



ELSEVIER

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Perception of predictable stress: A cross-linguistic investigation

Sharon Peperkamp*, Inga Vendelin, Emmanuel Dupoux

Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole des Hautes Etudes en Sciences Sociales, Département d'Etudes Cognitives-Ecole Normale Supérieure, Centre National de la Recherche Scientifique, 29, Rue d'Ulm, 75005 Paris, France

ARTICLE INFO

Article history:

Received 26 August 2009

Received in revised form

25 March 2010

Accepted 13 April 2010

ABSTRACT

Previous studies have documented that speakers of French, a language with predictable stress, have difficulty distinguishing nonsense words that vary in stress position solely (stress “deafness”). In a sequence recall task with adult speakers of five languages with predictable stress (Standard French, Southeastern French, Finnish, Hungarian and Polish) and one language with non-predictable stress (Spanish), it was found that speakers of all languages with predictable stress except Polish exhibited a strong stress “deafness”, while Spanish speakers exhibited no such “deafness”. Polish speakers yielded an intermediate pattern of results: they exhibited a weak stress “deafness”. These findings are discussed in light of current theoretical models of speech perception.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

It is well-known that adult listeners experience difficulty in perceiving phonological contrasts that are not used in their native language. For instance, Japanese has only one liquid consonant, /r/, and Japanese listeners have difficulty perceiving the contrast between the English liquids /r/ and /l/ (Goto, 1971; Miyawaki et al., 1975), even if they are learners of English (Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Logan, Lively, & Pisoni, 1991; Takagi & Mann, 1995). Similar difficulties have been reported for other cases of non-native consonantal and vocalic contrasts (Pallier, Bosch, & Sebastián-Gallés, 1997; Werker & Tees, 1984a). Moreover, they are known to emerge during the first year of life (Werker & Tees, 1984b; Werker & Tees, 1999). From a theoretical viewpoint, difficulties with the perception of non-native contrasts shed light on an important question: how does linguistic knowledge impact on our speech perception system?

Two types of accounts have been proposed regarding language-specific effects in the perceptual processing of consonants and vowels. We call them functional role accounts and lexical statistics accounts, respectively. In functional role accounts, dimensions that play a functional role in the language are amplified, and dimensions that are non-functional are attenuated, resulting in good perception of the former and poor perception of the latter. In several speech perception models, dimensions are expressed either in terms of acoustic or phonetic cues (Francis & Nusbaum, 2002; Iverson et al., 2003; Jusczyk, 1997) or in terms of phonological features (Brown, 1998, 2000; Flege, 1995; Lado,

1957). Lexical statistics accounts, by contrast, posit that the phonological grammar emerges from generalizations about phonological regularities across the lexicon. In particular, the more regular or predictable a pattern is, the less it needs to be specified in the lexical representation. For instance, if vowel nasality is entirely predictable, as in English where nasal vowels only occur before nasal consonants, it does not need to be specified lexically; similarly for predictable stress. Two types of lexical statistics accounts are exemplar-based models (Pierrehumbert, 2001, 2003) and connectionist models (Gaskell, 2003; Gaskell, Hare, & Marslen-Wilson, 1995).

Suprasegmental contrasts (i.e., contrasts concerning tone, length, stress, and pitch accent) likewise give rise to perceptual difficulties in speakers of languages that lack such contrasts. For instance, English speakers have difficulty perceiving the tonal contrasts of Mandarin (Gottfried & Suiter, 1997; Hallé, Chang, & Best, 2004; Kiriloff, 1969), and French speakers have difficulty perceiving the vowel length contrast of Japanese (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999). French speakers also have difficulty perceiving the stress contrast of Spanish (Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001), even if they are advanced learners of Spanish (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). These difficulties with perceiving suprasegmental contrasts, like those concerning segmental contrasts, arise early: At nine months of age, English-learning infants already differ from Chinese ones and fail to discriminate a tonal contrast (Mattock & Burnham, 2006). Likewise, French-learning infants of the same age differ from Spanish-learning infants and fail to discriminate a stress contrast (Skoruppa et al., 2009).

Unlike in the case of segmental contrasts, little is known about the mechanisms that yield difficulties perceiving suprasegmental

* Corresponding author. Tel.: +33 1 44 32 26 24; fax: +33 1 44 32 26 30.
E-mail address: sharon.peperkamp@ens.fr (S. Peperkamp).

contrasts. In this article, we address this issue in a cross-linguistic experiment concerning the perception of stress. Following our previous work on this issue (Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux et al., 2001, 2008; Peperkamp & Dupoux, 2002), we use the shorthand stress “deafness” to refer to the difficulty with the perception of stress at a phonological level.

Besides speakers of French, we test speakers of four more languages with predictable stress, i.e. Southeastern French (which – as we will show below – differs from Standard French in the distribution of stressed and unstressed syllables), Finnish, Hungarian, and Polish. As in our previous work, our control population consists of speakers of Spanish, whose language has contrastive stress, as evidenced by minimal pairs such as /'bebe/ “(s/he) drinks” – /be'be/ “baby”. Our aim is to identify which factor or factors govern(s) the emergence of stress “deafness”. We begin by reviewing the predictions of functional role accounts and lexical statistics accounts for the perception of stress.

As indicated above, in functional role accounts the weighting of acoustic/phonetic cues or phonological features can be reduced or enhanced, depending on their functional roles in the language (Brown, 1998, 2000; Flege, 1995; Francis & Nusbaum, 2002; Iverson et al., 2003; Jusczyk, 1997; Lado, 1957). It should be noted, however, that contrary to segments, suprasegmentals may have functional roles at more than just the lexical level. For instance, stress is instantiated by three phonetic cues: duration, F0 and intensity (Lehiste, 1970). In addition to lexical stress, these cues can signal prosodic constituent boundaries and grammatical functions (interrogative versus affirmative mode, focus, etc.). These prosodic and grammatical levels should clearly not be included in the definition of functional role, since in French, duration, F0 and intensity are known to signal boundaries and to serve grammatical functions; hence, it is only by restricting the definition of functional role to the lexical level that we can account for the reduction of sensitivity to stress cues in French speakers. Under this definition of functional role, two factors can be identified that might play a role in the emergence of stress “deafness”:

The first factor concerns the domain of stress assignment. Stress is typically a word-level property, in that all content words have one main stress. French (both its standard and its Southeastern variety) is an exception, in that stress is argued to be assigned at the phrase level (Grammont, 1965). Perhaps this is what induces stress “deafness” in French as opposed to Spanish speakers: they can completely ignore stress for the purposes of word recognition. Consequently, we would expect speakers of Finnish, Hungarian, and Polish, which all have a word stress rule, to be like speakers of Spanish and exhibit no stress “deafness”, while speakers of Southeastern French should be like speakers of Standard French and exhibit stress “deafness”.

The second factor concerns the lexical use of one or more phonetic correlates of stress. Duration and F0 are not only phonetic correlates of stress but are also used for other suprasegmentals: duration is the phonetic correlate of contrastive length, while F0 is the correlate of tone and pitch accent. Among our test languages with predictable stress, Finnish and Hungarian have contrastive vowel length (Karlsson, 1999; Tompa, 1972). Thus, Finnish and Hungarian speakers use duration lexically. French and Polish speakers, by contrast, do not use any of the stress cues lexically, since their languages do not make lexical use of length, tone, or pitch accent. Maybe the perception of stress correlates with the lexical use of stress cues in the listener’s native language. We should then obtain three levels of stress perception: Standard French, Southeastern French, and Polish speakers should exhibit a strong “deafness” (no lexical use of stress cues at all), Finnish and Hungarian speakers a weak “deafness” (lexical use of duration for contrastive vowel length), and Spanish speakers no

stress “deafness” at all (lexical use of duration, F0 and intensity for contrastive stress). An additional prediction is that when duration is not part of the cues by which stress is realized, the performance of Finnish, Hungarian, and Spanish speakers should decrease more than that of Standard French, Southeastern French and Polish speakers. Indeed, whereas all populations should have more difficulties when one of the phonetic correlates is absent, this disadvantage should be larger for those who make lexical use of the missing correlate in their native language.

Concerning lexical statistics accounts, we mentioned above that they focus on the existence of phonological distributional regularities across lexical items, with more regular patterns being encoded less precisely (Gaskell, 2003; Gaskell et al., 1995; Pierrehumbert, 2001, 2003). For these accounts, two different factors can be identified that might play a role in the emergence of stress “deafness”.

First, languages differ with respect to the variability in the position of stress. In Standard French, stress surfaces at a fixed position within its phrasal domain: stress always falls on the phrase-final syllable (even when the phrase ends in a function word). This is where Standard French diverges from Southeastern French. In Southeastern French, phrases can end in the vowel schwa, which surfaces unstressed (Durand, 1976). Consequently, stress is not on a fixed position in Southeastern French: it is final unless the final vowel is schwa, in which case it is penultimate. As to the other languages with predictable stress, stress is assigned within words. Finnish and Hungarian have stress on a fixed position within this domain, namely the initial syllable (Karlsson, 1999; Tompa, 1972). In Polish, by contrast, stress is not on a fixed position. Specifically, stress is regularly assigned to the word’s penultimate syllable (Comrie, 1967), but not all words end in a sequence of a stressed syllable followed by an unstressed one. This is because Polish has a large number of monosyllabic content words, which are stressed too. Finally, although Spanish has contrastive stress, the position of stress is confined to one of the word’s last three syllables (Harris, 1983). Maybe it is the complete absence of variability in the position of stress that yields stress “deafness” in speakers of Standard French. In that case, we expect the same result for speakers of Finnish and Hungarian, but less difficulties with the perception of stress in speakers of Southeastern French and Polish. The position of stress being the most variable in Spanish, speakers of this language should, moreover, outperform speakers of all other languages.

Second, languages differ with respect to the presence of lexical exceptions. Indeed, some languages with predictable stress have lexical exceptions to their stress rule, others do not. French is an example of the latter: all utterances end in a stressed syllable, even if the last word is a loanword. Southeastern French, Finnish and Hungarian are also exceptionless. Polish, however, does have a small number of exceptional words with irregular stress on either the antepenultimate or the final syllable (Comrie, 1967). These words are most often words of Greek or Latin origin (*gim'nastyka*) or foreign proper names (*'Everest*).¹ Finally, even though Spanish is described as having contrastive stress, it has a default stress pattern: in polysyllables, stress tends to fall on the penultimate syllable if the word ends in /n/, /s/ or a vowel, and on the last syllable otherwise (Harris, 1983). In the Spanish database of Santiago, Justicia, Palma, Huertas, and Gutiérrez (1996), 83% of the lexemes obey this default pattern (we consider monosyllabic content words to have regular stress, regardless of their final

¹ We conducted a search for the 30 commonest exceptional roots in a corpus of spontaneous speech in Polish. This corpus contained 436,972 words spoken by adults, extracted from the Polish CHILDES corpora “Szuman” (Smoczyńska, 1985) and “Weist” (Weist, Wysocka, Witkowska-Stadnik, Buczowska, & Konieczna, 1984). This search revealed that around 0.1% of the words have exceptional stress.

Table 1
Stress-related properties of six languages.

| Language | Domain of stress | Lexical use of stress cues | Variability in stress position | Lexical exceptions (%) |
|---------------------|------------------|----------------------------|--------------------------------|------------------------|
| Standard French | Phrase | None | No ^a | 0 |
| Southeastern French | Phrase | None | Moderate ^b | 0 |
| Finnish | Word | Duration | No ^c | 0 |
| Hungarian | Word | Duration | No ^c | 0 |
| Polish | Word | None | Moderate ^d | 0.1 |
| Spanish | Word | Duration, F0 and intensity | High ^e | 17 |

^a Final.

^b Last non-schwa syllable.

^c Initial.

^d Penultimate in polysyllables, and on the only syllable of monosyllables.

^e One of the last three syllables.

Table 2
Predictions stemming from four factors for the perception of stress (“<” means worse performance).

| | |
|--|---|
| Factor 1. Domain of stress | Standard French, Southeastern French < Finnish, Hungarian, Polish, Spanish |
| Factor 2. Lexical use of stress cues | Standard French, Southeastern French, Polish < Finnish, Hungarian < Spanish |
| Factor 3. Variability in stress position | Standard French, Finnish, Hungarian < Southeastern French, Polish < Spanish |
| Factor 4. Presence of lexical exceptions | Standard French, Southeastern French, Finnish, Hungarian < Polish < Spanish |

segment). Hence, we might re-describe Spanish as having predictable stress with 17% of lexical exceptions. Obviously, the more numerous the lexical exceptions are, the more advantageous it is to reliably encode stress in the phonological representation. This yields the prediction that there are three levels of performance as far as stress perception is concerned: a strong “deafness” should be found in speakers of Standard French, Southeastern French, Finnish and Hungarian, a weak “deafness” in Polish speakers, and no stress “deafness” in Spanish speakers.

To summarize, our six test languages vary with respect to four factors that might influence stress perception (see Table 1): (1) the domain of stress; (2) the lexical use of stress cues (for contrastive stress or some other suprasegmental); (3) the variability in stress position; and (4) the existence of lexical exceptions.

A summary of the predictions is shown in Table 2. It should be noted that no two factors make the same predictions.

Stress “deafness” in French speakers has been observed both in an ABX discrimination task (Dupoux et al., 1997) and in a sequence recall task (Dupoux et al., 2001). In order to assess the perception of stress by speakers of the six languages in Table 1, we use the sequence recall task, which has been shown to provide a more robust indicator of stress “deafness”. Indeed, it was found to yield a complete separation between French and Spanish monolinguals (Dupoux et al., 2001). Moreover, it is largely insensitive to experience effects, such as the late acquisition of Spanish by French speakers (Dupoux et al., 2008). In this paradigm, participants have to recall sequences of two auditorily presented non-words that differ either in the position of stress or in the middle consonant, with the various tokens of each non-word being phonetically different. The combination of a memory load and phonetic variability ensures that this task taps an abstract phonological processing level, and cannot be performed using a low-level acoustic strategy.

In Dupoux et al. (2001), speakers of Standard French but not those of Spanish were found to exhibit stress “deafness”: they had

severe difficulty perceiving the stress contrast compared to the consonantal contrast. The same paradigm was used by Peperkamp and Dupoux (2002) in two experiments testing, respectively, (Standard) French, Finnish and Spanish (Experiment 1), and Hungarian and Polish (Experiment 2). However, this study did not allow for firm conclusions regarding the effect of the four factors outlined above, for the following reasons. As shown in Table 2, Factors 1 and 2, unlike Factors 3 and 4, predict a stronger stress “deafness” in Standard French speakers than in Finnish and Hungarian speakers. Peperkamp and Dupoux reported an inconclusive, marginal, difference in the expected direction between Standard French and Finnish speakers, and a significant difference – also in the expected direction – between French and Hungarian speakers. However, the latter difference was difficult to interpret, due to the fact that different sets of stimuli and slightly different procedures were used to test the French and Hungarian groups; most importantly, the strength of the durational cue and the length of the test sequences differed. Table 2 also shows that Factors 2, 3, and 4, unlike Factor 1, predict a difference between Polish and Spanish speakers. Peperkamp and Dupoux found that Spanish speakers perceive stress better than Polish speakers, but again, this effect was only marginally significant and the stimuli and procedure used to test the Polish and the Spanish groups were different.

In the present study, we report on a novel perception experiment with the sequence recall task, which uses a single set of stimuli across all languages. To test the effect of the presence versus absence of the durational cue (see Table 2, Factor 2), there are two versions of the stress contrast: in one version, stress is instantiated by all three stress cues (duration, F0, and intensity); in the other one, the durational cue is removed.

2. Experiment

We use a variant of the sequence recall task described in Dupoux et al. (2001). The experiment is divided into three parts. In each part, participants are required to learn two CVCV non-words that are a minimal pair differing only in one phonological characteristic: place of articulation of the middle consonant in the first part, and location of stress in the second and third part. In each part, participants are taught to associate the two non-words to the keys [1] and [2], respectively, of a computer keyboard. After some training with an identification task, participants listen to random sequences consisting of the two items that they are required to recall and transcribe as sequences of [1] and [2]. The phonemic contrast in the first part is meant to be equally easy for speakers of all test languages, and is used to establish baseline performance. The second and the third part with the stress

contrasts use identical items, but different tokens: in one of them stress is instantiated by duration, F0 and intensity, in the other one it is instantiated by F0 and intensity only.

2.1. Method

2.1.1. Materials

Two minimal pairs were constructed, one involving a phonemic contrast, i.e. /'muku/ - /'munu/, and the other one involving a stress contrast, i.e. /'numi/ - /nu'mi/. None of the items is a real word in French, Finnish, Hungarian, Polish or Spanish, but all the segments making up the items are phonemic and appear in a phonotactically legal combination in these languages.

The items were recorded 10 times each by a trained female phonetician who is a native speaker of Dutch. She realized the stress contrast using the three acoustic cues, duration, F0 and intensity in an exaggerated manner (for details, see below), such that it would be easy to discriminate at an acoustic processing level. Six recordings of each item were selected, leading to six pairs of tokens with the phonemic contrast and six pairs of tokens with the stress contrast. Their mean durations were 432 and 570 ms, respectively. In addition, the word "OK" was recorded once by a male talker. All recorded items were digitized at 16 kHz at 16 bits, digitally edited, and stored on a computer disk. For each item, the global pitch of the six tokens was multiplied with the

values .95, .97, .99, 1.01, 1.03, and 1.05, respectively. This modification was applied in order to introduce more phonetic variation into the stimuli (see Dupoux et al. (2001) for discussion). Furthermore, in order to reduce the possibility of using a low-level acoustic strategy (as opposed to a more abstract phonological one), the stimuli were time-compressed, making it hard to consciously track acoustic properties of the tokens and mentally translate the sequences in terms of the number keys [1] and [2] while listening to them. Specifically, the stimuli with the phonemic contrast were compressed to 65% of the original duration, yielding a mean duration of 281 ms, and those with the stress contrast were compressed to 50%, yielding a mean duration of 285 ms. The compression thus resulted in about equal durations for the stimuli with the phonemic and the stress contrast.

In the stimuli with the stress contrast, stressed vowels were on average 12.4 ms longer than unstressed ones ($F(1,11)=5.691$, $p < .036$). The difference between the durations of stressed and unstressed vowels was highly significant in initial syllables (19.9 ms, $F(1,11)=34.9$, $p < .0001$), but not significant in final syllables (4.9 ms, $F(1,11)=3.6$, $.05 < p < .1$) (Fig. 1a). Stressed and unstressed vowels also had different F0 contours and intensity values (Fig. 1b and c). Specifically, stressed vowels were on average 86.5 Hz higher than unstressed ones, corresponding to a difference of 7 semitones ($F(1,11)=116.5$, $p < .0001$), and they were on average 5.5 dB louder ($F(1,11)=71.4$, $p < .0001$).

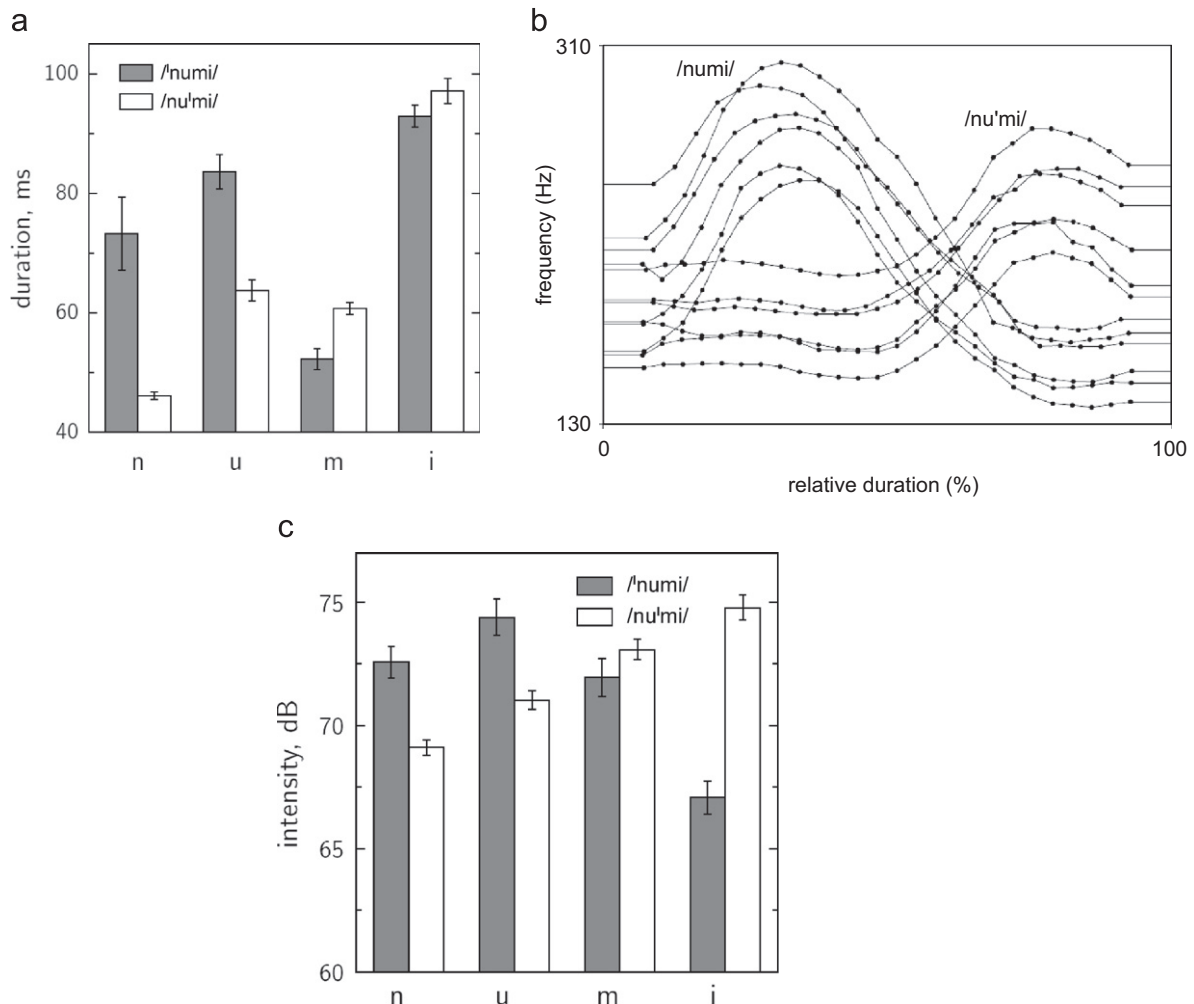


Fig. 1. Acoustic measurements of six tokens /'numi/ and six tokens /nu'mi/: duration of individual segments (a), F0 of tokens as a function of normalized duration (b), and intensity of individual segments (c).

On the basis of this second set, a third set of stimuli was created in which the durational differences of the stress-initial versus stress-final stimuli were neutralized and only F0 and intensity were kept as stress cues. To this effect, each one of the stress-initial tokens was randomly paired with one of the stress-final ones. In each pair of tokens, the durations of the individual syllables were then changed, such that the duration of the initial syllables was the mean of the durations of the original two initial syllables, and that of the final syllables was, likewise, the mean of the original two final syllables. As a result, both the initial and the final syllables within each stress pair had equal durations. (For instance, suppose that in a given pair, the syllable durations of /'numi/ were 340 and 300 ms, respectively, and those of /nu'mi/ were 300 and 500 ms, respectively; then in the modified stimuli the syllable durations were 320 and 400 ms, respectively, in both /'numi/ and /nu'mi/.)

Finally, for each minimal pair, we constructed four sequences composed of two items each and 20 sequences composed of five items each (see Appendix).

2.1.2. Procedure

The experiment consisted of three parts. During the first part, participants were tested on the phonemic contrast, and during the second and the third parts on each of the two versions of the stress contrast. The order of the latter parts was counterbalanced across participants. That is, half of the participants first listened to tokens with the durational cue and then to those without, while for the other half of the participants the order was the reverse. For the phonemic contrast, the item /'muku/ was associated to the number key [1], while its counterpart /'munu/ was associated to key [2]; for the stress contrasts, the item with stress on the first syllable, /'numi/, was associated to key [1], while its counterpart with stress on the second syllable, /nu'mi/, was associated to key [2].

During each part, participants were told that they were going to learn two words in a foreign language. They were first asked to press the number key [1], upon which they heard all tokens of the first item. They were then asked to press the number key [2], upon which they heard all tokens of the second item. Subsequently, participants could continue listening to the various tokens of the two items by pressing the associated keys; when they pressed one of the two number keys [1] or [2], the corresponding item was played. Participants could thus hear as many tokens of the two items as they wished. After they had indicated having learned the distinction between the two items, they heard a token of one of the items and had to press the associated key, [1] or [2]. A message on the screen informed them whether the response was correct. The message was “OK!” or “ERROR!”, and was displayed for 800 ms. This procedure was repeated until seven correct responses in a row had been given. Items were chosen semi-randomly, in that the maximum number of tokens of the same item played in a row was four. After having passed the success criterion, indicating that they had understood the task and memorized the association between the items and the number keys, participants turned to the main part of the experiment.

During the main part, participants first listened to a warm-up block of four sequences consisting of two repetitions of the two items, and then to a test block of 20 sequences consisting of five repetitions of the two items. Their task was to reproduce each sequence by typing the associated keys in the correct order. For each participant, the order of the sequences within the two blocks was randomized. Within each sequence, the items were randomly instantiated by one of the six tokens, with the proviso that a single token could not appear more than once. Each trial consisted of a

sequence in which tokens were separated by an 80 ms silence, and ended by the word “OK”. Participants could not begin typing their response until they had heard this word; they were thus prevented from using echoic memory (Morton, Crowder, & Prussin, 1971). A 1500 ms pause separated each response from the next trial. Participants were warned whenever they entered a sequence with a length that did not correspond to the length of the input string and asked to enter their reply again. In the warm-up block, participants received feedback as to whether their responses were correct or incorrect. In case of an incorrect response, the sequence was repeated until the correct response was provided. In the test block, participants received no feedback.

On average, the entire experiment lasted about 20 min. Responses were recorded on a computer disk and classified as follows. Responses that were a 100% correct transcription of the input sequence were coded as correct; all other responses were coded as incorrect. Among the incorrect responses, those that were a 100% incorrect transcription – i.e. with each token of the sequence labeled incorrectly – were coded as reversals. Participants with more reversals than correct responses in one or more of the three conditions were rejected, the high percentage of reversals suggesting that they might have confused the number key associated to the first item with the one associated to the second item.

2.1.3. Participants

Twelve native speakers of Standard French, Southeastern French, Finnish, Hungarian, Polish, and Spanish participated in the experiment. The Standard French speakers were aged between 23 and 28 (mean: 25), the Southeastern French between 18 and 28 (mean: 22), the Finnish between 19 and 53 (mean: 29), the Hungarian between 20 and 28 (mean: 24), the Polish between 21 and 31 (mean: 25), and the Spanish between 18 and 44 (mean: 31). Most of the participants of the languages with predictable stress had some basic knowledge – learned in school – of one or more foreign languages that use stress contrastively (e.g. English, German, or Russian). None of them had started to learn a foreign language before age 10, and none of them had a known hearing deficit.

One of the Finnish speakers was tested in Helsinki, and all but one of the Southeastern French speakers were tested in Marseilles; all other participants were tested in Paris. None of the non-French participants tested in Paris reported being French-dominant.

A number of additional participants were tested and excluded from the results on the basis of the rejection criterion defined above. This concerned six Standard French participants (too many reversals in one of the stress conditions), eight Southeastern French participants (phoneme: $N=2$; stress: $N=6$), six Finnish participants (phoneme: $N=3$; stress: $N=2$; both: $N=1$), three Hungarian participants (phoneme: $N=2$; stress: $N=1$), five Polish participants (phoneme: $N=1$; stress: $N=4$), and four Spanish participants (phoneme: $N=2$; stress: $N=2$).

2.2. Results and discussion

Mean error rates for the six groups of participants for the phoneme contrast and the two versions of the stress contrast are shown in Table 3.

For the analyses, we use the mean of the errors rates for the two versions of the stress contrast. Henceforth, “the error rate for the stress contrast” is shorthand for “the mean of the error rates for the two versions of the stress contrast”.

We submitted the error rates for the phoneme contrast and the stress contrast to an ANOVA with the within-participant factor

Contrast (phoneme vs. stress) and the between-participant factor Language (Standard French vs. Southeastern French vs. Finnish vs. Hungarian vs. Polish vs. Spanish). There was a main effect of Contrast ($F(1,66)=229.5, p < .0001$), with participants making more errors in the stress condition than in the phoneme condition, and a main effect of Language ($F(5,66)=3.6, p < .007$). The interaction between these two factors was also significant ($F(5,66)=12.6, p < .0001$). This interaction was due to the fact that there was an effect of Contrast for all participants except the Spanish ones (Standard French: $F(1,11)=227.2, p < .0001$; Southeastern French: $F(1,11)=127.9, p < .0001$; Finnish: $F(1,11)=66.4, p < .0001$; Hungarian: $F(1,11)=77.3, p < .0001$; Polish: $F(1,11)=8.0, p < .02$; Spanish: $F < 1$).

Next, we defined an individual stress “deafness” index as the difference between the error rate for the stress contrast and the error rate for the phoneme contrast. Boxplots of the stress “deafness” indices of the Standard French, Southeastern French, Finnish, Hungarian, Polish, and Spanish groups are shown in Fig. 2.

In order to quantitatively evaluate the differences in distribution of “deafness” indices among the language groups, we carried

out a series of pairwise one-way ANOVAs on the individual “deafness” indices. The results are shown in Table 4.

These results show that the distribution of individual “deafness” indices among the Spanish participants was significantly different from those of the other participant groups. The Polish group, furthermore, differed from the Standard French, Southeastern French, Finnish, and Hungarian groups. No other comparison reached significance.

In order to test the effect of the absence versus presence of the durational cue on the perception of stress, we submitted the error rates for the two versions of the stress contrasts to an ANOVA with the within-participant factor Contrast (stress-with-duration vs. stress-without-duration) and the between-participant factors Language-type (durational-cue vs. no-durational-cue). This latter factor divided the participants into two groups, according to whether their language uses duration lexically for either stress or length (Finnish, Hungarian, Spanish) or does not make lexical use of duration at all (Standard French, Southeastern French, Polish). There was a main effect of Contrast only ($F(1,70)=4.2, p < .05$), with participants making fewer errors in the condition stress-with-duration than in the condition stress-without-duration.

Table 3

Mean error rates and standard errors for the Standard French, Southeastern French, Finnish, Hungarian, Polish, and Spanish participants for the phoneme contrast and the two versions of the stress contrast.

| | Phoneme | | Stress-with-duration | | Stress-without-duration | |
|-----------------|---------|-----|----------------------|-----|-------------------------|-----|
| | % error | SE | % error | SE | % error | SE |
| Standard French | 34.2 | 3.9 | 77.5 | 4.4 | 84.6 | 2.2 |
| SE French | 45.8 | 4.3 | 85.4 | 2.0 | 86.7 | 2.3 |
| Finnish | 49.6 | 5.3 | 90.0 | 2.3 | 87.1 | 2.2 |
| Hungarian | 45.4 | 6.0 | 80.4 | 4.3 | 85.4 | 3.9 |
| Polish | 42.5 | 6.4 | 59.2 | 7.9 | 65.4 | 8.0 |
| Spanish | 47.5 | 5.6 | 47.1 | 5.4 | 48.8 | 7.4 |

Table 4

F- and p-values for pairwise one-way ANOVAs on the individual “deafness” indices across the Standard French, Southeastern French, Finnish, Hungarian, Polish, and Spanish groups.

| | St. French | SE-French | Finnish | Hungarian | Polish |
|-----------|------------|-----------|---------|-----------|--------|
| Spanish | | | | | |
| $F(1,22)$ | 48.2 | 10.0 | 25.6 | 25.8 | 4.5 |
| $p <$ | .0001 | .006 | .0001 | .0001 | .05 |
| Polish | | | | | |
| $F(1,22)$ | 12.6 | 6.8 | 5.1 | 4.7 | – |
| $p <$ | .002 | .02 | .03 | .04 | – |
| Hungarian | n.s. | n.s. | n.s. | – | – |
| Finnish | n.s. | n.s. | – | – | – |
| SE-French | n.s. | – | – | – | – |

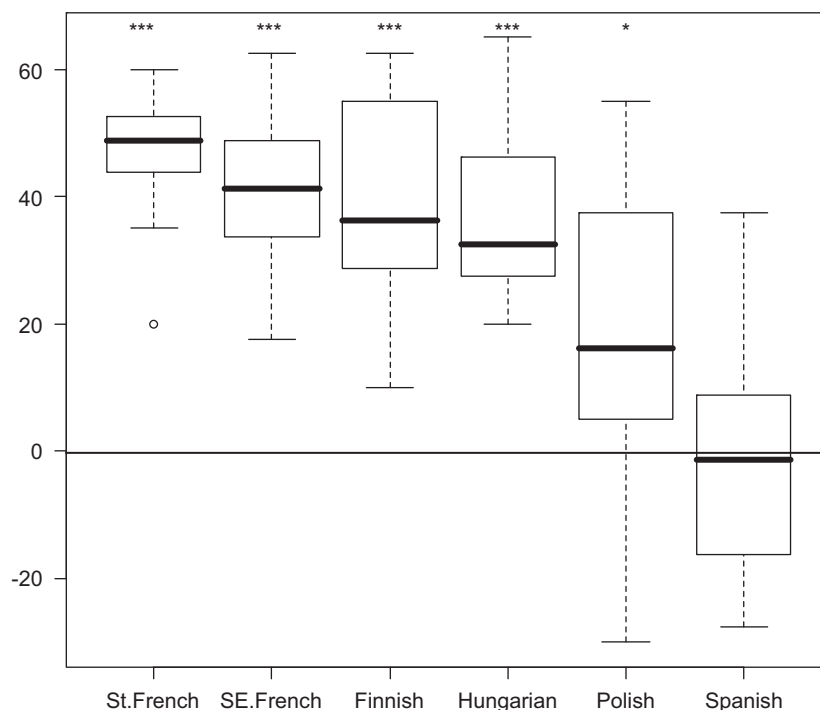


Fig. 2. Boxplots of the stress “deafness” indices of the Standard French, Southeastern French, Finnish, Hungarian, Polish, and Spanish groups. * $p < .02$; *** $p < .0001$.

Finally, we assessed the predictive power of the different factors listed in Table 1 by means of a series of regressions on the individual “deafness” indexes of the different participant groups. Five binary regressors were used, as follows. *Domain of stress* was modeled with a variable coded as 0 for languages with phrasal stress, and 1 for languages with word stress. *Lexical use of stress cues* was coded as 0 for languages that use none of the stress cues, and 1 for the others. *Variability in stress position* was coded as 0 for languages with no variability in stress position and 1 for the others. *Lexical exceptions* was coded as 0 for languages with no lexical exceptions at all and 1 for the others. As can be seen in Table 1, all factors except *Domain of stress* actually divide the languages into three rather than two groups, with Spanish forming in each case a singleton group. A fifth binary regressor was therefore introduced to distinguish Spanish – the only language with contrastive stress – from the other languages; this variable, *Contrastive stress*, was coded as 1 for Spanish and 0 for the remaining languages. The individual regressions, Bonferroni-corrected for multiple comparison, showed a significant correlation for *Variability in stress position* ($R=.34$, $F(1,70)=9.3$, corrected $p < .02$), *Lexical exceptions* ($R=.53$, $F(1,70)=27.3$, corrected $p < .0005$), and *Contrastive stress* ($R=.44$, $F(1,70)=16.7$, corrected $p < .0005$), while no significant correlation was obtained for *Domain of stress* ($R=.18$, $F(1,70)=2.4$, uncorrected $p > .1$) and *Lexical use of stress cues* ($R=0$, $F(1,70) < 1$, uncorrected $p > .1$). A multiple step-wise regression analysis rejected all variables except one: *Lexical exceptions*. Thus, once lexical exceptions are taken into account the other variables do not add any more information.

3. General discussion

Using a short-term memory sequence recall task, we found that speakers of four languages with predictable stress, namely, Standard French, Southeastern French, Finnish and Hungarian, exhibit a strong stress “deafness”; that is, they have difficulty perceiving a stress contrast as compared to a phoneme contrast. The “deafness” effect is numerically large and does not differ among the languages. These results clarify the potential difference between French and Hungarian speakers found previously (Peperkamp & Dupoux, 2002) that was difficult to interpret because the two languages had been tested with different stimuli and procedures. It is now clear that when identical stimuli and procedure are used, there is no such difference: French and Hungarian speakers display a robust stress “deafness” of comparable magnitude. We also found that speakers of Polish exhibit an intermediate pattern, with a small but significant stress “deafness” effect that differs from both the indices of speakers of the other languages with predictable stress and from the index of Spanish speakers. This clarifies the trend in the same direction observed in Peperkamp and Dupoux (2002) with different stimuli: we now confirm that Polish is really an intermediate case between French and Spanish. Finally, we found a small but significant effect of durational cues, in that more errors were made in the absence of durational cues than in their presence. This difference was homogeneous across, on the one hand, speakers of languages in which duration is used as a cue to either stress or length (i.e. Finnish, Hungarian, and Spanish), and, on the other hand, languages in which duration is not used at all (i.e. Standard French, Southeastern French, and Polish).

In the Introduction, we outlined four factors that could govern the presence or absence of stress “deafness”, i.e. domain of stress assignment, lexical use of stress cues, variability in stress position, and lexical exceptions (see Table 1), and we formulated corresponding predictions for the perception of stress (see Table 2). The present results clearly rule out the first three of

these factors, whereas they provide evidence in favor of the fourth one. Factor 1, the domain of stress assignment, predicted stress “deafness” to be exhibited only by speakers of languages with phrasal stress, i.e. Standard and Southeastern French, contrary to fact. Factor 2, the lexical use of stress cues, predicted that Finnish and Hungarian speakers (whose languages have contrastive vowel length) should exhibit less stress “deafness” than speakers of the remaining languages with predictable stress. This prediction is not borne out, since Polish speakers performed better than all other groups (except the Spanish). Factor 2 also predicted that when duration is not part of the cues by which stress is realized, the performance of Finnish, Hungarian, and Spanish speakers (in whose languages duration is used as a cue to either stress or length) decreases more than that of Standard French, Southeastern French and Polish speakers (whose languages do not use duration at all). This prediction is not borne out either. Factor 3, the variability in stress position, predicted three levels of stress perception, corresponding to the three degrees of variability found in the six languages. In particular, Southeastern French and Polish speakers were predicted to show the same, intermediate performance, contrary to fact. This leaves us with Factor 4, the presence of lexical exceptions to the stress regularity. The present data are in accordance with the prediction yielded by this factor. That is, speakers of Polish, the only language with predictable stress that has a small number of lexical exceptions, perform better than speakers of the other languages with predictable stress (who do not differ from one another) and worse than speakers of Spanish, a language with a large percentage of exceptions to its default stress pattern.

Based on the results of our experiment, we propose the following empirical generalization: as far as the perception of stress contrasts is concerned, listeners fall into three classes, according to their native language: First, a strong “deafness” is exhibited by speakers of languages with predictable stress and no lexical exceptions, irrespective of the domain of stress assignment and variability of stress position. Second, a weak “deafness” is displayed by speakers of languages with predictable stress and few lexical exceptions. Third, no “deafness” is found in speakers of languages with unpredictable stress, that is, languages like Spanish, in which the number of lexical exceptions is very large compared to the regular pattern, if any. One caveat concerning this generalization is in order: The difference between languages yielding a strong “deafness” and those yielding a weak one is clear-cut, in that the presence of only a few lexical exceptions has a large consequence for the perception of stress. By contrast, our results do not allow us to infer whether the difference between languages yielding a weak “deafness” and those yielding no “deafness” at all is likewise a qualitative difference; that is, the amount of “deafness” might increase linearly with the percentage of lexical exceptions.

What does this empirical generalization mean for accounts of how linguistic knowledge impacts on speech processing? In the Introduction, we divided such accounts into two types: functional role accounts and lexical statistics accounts. In functional role accounts, the speech processing system encodes only those acoustic/phonetic dimensions or phonological features that have a functional (here: lexical) role in the language. For speakers of languages with predictable stress, these accounts predict less or no stress “deafness”, firstly, if stress is assigned at the word level (as opposed to at the phrasal level, Factor 1), and secondly, if at least one other suprasegmental is used lexically (Factor 2). As explained above, neither of these predictions is borne out by our data.

Lexical statistics accounts, by contrast, focus on the existence of phonological distributional regularities across lexical items. According to these accounts, the more predictable the stress position, the less stress is encoded and hence the more difficulty

listeners have perceiving stress contrasts. This prediction is borne out by our data, at least if predictability is construed as being related to the number of lexical exceptions (Factor 4). In light of this finding, it would be very interesting to test languages that are intermediate between Polish (.1% of exceptions) and Spanish (17% of exceptions).

Overall, then, lexical statistics accounts fare better for the present data. However, they face a challenge from acquisition studies: as mentioned before, French but not Spanish nine-month-old infants fail to discriminate a stress contrast (Skoruppa et al., 2009). In other words, stress “deafness” already emerges at nine months. At this age, however, it is unlikely that infants have acquired a large enough lexicon to start extracting meaningful statistics. To take up this challenge, it would be important to assess the developmental pathway of Polish infants with respect to the perception of stress. Given the very low frequency of lexical exceptions in Polish, it is likely that at nine months of age, they have not learned or even heard words with exceptional stress. Lexical statistics accounts thus predict that Polish nine-month-olds should perform initially like French nine-month-olds, and that they should improve their perception of stress by the time they learn lexical exceptions (which are usually taught in school). However, a question that should be examined in parallel – using statistical analyses of infant-directed speech – is whether a lexicon is really needed to acquire the distinction between predictable and unpredictable stress. For instance, Peperkamp and Dupoux (2002) claimed that this distinction can be acquired on the basis of utterance-level statistics only: given that utterance beginnings and endings necessarily correspond to word beginnings and endings, respectively, infants might infer whether stress is predictable or not by looking for regularities in the distribution of stresses at utterance edges. Infants as young as seven months are sensitive to utterance boundaries (Hirsh-Pasek et al., 1987). Therefore, if the distinction between predictable and contrastive stress can be acquired without a lexicon, we predict that Polish nine-month olds already have converged on the adult pattern.

To conclude, we have shown that language-specific perception of stress is best captured by lexical statistics accounts. Yet, more cross-linguistic research is necessary to investigate whether lexical statistics accounts are overall better suited to explain language-specific perception; such research should focus on the perceptual processing of other suprasegments as well as on that of segments.

Acknowledgments

This research was supported by grants from the Centre National de la Recherche Scientifique (Aide à Projet Nouveau, and Cognition et Traitement de l'Information, CTI02-15), the Agence Nationale pour la Recherche (ANR-05-BLAN-0065-01), and the European Union (FP6 012738). We would like to thank Anne Christophe for comments and discussion.

Appendix. Sequences used in the experiment

Two-word sequences:

11, 12, 21, 22

Five-word sequences:

11121, 11211, 11212, 11221, 12111, 12112, 12122, 12211, 12212, 12221

21112, 21121, 21122, 21211, 21221, 22112, 22121, 22122, 21222, 22212

References

- Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, 14, 136–193.
- Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Oxford: Blackwell.
- Comrie, B. (1967). Irregular stress in Polish and Macedonian. *International Review of Slavic Linguistics*, 1, 227–240.
- 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., Sebastián, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language* 36, 406–421.
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust paradigm to study stress “deafness”. *Journal of the Acoustical Society of America*, 110, 1606–1618.
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress “deafness”: The case of French learners of Spanish. *Cognition*, 106, 682–706.
- Durand, J. (1976). Generative phonology, dependency phonology and southern French. *Lingua e Stile*, 11, 3–23.
- Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.
- Francis, A., & Nusbaum, H. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 349–366.
- Gaskell, M. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31, 447–463.
- Gaskell, M., Hare, M., & Marslen-Wilson, W. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19, 407–439.
- Grammont, M. (1965). *Traité de phonétique*. Paris: Delagrave.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r”. *Neuropsychologia*, 9, 317–323.
- Gottfried, T., & Suiter, T. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25, 207–231.
- Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, 107, 2711–2724.
- Hallé, P., Chang, Y.-C., & Best, C. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395–421.
- Harris, J. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.
- Hirsh-Pasek, K., Kemler-Nelson, D., Jusczyk, P., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269–286.
- Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–B57.
- Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Karlsöon, F. (1999). *Finnish: An essential grammar*. London: Routledge.
- Kirilloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, 20, 2–4.
- Lado, R. (1957). *Linguistics across cultures: Applied Linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
- Logan, J., Lively, S., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 874–886.
- Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10, 241–265.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331–340.
- Morton, J., Crowder, R., & Prussin, H. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*, 91, 169–190.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–17.
- Peperkamp, S., & Dupoux, E. (2002). A typological study of stress “deafness”. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 203–240). Berlin: Mouton de Gruyter.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Joan L. Bybee, & Paul Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Philadelphia, PA: John Benjamins.
- Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177–228). Cambridge, MA: MIT Press.
- Santiago, J., Justicia, F., Palma, A., Huertas, D., & Gutiérrez, N. (1996). LEX I and II: Two databases of surface word forms for psycholinguistic

- research in Spanish. *Behavior Research Methods, Instruments & Computers*, 28, 418–426.
- Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., & Sebastián-Gallés, N., et al. (2009). Language-specific stress perception of stress by nine-month-old French and Spanish infants. *Developmental Science*, 12, 914–919.
- Smoczynska, M. (1985). The acquisition of Polish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 595–686). Hillsdale, NJ: Lawrence Erlbaum.
- Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, 16, 379–405.
- Tompá, J. (1972). *Kleine ungarische Grammatik*. Budapest: Akadémiai Kiadó.
- Weist, R., Wysocka, H., Witkowska-Stadnik, K., Buczowska, E., & Konieczna, E. (1984). The defective tense hypothesis: On the emergence of tense and aspect in child Polish. *Journal of Child Language*, 11, 347–374.
- Werker, J., & Tees, R. (1984a). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866–1878.
- Werker, J., & Tees, R. (1984b). Cross language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.
- Werker, J., & Tees, R. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509–535.