cues condition of Experiment 1) should have presented the best ALL performance, both in comparison to the other stressed conditions and to the TP-unstressed, single-cue condition.

This was obviously not the case. The overall ALL patterns found in Experiment 2B do not parallel those found in Experiment 1, and the impact of the prosodic cues in each experiment was quite different. In Experiment 1, not only listeners considered the (universal) prosodic cue (Ips) more reliable than TPs, but when Ips and TPs were congruent there was a redundancy gain. In Experiment 2B, the presence of a language-dependent prosodic cue only impacted on participants' performance in noisy conditions and narrowed the selection of statistical segmentation outputs to the ones with the strongest support; that is to the outputs supported by the conjunction of the two available segmentation cues. This led to the exclusive extraction of high-TP--words obeying to the default stress pattern (paroxytone) in Portuguese.

The present results thus suggest that universal and language-specific prosodic cues do have different roles in (Portuguese) speech segmentation. Also, they show that speech segmentation outputs are not simply the product of one preponderant prosodic cue acting as a filter over the outputs of another lower-weighted cue. Instead, a more parsimonious account would suggest that the outcome of speech segmentation is largely the result of the conjunction of the available cues on the grounds of their nature and weighted reliability. On the one hand, as long as a reliable cue is available in the speech stream (e.g., in intact speech, coarticulation: Fernandes et al., 2007; in any listening condition, universal prosodic cues like Ips: Shukla et al., 2007, and Experiment 1 of the present study), its byproducts are considered reliable potential words. If these units are also supported by a lower-weighted cue,

such as the domain-general TPs, a redundancy gain will probably be observed, thus leading to the optimization of speech segmentation processing. Note that this optimization does not necessarily correspond to the additive sum of the weight of the available cues (see e.g., Christiansen et al., 1998). On the other hand, if the available cues are weakly reliable in speech segmentation, only the segmentation outputs strongly supported by the conjunction of the several available cues will be considered as "word" units, and hence no redundancy gain will be observed. This seemed to have been the case in Experiment 2B with lexical stress, probably because in Portuguese stress does not have high word-boundary predictability (since the stressed syllable does not often indicate – either a right or a left- lexical edge in this language).

The idea that a cue that insures a deeper encoding of structural regularities of the input can enable reliance on more subtle aspects of the input for making correct predictions is not new (Christiansen et al., 1998; Christiansen & Curtin, 2005). Consequently, the integration of different cues does not necessarily promote a *quantity* gain (i.e., more units parsed from the stream). Instead, it can promote a *qualitative* gain (i.e., the correct parsing and deeper encoding of highly supported units), leading to the correct extraction of the byproducts of the conjunction (i.e., intersection) of the available sources of information. This would minimize errors and unwanted over-generalizations. Curtin et al. (2005; see also Creel et al., 2006) already suggested that lexical stress could benefit the learner by changing the representational landscape, providing a qualitative gain on information to be used in speech segmentation. The role of lexical stress could thus be one of reducing confusability in learning.

In sum, the nature of the available segmentation cues – their domain--generality and their role in specific languages – can differentially impact speech segmentation. Both domain-general and universal prosodic cues are highly resilient to physical degradation of the signal, with the latter occupying a preponderant role in segmentation over the former. The role of language--dependent prosodic segmentation cues is, on the contrary, highly dependent of listening conditions (Fernandes et al., 2007; Mattys et al., 2005). It thus seems that both the available sources of information and the listening conditions (cf. Mattys et al., 2005) do delineate the future of speech segmentation outputs.

References

Abaurre, M. B., & Galves, C. (1998). Rhythmic differences between European and Brazilian Portuguese: an optimalist and minimalist approach. *D.E.L.T.A., 14,* 377--403.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probabilities by 8-Month-Old Infants. *Psychological Science, 9*, 321-324.

Bagou, O., Fougeron, C., Fauenfelder, U. H. (2002). Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 59–62). Aix-en-Provence: Laboratoire Parole et Langage.

Böcker, K. B. E., Bastiaansen, M. C. M., Vroomen, J., Brunia, C. H. M., & de Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophisiology, 36,* 706-720.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbum, K. (2005). Mommy and Me: Familiar Names Help Launch Babies Into Speech-Stream Segmentation. *Psychological Science, 16,* 298-304.

Christiansen, M. H., Allen, J. A., Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes, 13,* 221-268.

Christiansen, M. H., & Curtin, S. (2005). Integrating multiple cues in language acquisition: A computational study of early infant speech segmentation. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 347-372). Hove, UK: Psychology Press.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: the role of prosody. *Journal of Phonetics, 31,* 585--598.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51,* 523-547

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 24-39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. *Psychological Science, 17,* 905-912.

Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of Lexical Stress on Learning an Artificial Lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 15-32.

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event--related brain potential study. *Brain Research,* 1123, 168-178.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition, 96,* 233--262.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech, 40,* 141-201.

D'Andrade, E., & Laks, B. (1996). Stress and Constituency: The Case of Portuguese. In J. Durant & B. Laks (Eds), *Current Trends in Phonology: Models and Methods, volume I.* (pp. 15-41). ESRI. Manchester: Universidade de Salford.

Cutler, A., & Norris, D. (1988). The role of Strong Syllables in Segmentation for Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113-121.

Dehaene-Lambertz, G., & Houston, D. (1998) Faster orientation latency toward native language in two-month-old infants. *Language and Speech, 41,* 21-43.

Delgado-Martins, M.R. (2002). *A Fonética do Português: Trinta anos de Investigação.* Editorial Caminho, Lisboa – Portugal.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*, 294--311.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion*? Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568-1578.

Dupoux, E., Pallier, C., Sebastian-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*, 406-421.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. *Proceedings of ICSLP'96*, Philadelphia, *3,* 1393-1396.

Fernandes, T., Ventura, P., & Kolinky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of quality of the signal. *Perception & Psychophysics, 69,* 856-864.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499-504.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cognitive Brain Research, 20,* 300-308.

Frota, S. & Vigário, M. (2001). On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus, 13,* 247-275.

Gomes, I., & Castro, S. L. (2003). PORLEX database in European Portuguese. *Psychologica, 32*, 91-108.

Grice, M. (2006). Intonation. In K. Brown (Ed.) *Encyclopedia of Language & Linguistics, second edition,* volume 5 (pp. 778-788). Oxford: Elsevier.

Grønnum, N. & Viana, M. C. (1999). Aspects of European Portuguese Intonation. In J. Ohala (Eds.) *Proceedings of the 14th International Congress of Phonetic Sciences, vol.3,* (pp. 1997-2000).

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78,* B53-B64.

Houston, D. M., Santelman, L. M., & Jusczyk, P. M. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19,* 97-136.

Iivonen, A., Niemi, T., & Paananen, M. (1998). Do F0 peaks coincide with lexical stresses? In S. Werner (Ed.), *Nordic prosody: Proceedings of the VIIth conference, Joensuu 1996* (pp. 141–158). Frankfurt am Main: Peter Lang.

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language, 44,* 548-567.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginning of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159–207.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence of a general learning mechanism. *Cognition, 83,* B35-B42.

Klatt, D. H. (1980). Speech perception: A model of acoustic- phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, N.J.: Erlbaum.

Kolinsky, R. Cuvelier, H., Goetry, V., Peretz, I., & Morais, J. (2009). Music training facilitates lexical stress processing. *Music Perception, 26*, 235-246

Liberman, A. M. Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. Leibowitz H.-L. Teuber (Eds.), *Handbook of sensory physiology: Perception* (VIII, 143-178). Berlin: Springer-Verlag.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62,* 253-265.

_____ (2004). Stress Versus Coarticulation: Toward an Integrated Approach to Explicit Speech Segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 397-408.

Mattys, S. L., Jusczyck, P. W. (2001). Phonotactic cues for segmentation on fluent speech by infants. *Cognition, 78,* 91-121.

Mattys, S. L., & Melhorn, J., F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America, 122,* 554-567.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477-500.

Mateus, M. H., & d'Andrade, E. (2000). *The Phonology of Portuguese.* Oxford, UK: Oxford University Press.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro S.-L., & Besson, M. (2009). Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cerebral Cortex, 19*, 712-723.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39,* 21-46.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in Spoken Word Recognition. Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621-638.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41,* 233–243.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in Spoken-word Recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1209-1228.

Ortega-Llebaria, M. (2006). Phonetic Cues to Stress and Accent in Spanish. In M. Díaz-Campos (Ed.), *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology* (pp 104-118). Cascadilla Press.

Peretz, I., & Hyde, K.L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7*, 362-367.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2*, 85-115.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 349-351.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical Learning by 8--Month-Old Infants. *Science, 274,* 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35,* 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental Language Learning: Listening (and Learning) out of the Corner of Your Ear. *Psychological Sicence, 8*, 101-105.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime references guide.* Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime user's guide.* Pittsburgh: Psychology Software Tools Inc.

Seidl, A. (2007) Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57,* 24–48.

Seidl, A. & Johnson, E. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9,* 565-573.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54,* 1-32.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50,* 86-132.

Thiessen, E. D., & Saffran, J. R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month Old Infants. *Developmental Psychology, 39*, 706-716.

Toro, J. M., Rodríguez-Fornells, A., Sebastián-Gallés, N. (2007). Stress placement and word segmentation by Spanish speakers. *Psicológica: Revista de metodología y psicología experimental, 28,* 167-176.

Toro, J. M., Sinett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97,* B25-B34.

Valian, V. & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. *Journal of Memory and Language*, *35*, 497-516.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The Roles of Word Stress and Vowel Harmony in Speech Segmentation. *Journal of Memory and Language, 38,* 133-149.

Werner, S., & Keller, E. (1994). Prosodic aspects of speech. In E. Keller (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges* (pp. 23-40). Chichester: John Wiley.

Acknowledgements

Correspondence concerning this article should be sent to Tânia Fernandes, Speech Lab – Lab. De Fala, R. do Dr. Manuel Pereira da Silva, 4200-392 Porto, Portugal. E-mail: tfernandes@fpce.up.pt. Preparation of this article was supported by a grant of Fundação para a Ciência e a Tecnologia – Ministério da Ciência, Tecnologia e Ensino Superior to T. Fernandes, ref SFRH / BPD / 46979 / 2008, as well as by a grant of Fundação para a Ciência e a Tecnologia

– Ministério da Ciência, Tecnologia e Ensino Superior – and European Community FEDER funding (Project PTDC/PSI/66077/2006, "Cognitive consequences of literacy"). Preparation of this article was also supported by Centro de Psicologia Clínica e Experimental – Desenvolvimento, Cognição e Personalidade of the Universidade de Lisboa, Portugal.

The third author is Senior Research Associate of the Fonds de la Recherche Scientifique- FNRS (Belgium).

Tânia Fernandes
Laboratório de Fala – Speech Lab
Universidade do Porto
Rua Dr. Manuel Pereira da Silva
4200-392 Porto, Portugal
tfernandes@fpce.up.pt

Paulo Ventura
Faculdade de Psicologia
Universidade de Lisboa
Alameda da Universidade
1649-013 Lisboa, Portugal
pvfv@fp.ul.pt

Régine Kolinsky
UNESCOG – Université Libre de Bruxelles, Belgium
Fonds de la Recherche Scientifique – FNRS
Av. Franklin Roosevelt, 50 B
1050 Brussels, Belgium
rkolins@ulb.ac.be