ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Cognition xxx (2007) xxx–xxx

COGNITION

www.elsevier.com/locate/COGNIT

Persistent stress 'deafness': The case of French learners of Spanish ☆

Emmanuel Dupoux [a,*], Núria Sebastián-Gallés [b], Eduardo Navarrete [b], Sharon Peperkamp [a,c]

[a] *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*

[b] *Grup de Recerca Neurociència Cognitiva, Parc Científic, Universitat de Barcelona, Spain*

[c] *Département des Sciences du Langage, Université de Paris VIII, Saint-Denis, France*

Received 22 December 2006; revised 4 April 2007; accepted 7 April 2007

Abstract

Previous research by Dupoux et al. [Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory Language 36,* 406–421; Dupoux, E., Peperkamp, S., & Sebastián-Gallés (2001). A robust method to study stress' deafness. *Journal of the Acoustical Society of America 110,* 1608–1618.] found that French speakers, as opposed to Spanish ones, are impaired in discrimination tasks with stimuli that vary only in the position of stress. However, what was called stress 'deafness' was only found in tasks that used high phonetic variability and memory load, not in cognitively less demanding tasks such as single token AX discrimination. This raised the possibility that instead of a perceptual

☆ Research for this paper was funded by grants from the Centre National de la Recherche Scientifique (Aide à Projet Nouveau, and Cognition et Traitement de l'Information, CTI02-15), the Groupement d'Intérêt Scientifique Sciences de la Cognition (99N35/0008), l'Agence Nationale pour la Recherche (ANR-05-BLAN-0065-01), the McDonnell Foundation (Bridging Mind, Brain and Behavior Program, JSMF 20002079), the Spanish Ministerio de Educacion y Ciencia (Grant Contract SEJ2004-07680-C02-01) and the Catalan government (2005SGR-01026). We are grateful to Katherine White for help in running the participants in Paris, and to Anne Christophe and Jacques Mehler for comments and discussion.

* Corresponding author.
E-mail address: dupoux@lscp.ehess.fr (E. Dupoux).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2007.04.001

problem, monolingual French speakers might simply lack a metalinguistic representation of contrastive stress, which would impair them in memory tasks. We examined a sample of 39 native speakers of French who underwent formal teaching of Spanish after age 10, and varied in degree of practice in this language. Using a sequence recall task, we observed in all our groups of late learners of Spanish the same impairment in short-term memory encoding of stress contrasts that was previously found in French monolinguals. Furthermore, using a speeded lexical decision task with word–nonword minimal pairs that differ only in the position of stress, we found that all late learners had much difficulty in the use of stress to access the lexicon. Our results show that stress 'deafness' is better interpreted as a lasting processing problem resulting from the impossibility for French speakers to encode contrastive stress in their phonological representations. This affects their memory encoding as well as their lexical access in on-line tasks. The generality of such a persistent suprasegmental 'deafness' is discussed in relation to current findings and models on the perception of non-native phonological contrasts.

Keywords: Stress 'deafness'; Second language acquisition; Plasticity

1. Introduction

After spending years in a foreign country, adults continue to experience important difficulties in dealing with several aspects of the foreign language. These difficulties are especially salient in the area of phonology, both in production (Flege, Frieda, & Nozawa, 1997; Piske, Flege, MacKay, & Meador, 2002) and in perception (MacKay, Meador, & Flege, 2001). Concerning difficulties in perception, they have been linked to the interference from the native-language phonology. Early work by Polivanov (1931) and Trubetzkoy (1939/1969) led to the conclusion that learners of a second language incorrectly apply their native set of phonological contrasts to parse those of the second language. More recently, a number of such interference effects have been documented in the psycholinguistic literature. Perhaps most well-known is the difficulty that native Japanese listeners, who use only a single liquid phoneme, experience in perceiving the English /r-l/ contrast (Goto, 1971; Miyawaki et al., 1981). These difficulties are persistent, as they are found in listeners with several years of exposure to English (Takagi & Mann, 1995). Moreover, Japanese monolinguals can improve their performance on /r-l/ discrimination over several weeks of intensive computerized training, but their performance remains significantly below that of English monolinguals (Lively, Logan, & Pisoni, 1993), and they continue to use different acoustic cues than native listeners (Takagi, 2002). Similar problems have been reported for the perception of vocalic contrasts, like the Catalan /e-ɛ/ contrast, which is difficult for native speakers of Spanish, (Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés, Echeverría, & Bosch, 2005), or the English /a-ʌ/ and /a-æ/ contrasts, which are difficult for native speakers of Italian (Flege & Mackay, 2004; Flege, MacKay, & Meador, 1999). It is noteworthy, however, that not all non-native contrasts are equally difficult. Best, McRoberts, and Sithole (1988) documented that an acoustically minimal contrast in Zulu clicks can be discriminated very well by English listeners despite the fact that English uses no such contrasts. For Japanese listeners, /b/ vs. /v/ is initially as difficult as /r/ vs. /l/; however, practice yields a large increase in performance in the former, not the latter (Guion, Flege, Akahane-Yamada, & Pruitt, 2000).

Several theoretical approaches have been proposed to account for these patterns of data. In some models, the acoustic perceptual space is 'warped' according to experience. This means that the weighting of certain acoustic dimensions or cues (the slope of certain formants, fine-grained timing properties, etc) is enhanced or reduced, depending on their functional values in the language (Francis & Nusbaum, 2002; Iverson et al., 2003; Jusczyk, 1997; Nosofsky, 1986). In these models, non-functional perceptual dimensions will yield poor perceptual discrimination. Other models are based on similar proposals, but rely on abstract phonetic (Flege, 1995) or even phonological features (Lado, 1957; Brown, 1998, 2000). In such models, when a non-native contrast cannot be parsed in terms of a contrastive feature of the native language, perceptual problems arise. Finally, a third class of models proposes that perceptual difficulties are due to the interference from native phoneme categories or prototypes. Phoneme categories reduce the multidimensional and detailed representations of a stimulus to a discrete linguistic label, plus a goodness rating indicating how far the stimulus is from the center of the category (Grieser & Kuhl, 1989; Best, 1994). Best (1994) proposes that upon exposure to a speech sound, the perceptual system automatically assimilates it to the closest category. If two sounds are assimilated to the same category, they will be very difficult to discriminate, unless they differ greatly in category goodness. Stimuli that are too distinct from any existing category will not be assimilated, and hence, their acoustic details remain available, yielding good discrimination (Flege, 1995).

Considerable effort has been devoted to explicitly distinguishing these models in the realm of vowel and consonant perception, generally supporting models based on perceptual warping and prototype formation (Guion et al., 2000; Kuhl, 2000; Best, McRoberts, & Goodell, 2001; Kingston, 2003). Less research has been directed towards the perception of suprasegmentals. Suprasegmentals are interesting, because they involve different types of acoustic cues. Indeed, whereas segments involve fast and fine-grained spectral changes, suprasegmentals involve slower and more global acoustic cues like $F0$, energy and duration (Lehiste, 1970). These cues are, moreover, also used to carry information regarding prosodic constituent structure, pragmatic content, emotional state, etc., adding to the variability in their distributions. The description in terms of discrete categories with a prototype structure may therefore be less pertinent for suprasegmentals than for segments. Hence, it is not clear that the above perceptual effects for segmental contrasts would also apply to suprasegmental contrasts.

Yet, recent research has pointed to interesting parallels between the perception of segmental and suprasegmental contrasts, in particular ones concerning tones. First, it has been found that native speakers of tone languages perceive their native tonal contrasts in a more categorical fashion than non-native speakers (Hallé, Chang, & Best, 2004; Xu, Gandour, & Francis, 2006). In addition, they process these contrasts in the left hemisphere, whereas speakers of non-tone languages process them either in the right hemisphere or bilaterally (Gandour, Wong, & Hutchins, 1998; Gandour et al., 2000; Klein, Zatorre, Milner, & Zhao, 2001). Furthermore, native speakers of English who learn Mandarin have much difficulty in learning the tones of this language, and rely on slightly different acoustic cues than native speakers (Stagray & Downs, 1993). As in the case of segments, not all tonal contrasts are equally difficult. This finding has been linked to the interference from existing suprasegmental contrasts in the native language, like stress for English learners of Mandarin (Kiriloff, 1969; White, 1981; Shen, 1989; Chen, 1997). Finally, as in the case of /r/ and /l/, English adults can be trained to increase their performance on tone perception, with performance reaching a suboptimal plateau (from 66% correct pre-training to 87% post-training, Wang, Spence, Jongman, & Sereno, 1999).

In this paper, we examine the perception of a stress contrast by French listeners. Stress is instantiated by three acoustic cues: *F*0, duration and energy, none of which is used in French to signal a phonological contrast. Indeed, French has no contrastive tone, pitch accent or stress, and duration is used contrastively in neither vowels nor consonants. Duration is only used allophonically, with vowels being lengthened before certain consonants (Casagrande, 1984). French is described as having a phrasal 'accent', which is realized as final syllable lengthening in prosodic groups with no increase in *F*0 or intensity (Rossi, 1980; Vaissière, 1991); this phrasal accent thus has a demarcative, not a contrastive function. Hence, a contrast between, for instance, Spanish /bébe/ '(s/he) drinks' and /bebé/ 'baby' cannot be represented by an existing contrastive phonological or phonetic feature in French.

Depending on the theoretical model, the lack of contrastive use of suprasegmentals in French yields two divergent predictions. According to a model where the existing inventory of linguistic features is determinant, French participants should have much difficulty in representing and acquiring contrastive stress. This would parallel the finding that native speakers of Spanish have more difficulties than native speakers of English to learn Swedish vowel duration contrasts (McAllister, Flege, & Piske, 2002); indeed, in Spanish, vowel duration is not used at all, whereas in English, it is a secondary cue to the contrast between tense and lax vowels. In contrast, according to a model where the limiting factor is the similarity to an existing prototype, French listeners should find it very easy to acquire contrastive stress. This would parallel findings that American listeners can distinguish Zulu clicks quite easily (Best et al., 1988), and that French listeners have some strong abilities to distinguish Mandarin tones, even though they behave in a less categorical fashion (Hallé et al., 2004).

So far, the data accumulated on the processing of contrastive stress by French listeners supports both types of models (Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001). Indeed, when French participants are tested with a standard AX discrimination task on a stress contrast, they do not differ from Spanish monolinguals. Yet, if memory load and/or talker variability is introduced, French participants display a very strong impairment to perceive stress contrasts (see Table 1).

Prima facie, this pattern of data can be interpreted as showing that French listeners have a true impairment concerning the processing of contrastive stress. Accord-

Table 1
Summary of error rates in discrimination or sequence recall tasks on minimal pairs that differ either in stress position or in a single phoneme, in French and Spanish participants

	French participants		Spanish participants	
	Stress	Phoneme	Stress	Phoneme
AX, low phonetic variability[a]	3.8%	4.0%	—	—
ABX, high phonetic variability[b]	*19%*	5.8%	4.0%	8.8%
Sequence recall, low phonetic variability[c]	27%	23%	19%	17%
Sequence recall, high phonetic variability[d]	*53%*	15%	20%	24%

Notes:
[a] Percent error in a same-different AX task with minimal pairs differing in stress position or in one phoneme. The stimuli were recorded by a single talker, and the "Same" tokens were physically identical (Dupoux et al., 1997; Experiment 4).
[b] Percent error in an ABX discrimination task with the same contrasts and stimuli. The A,B, and X stimuli were recorded by different talkers (Dupoux et al., 1997; Experiments 1 and 2).
[c] Percent error in a sequence recall task using binary sequences of minimal pairs that differ in stress position or in one phoneme. Only one physical token per member of each minimal pair was used (Dupoux et al., 2001; Experiment 5).
[d] Percent error in a sequence recall task using binary sequences of minimal pairs that differ in stress position or in one phoneme. Six tokens per member of each minimal pair were used (Dupoux et al., 2001; Experiment 3).

ing to this interpretation, their poor performance in the high variability experiments shows that they lack a proper phonological representation of stress that they could use to encode stress contrasts in an on-line fashion. It is only in the low variability tasks that they have enough time and resources to discriminate the stress contrast, using an ad hoc acoustic strategy. However, there is an alternative explanation which needs to be addressed. Perhaps French listeners are not impaired in their phonological representation of contrastive stress, but rather have difficulty with the *metalinguistic* access to this contrast. Note that the short-term memory task requires participants to explicitly associate two stimulus classes to two response keys. In order for this association to be learned efficiently, metalinguistic access to these two classes is required. Yet, monolingual French listeners are likely not to have a metalinguistic representation of Spanish stress, contrary to Spanish, contrastive stress does not exist in French, stress is not marked in the orthography, and it is not taught in school. Would late learners, who surely have metalinguistic knowledge of Spanish stress, show the same limitation as French monolinguals in their processing of contrastive stress?

The aim of this paper is thus to examine whether the performance pattern of French listeners is due to a true processing deficit or to metalinguistic limitations. We test French learners of Spanish, who have learned about contrastive stress as part of their explicit L2 training. If the problem is a true processing limitation, beginning learners of Spanish should not differ from monolinguals, and only more advanced learners should start to show an enhancement in performance. If, in contrast, the problem is metalinguistic, even a limited exposure to Spanish should enable the French participants to use the stress labels and hence enhance their performance Furthermore, in order to evaluate whether stress 'deafness' is limited to tasks requiring metalinguistic access to stress categories, we contrast two tasks: the first one is the sequence recall task with high phonetic variability of Dupoux et al. (2001), the second one is a lexical decision task using word–nonword minimal pairs that vary only in stress position. Again, if French listeners have a real processing deficit as far as stress is concerned, then French learners of Spanish should have difficulty rejecting nonwords that differ from a real word with respect to the position of stress only. By contrast, if our previous findings with French monolinguals are due to metalinguistic limitations, then French learners of Spanish should have no difficulty in encoding stress in lexical items, and hence correctly reject such nonwords.

2. Sample

Thirty-nine native speakers of French who were late learners of Spanish participated in this study. There were 11 men and 28 women, aged between 20 and 57 (mean = 28). Six late learners lived and were tested in Paris; they had been recruited from among university students in a Spanish language department. The remaining 33 had lived in Barcelona for at least 6 months and were tested there; they had been recruited from among undergraduate students who spent a year on an exchange program. Only four of them had some knowledge of Catalan, the other official language spoken in Barcelona.[1] None of the participants had started to learn Spanish before age 11 or any other foreign language before age 10.

The control samples consisted of 20 native speakers of French, who were tested in Paris, and 20 native speakers of Spanish, who were tested in Barcelona. The native French participants, 7 men and 13 women, were aged between 18 and 28 (mean = 24); none of them had learned Spanish and none of them had started to learn a foreign language before age ten. The majority of them had some basic knowledge of English and German, learned in school. The native Spanish participants, 3 men and 17 women, were aged between 18 and 25 (mean = 21). All had been born in monolingual Spanish families in Barcelona but were fluent in Catalan as well.

The 39 late learners filled out a questionnaire concerning their Spanish language background (age/place/manner of acquisition, length of residence in a Spanish-speaking country) and their current usage of Spanish (visits to Spanish-speaking countries, private and professional usage of Spanish). On the basis of these questionnaires (and without knowing the participants' scores on the various stress perception tasks or their performance on the production task to be discussed below), three of the authors (ED, NS, and SP) independently classified each participant as being a beginner, intermediate, or advanced learner. They then compared their classifications and discussed cases of disagreement. If the disagreement persisted, the final classification was the one that had been proposed by two out of three authors (there were no cases in which each author had proposed a different classification). According to this final classification, 14 late learners were beginners, 14 were intermediate and 11 were advanced learners.

[1] The stress systems of Spanish and Catalan are very similar: stress falls on one of the last three syllables in the word, with penultimate stress being the default.

Table 2
Biographic data and language background of three groups of French late learners of Spanish (Beginner, Intermediate, and Advanced)

	Degree of practice with Spanish			Mean
	Beginner ($N = 14$)	Intermediate ($N = 14$)	Advanced ($N = 11$)	
		Biographic data		
Mean age	27	26	30	*28*
Country of testing				
France	$N = 2$ (14%)	$N = 2$ (14%)	$N = 2$ (18%)	$N = 6$ (15%)
Spain	$N = 12$ (86%)	$N = 12$ (86%)	$N = 9$ (82%)	$N = 33$ (85%)
		Language background		
Mean age of first instruction	17.5 years	17.8 years	15.1 years	*16.9*
Mean length of residence in a Spanish-speaking country	0.7 years	2 years	4.3 years	*2.2*
Learned Spanish in school	71%	64%	82%	*72%*
Learned Spanish at university	71%	57%	82%	*69%*
Regularly visits Spanish-speaking countries	4%	43%	100%	*45%*
Regularly speaks Spanish in private life	7%	61%	68%	*44%*
Regularly speaks Spanish in professional/student life	32%	50%	64%	*47%*

Table 2 summarizes the biographic data and Spanish language background of the three groups of late learners.

As can be seen, majorities in all three groups learned Spanish in school and at the university. The three groups vary with respect to the remaining variables as follows. Advanced learners started to learn Spanish at a younger age than beginners (at 15.1 vs. 17.5 years), and likewise they spent more time living in a Spanish-speaking country (4.3 vs. 0.7 years). Higher percentages of advanced learners than of beginners often visit Spanish-speaking countries (100% vs. 4%) and regularly speak Spanish in their private lives (68% vs. 7%) as well as during professional and/or student activities (64% vs. 32%). As to the intermediate learners, they are practically identical to the beginners with respect to the age of first instruction, whereas they are more similar to the advanced learners with respect to the percentage of individuals who regularly visit Spanish-speaking countries; with respect to the remaining three variables they are truly intermediate between the beginners and the advanced learners.

3. Language assessment

3.1. Self-evaluation

All late learners rated their Spanish competence in several domains, i.e. pronunciation, grammar, and vocabulary, on a scale from one to ten, as well as the importance of Spanish in their own lives (Table 3).

Please cite this article in press as: Dupoux, E. et al., Persistent stress 'deafness': The case of French ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.001

Table 3
Self-evaluation of Spanish competence and importance of Spanish for three groups of French late learners of Spanish (Beginner, Intermediate, and Advanced)

	Degree of practice with Spanish			Mean
	Beginner ($N = 14$)	Intermediate ($N = 14$)	Advanced ($N = 11$)	
Pronunciation (1: Extremely poor –10: Perfect)				
Mean	6.00	6.07	7.55	*6.46*
SE	0.53	0.43	0.25	*0.27*
Grammar (1: Extremely poor – 10: Perfect)				
Mean	6.71	6.93	7.73	*7.08*
SE	0.49	0.34	0.33	*0.24*
Vocabulary (1: Extremely poor – 10: Perfect)				
Mean	6.79	6.93	7.82	*7.13*
SE	0.33	0.43	0.23	*0.21*
Importance of Spanish (1: No importance – 10: Extremely important)				
Mean	8.86	9.14	9.36	*9.10*
SE	0.31	0.29	0.15	*0.16*

These data were subjected to a series of analyses of variance (ANOVA) with the between-participant factor Practice (Beginner vs. Intermediate vs. Advanced) and the dependent variables Pronunciation, Grammar, Vocabulary, and Importance, respectively. For Pronunciation there was a significant effect of Practice ($F(2,36) = 3.5$, $p < .04$). The beginners gave significantly lower estimates of their pronunciation than the advanced learners ($F(1,23) = 5.8$, $p < .03$), with the intermediate learners falling in-between. For Grammar, the effect of Practice was not significant ($F(2,36) = 1.6$, $p > .1$). For Vocabulary, it was not significant either ($F(2,36) = 2.3$, $p > .1$), although the beginners gave significantly lower estimates of their vocabulary than the advanced learners ($F(1,23) = 5.8$, $p < .03$). Finally, for Importance, the effect of Practice again did not reach significance ($F < 1$).

3.2. Native-like accent

In order to evaluate the productive competence of the late learners in Spanish, we examined their native-like accent in a reading task. In this test, participants had to read aloud a short text (a story for children) that, when read by a native Spanish speaker, lasts about 30 s. They were instructed to first read the text silently to get acquainted with it and then to read it aloud as quickly as possible. All late learners as well as 18 Spanish controls participated in this test. The recordings were digitized at 16 kHz at 16 bits and stored on a computer disk. Due to a technical problem, the data of three late learners were lost. The individual sound files were then presented to five native speakers of Spanish in Barcelona who were neither phoneticians nor French language specialists. These naive judges rated the native-like Spanish accent of the speakers on a scale from 1 (*extremely poor*) to 6 (*perfect*) (Table 4).

The ratings for all participants were subjected to an analysis of variance (ANOVA) with the between-participant factor Native Language (French vs. Span-

Table 4
Native-like accent of three groups of French late learners of Spanish (Beginner, Intermediate, and Advanced) and Spanish controls on a scale from 1 (*extremely poor*) to 6 (*perfect*)

	Mean	SE
Late learners ($N = 36$)		
Beginner ($N = 12$)	*1.95*	*0.26*
Intermediate ($N = 14$)	*2.77*	*0.27*
Advanced ($N = 10$)	*3.42*	*0.35*
Mean	2.68	0.19
Spanish ($N = 18$)	5.67	0.09

ish). The late learners had a significantly poorer accent than the Spanish controls ($F(1,52) = 117.1$, $p < .0001$). An analysis of the late learners alone revealed a significant effect of Practice on accent ($F(2,33) = 6.0$, $p < .007$). The beginners had a significantly poorer accent than the advanced learners ($F(1,20) = 11.9$, $p < .004$), and the intermediate learners fell in between. None of the late learners approached the distribution of the Spanish control sample by less than two standard deviations.[2]

3.3. Summary

We selected a group of late learners of Spanish who varied in their degree of practice with this language. Practice was found to significantly affect the participants' self-evaluation of pronunciation, and the judge-evaluation of their degree of native-like accent.

4. Experiment 1: Stress encoding in short-term memory

In this experiment, we use a modified version of the high phonetic variability sequence recall task of Dupoux et al. (2001). This task is meant to make the acoustic level of representation not accessible and hence to highlight the phonological level. The present version of the task differs from the one used by Dupoux et al. (2001) in three respects. First, it uses sequences of length four only, thus avoiding both floor and ceiling effects. Second, the stimuli are produced by several talkers. Although Dupoux et al. (2001) observed no difference between using one male and one female talker and using a single female talker with variable base $F0$, we reasoned that the presence of multiple talkers (3 men, 3 women) would prevent the possibility of using acoustic cues. Third, the speed of presentation is increased in that the stimuli are slightly compressed and the ISI is reduced. The reasoning is that at fast presentation rate, only automatic phonological encoding is expected to be operative.

[2] We also counted the number of stress errors that the participants made in the 64 polysyllabic words contained in the passage. Beginners, intermediate learners, and advanced learners made on average 7.2%, 7.3%, and 7.8% errors, respectively. None of the differences between these groups was significant.

Please cite this article in press as: Dupoux, E. et al., Persistent stress 'deafness': The case of French ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.001

4.1. Method

4.1.1. Materials and design

Two minimal pairs were constructed, one involving a consonantal contrast, i.e. /fiku–fitu/, the other one involving a stress contrast, i.e. /númi–numí/. All items are nonwords in both French and Spanish. They were recorded several times by six native speakers of French, 3 men and 3 women, who were trained to produce Spanish stress patterns. In Spanish, the main stress cues are $F0$ and duration (Ortega-Llebaria, 2006); all stimuli were judged by a Spanish phonetician and only tokens with unambiguous stress patterns were used. For each talker, one recording of each item was selected. All recorded items were digitized at 16 kHz at 16 bits, digitally edited, and stored on a computer disk. They were then time compressed to 84% of their original duration, yielding a mean duration of 345 and 338 ms, for the phoneme and the stress contrasts, respectively.

As to the tokens with the stress contrast, stressed vowels were on average 35 ms longer than unstressed vowels, a significant difference ($F(1,5) = 10.9$, $p < .03$). Stressed vowels also had a higher $F0$ than unstressed vowels: the maximum $F0$ value of the stressed vowels was on average 96 Hz higher than that of the unstressed vowels, corresponding to a significant difference of 8.5 semitones ($F(1,5) = 141.0$, $p < .0001$). Finally, stressed vowels were on average 6.1 dB louder than unstressed vowels, again a significant difference ($F(1,5) = 143.1$, $p < .0001$).

For each minimal pair, one warm-up block and one experimental block were constructed, consisting of sequences of the two nonwords. The warm-up block contained four two-item sequences; the experimental block contained 28 four-item sequences (see Appendix A).

4.1.2. Procedure

Participants were tested individually in sound-attenuated booths. They were first tested on the minimal pair containing the segmental contrast. They were told that they were going to learn two words in a foreign language, and were first asked to press the number key [1], upon which they heard all tokens of the first item, /fiku/. They were then asked to press the number key [2], upon which they heard all tokens of the second item, /fitu/. Subsequently, they could continue listening to the various tokens of the two items by pressing the associated keys; pressing each one of these keys resulted in the playing of one token of the corresponding item. Participants could thus hear as many tokens of the two items as they wished. When they indicated they were ready to move on, they were trained on the distinction between the two items as well as the correct association between the items and the number keys. 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 participants whether their responses were correct. We defined a success criterion of seven correct responses in a row. After having reached this criterion, participants turned to the main experiment, consisting of four warm-up trials and 28 test trials.

Warm-up and test trials consisted in the oral presentation of a sequence constituted by two or four repetitions, respectively, of the two items. The task was to reproduce each sequence by typing the associated keys in the correct order. For each participant, the order of sequences was randomized, and each item was instantiated randomly by one of the six recorded tokens with the proviso that a single token could not appear more than once in a sequence. In order to diminish the likelihood that participants mentally translate the words into the associated numbers while listening to the sequence, the silent period between the items in a sequence was kept very short, i.e. 50 ms. Each trial consisted of a sequence followed by the word 'OK', which was used to prevent participants from using echoic memory (Morton, Crowder, & Prussin, 1971; Morton, Marcus, & Ottley, 1981), and participants could not begin typing their response until they had heard this word. 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. During the warm-up phase, incorrect responses gave rise to an error message that was displayed for 800 ms, after which the same sequence was repeated until the correct response was given. During the test phase, no feedback was provided and no sequences were repeated. A 1500 ms pause separated each response from the next trial.

The whole procedure was repeated with the minimal pair containing a stress contrast. The nonword with stress on the first syllable, /númi/, was associated with key [1], while its counterpart with stress on the second syllable, /numí/, was associated with key [2].

On average, the experiment lasted around 15 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 either the phoneme or the stress condition 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.

4.1.3. Participants

Twenty French controls, 20 Spanish controls, and the 39 late learners participated. Three late learners were rejected due to too many complete reversals in the stress condition. Four French controls were rejected and replaced, due to too many complete reversals in either the phoneme condition (one participant) or the stress condition (three participants). Finally, one Spanish participant was rejected and replaced, due to too many errors in the phoneme condition.

4.2. Results

The mean lengths of the training phase and the mean error rates in the test phase for each group and contrast are presented in Table 5.

Table 5
Mean length of training phase and mean error rates with standard errors in the test phase for French controls, three groups of French late learners of Spanish (Beginner, Intermediate, and Advanced), and Spanish controls on the phoneme and the stress contrast in Experiment 1

		Training length	Test % errors (SE)
French ($N = 20$)	Stress	29.5	78.7 (3.0)
	Phoneme	7.6	27.0 (3.6)
	Difference[a]	21.9[*]	51.7[***]
Late learners ($N = 36$)			
Beginner ($N = 14$)	Stress	*19.5*	*66.1 (6.0)*
	Phoneme	*7.1*	*22.4 (5.6)*
	Difference	*12.4* ~	43.7[***]
Intermediate ($N = 13$)	Stress	*9.4*	*78.8 (4.4)*
	Phoneme	*8.3*	*20.6 (4.0)*
	Difference	*1.1* ns	58.2[***]
Advanced ($N = 9$)	Stress	*8.5*	*74.2 (2.6)*
	Phoneme	*7.4*	*25.4 (7.9)*
	Difference	*1.1* ns	48.8[***]
Mean	Stress	13.1	72.7 (3.0)
	Phoneme	7.6	22.5 (3.2)
	Difference	5.5[*]	50.2[***]
Spanish ($N = 20$)	Stress	7.2	19.6 (2.2)
	Phoneme	7.8	25.0 (4.6)
	Difference	−0.6 ns	−5.4

Note. ns: $p > .1$; ~: $.05 < p < .1$.
[a] The analyses were done on the log transform of the training length.
[*] $p < .05$.
[***] $p < .001$.

4.2.1. Training phase

We analyzed the number of trials required to reach criterion in the training phase. Inspection of individual responses revealed a lot of variability in training lengths, with some very long ones (over 50 trials). Such a distribution is to be expected given the nature of the criterion: if the error rate is above 1/7, the training phase can last for a very long time. To analyze such a time series which is analogous to a Poisson process, we took the log of the training length for each participant and condition and then ran separate ANOVAs for each group. In the French group, there was an effect of contrast, due to a higher error rate in the stress than in the phoneme condition ($F(1,19) = 5.2$, $p < .04$). The late learners showed a smaller but significant effect ($F(1,35) = 4.8$, $p < .05$), and the Spanish group showed no effect ($F < 1$). Within the late learners, the difference between stress and phoneme diminishes with practice, and only the beginners had a marginal difference between the two conditions ($F(1,13) = 3.4$, $p = .087$).

4.2.2. Test phase, error rates

The mean error rates were analyzed with ANOVAs in each group separately and then with a combined ANOVA. Both French monolinguals and late learners showed significantly more errors in the stress than in the phoneme conditions ($F(1,19) = 122.5$, $p < .001$; $F(1,35) = 174.5$, $p < .001$, respectively). The two groups did not differ, as shown by the absence of an effect of Group ($F(1,54) = 1.9$, $p > .1$), and of an interaction between Group and Contrast ($F < 1$). In Spanish monolinguals, there was no significant effect of Contrast ($F(1,19) = 2.1$, $p > .1$) and there was an interaction between Contrast and Group when compared to the French monolinguals ($F(1,38) = 91.2$, $p < .001$) and the late learners ($F(1,54) = 91.3$, $p < .001$), respectively. Within the late learners, we found no effect of Practice ($F < 1$) and no interaction between Practice and Contrast ($F(2,33) = 1.4$, $p > .1$). All subgroups of late learners showed a significant difference between stress and phoneme (beginners: $F(1,13) = 45.5$, $p < .001$; intermediate: $F(1,12) = 116.9$, $p < .001$; advanced: $F(1,8) = 37.4$, $p < .001$).

4.2.3. Test phase, mutual information

The analysis in terms of error rates may not be very sensitive, because a single mistake counts as much as a series of mistakes in a sequence. Possibly, global error rates may be hiding a slight improvement in the encoding of stress between French monolinguals and French learners of Spanish. In order to test this, we computed the mutual information between the individual binary targets (e.g., /númi/ vs. /numí/) and the individual binary responses ([1] vs. [2]) over the entire test trials, for each participant and condition. Mutual information can be interpreted as the amount of information (in bits) about the response that can be gained by observing the target stimulus (or vice versa). Alternatively, if one views a participant as a noisy channel, it is a measure of the capacity of that channel to transmit information. In the formula below, the stimulus is denoted by the binary random variable S and the response by the binary random variable R. The probabilities associated to these variables are estimated over the entire sequences of responses made by a given participant for a given condition.

$$I(S,R) = \sum_{s \in S} \sum_{r \in R} p(s,r) \cdot \log_2 \frac{p(s,r)}{p(s)p(r)}$$

The average mutual information across conditions and groups is shown in Fig. 1.

ANOVAs revealed that both French monolinguals and late learners had a lower mutual information in the stress compared to the phoneme condition ($F(1,19) = 38.9$, $p < .001$ and $F(1,35) = 84.3$, $p < .001$, respectively); the two groups did not differ, as shown by the absence of both an effect of Group and an interaction between Group and Contrast (both $F < 1$). In Spanish monolinguals there was a tendency for an effect in the other direction, i.e. higher mutual information for stress than for phoneme ($F(1,19) = 4.7$, $p < .05$), but there was an interaction between Group and Contrast when compared to French monolinguals ($F(1,38) = 34.4$, $p < .001$) and late learners ($F(1,54) = 59.5$, $p < .001$), respectively. Within the late learners, there was no significant effect of Practice and no interaction between Practice and Contrast (both $F < 1$).

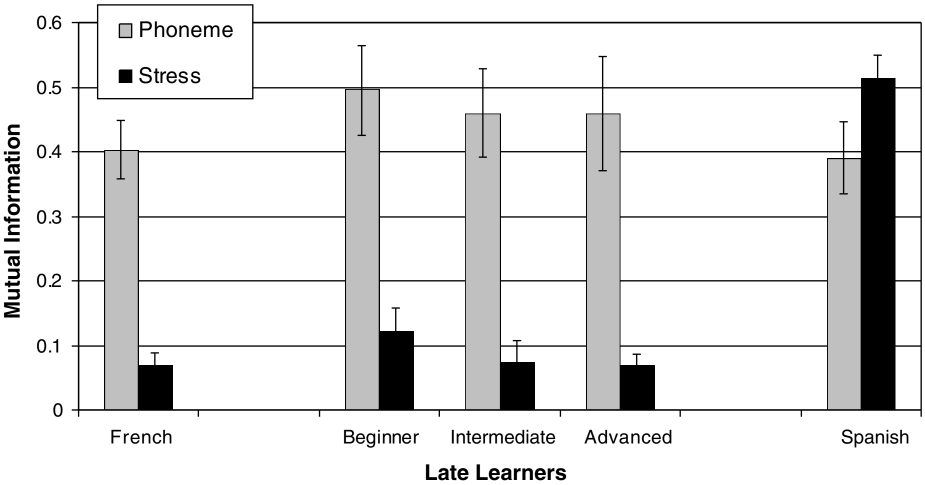

Fig. 1. Mutual information scores for the three groups of French late learners of Spanish as well as Spanish and French controls, in the stress versus phoneme condition in Experiment 1. Error bars indicate one standard error around the mean.

4.3. Discussion

As in Dupoux et al. (2001), the sequence recall test showed a robust stress 'deafness' effect in the French controls as well as in the late learners. They made 2.9 and 3.2 times more errors in the stress condition, respectively, than the Spanish controls, and there was no difference between the two groups. The bit-wise information-theoretic analysis of the response sequences confirmed this result. Importantly, there was no significant improvement between the subgroups of late learners with different degrees of practice with Spanish. That is, even the advanced learners did not perform better than the French controls who had no practice with Spanish at all. This result enables us to discard the hypothesis that the stress 'deafness' found previously was an artifact due to the fact that French monolinguals had no familiarity with languages with contrastive stress. Stress 'deafness' emerges here as a robust processing limitation, which cannot be eliminated with a significant exposure to a language with contrastive stress.

5. Experiment 2: Stress encoding in the lexicon

In order to explore whether the difficulty with the perception of stress is not limited to short-term memory tasks, but extends to the use of stress for on-line lexical access, we examined the use of stress in lexical representations by means of a lexical decision task. In this experiment, we presented real Spanish words, as well as nonwords obtained from these words by changing the location of stress. For instance, *górro* is a word ('hat'), but *gorró* is not. Participants had to make a speeded lexical decision; both reaction times and error rates were measured. As a control condition, we used word–nonword minimal pairs differing by a single phoneme.

5.1. Method

5.1.1. Materials and design

One hundred and twelve word–nonword pairs that differed only in the position of stress were selected as test items (see Appendix B). Ninety-six of them were disyllabic, half of the real words having initial stress and the other half final stress. The remaining 16 items were trisyllabic (2 words with initial stress, 6 with medial stress and 8 with final stress). An additional 40 word–nonword pairs that differed only in one phoneme were selected as control items. Five items were monosyllabic, 22 were bisyllabic, 10 were trisyllabic, and 3 were quadrisyllabic. All the stimuli were recorded by a female Spanish native speaker. All recorded items were digitized at 16 kHz, digitally edited, and stored on a computer disk.

Two stimuli lists (A and B) composed of 56 test words, 56 test nonwords, 20 control words and 20 control nonwords were created. These lists were compiled such that if a real word was included in one list, the corresponding nonword was included in the other list.

5.1.2. Procedure

Participants were tested individually in sound-attenuated booths. They were instructed to decide as quickly and accurately as possible if the stimuli presented were words or not by pressing one of two labeled buttons with their dominant hand. The late learners were informed that some nonwords had the very same sounds of an existing word but with a change in the position of stress. Each participant was tested with just one list. The order of presentation of the stimuli was randomized for each participant. Preceding the test phase, 30 warm-up trials were presented. None of the warm-up items was included in either experimental list, and none of them was a nonword differing from a real word in the position of stress only. During the warm-up phase, feedback was given after each response with the words ‘correct’ and ‘incorrect’. During the test phase, no feedback was given. A trial started with the presentation of a fixation point, an asterisk in the centre of the screen, during 500 ms. The auditory stimulus was presented 500 ms after the disappearance of the fixation point. The next trial began 3500 ms after the end of the auditory stimulus. Instructions were given in Spanish.

The test lasted for approximately 20 min.

5.1.3. Participants

Twenty Spanish controls and the 39 late learners participated. All of them had also participated in the sequence recall test. Three of the late learners, however, had been rejected in this test.

5.2. Results

Mean reaction times for correct responses and error rates for the three groups of French late learners of Spanish and the Spanish controls on words and nonwords in the test and the control conditions are shown in Table 6.

The test condition and the control condition were analyzed separately.

Please cite this article in press as: Dupoux, E. et al., Persistent stress ‘deafness’: The case of French ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.001

Table 6
Mean reaction times for correct responses and error rates for three groups of French late learners of Spanish (Beginner, Intermediate, and Advanced) and for Spanish controls on words and nonwords in the test and control conditions of Experiment 2

			RT (ms)	SE	Error (%)	SE
Late learners ($N = 39$)						
Beginner ($N = 14$)	*Test*	*Word*	*1365*	*66*	*31.7*	*2.0*
		Nonword	*1644*	*89*	*55.9*	*4.1*
	Control	*Word*	*1235*	*67*	*11.1*	*1.7*
		Nonword	*1497*	*75*	*18.2*	*4.9*
Intermediate ($N = 14$)	*Test*	*Word*	*1341*	*62*	*23.1*	*2.4*
		Nonword	*1635*	*83*	*61.6*	*3.4*
	Control	*Word*	*1163*	*54*	*5.0*	*1.5*
		Nonword	*1479*	*64*	*16.1*	*3.1*
Advanced ($N = 11$)	*Test*	*Word*	*1104*	*53*	*16.4*	*3.1*
		Nonword	*1460*	*67*	*56.1*	*5.7*
	Control	*Word*	*1055*	*40*	*0.9*	*0.6*
		Nonword	*1452*	*72*	*10.9*	*2.2*
Mean	Test	Word	1283	39	24.3	1.7
		Nonword	1589	48	58.0	2.5
	Control	Word	1158	34	6.0	1.0
		Nonword	1478	40	15.4	2.2
Spanish ($N = 20$)	Test	Word	917	17	3.9	0.9
		Nonword	1085	23	4.5	1.0
	Control	Word	876	17	1.0	0.6
		Nonword	1123	31	2.5	0.9

5.2.1. Test condition

Two ANOVAs were run on the error rates in the test condition, one by participants ($F1$) and one by items ($F2$), with between-participants factor Group (Spanish vs. late learners) and within-participant factor Lexical Status (word vs. nonword). These ANOVAs revealed main effects of Lexical Status ($F1(1,57) = 110.8$, $p < .001$; $F2(1,103) = 76.3$, $p < .001$) and Group ($F1(1,57) = 293.6$, $p < .001$; $F2(1,103) = 1175.4$, $p < .001$), as well as an interaction between these two factors ($F1(1,57) = 62.8$, $p < .001$; $F2(1,103) = 102.3$, $p < .001$). Restricted analyses revealed that the interaction was due to the fact that there was no effect of Lexical Status in the Spanish group ($F1(1,18) = 2.0$, $p > .1$; $F2 < 1$, $p > .1$), but a very large effect of Lexical Status in the late learners, who made almost twice as many errors with the nonwords than with the words (58.0% vs. 24.3%; $F1(1,37) = 104.7$, $p < .001$; $F2(1,103) = 608.3$; $p < .001$). That is, the late learners tended to misidentify the nonwords as words much more often than that they rejected real words. The analysis of the reaction times was not performed because of the very high error rates of the late learners. However, the numerical pattern was very similar, with the late learners having longer reaction times than the Spanish controls.

5.2.2. Control condition

The same ANOVAs were run on both the error rates and the reaction times in the control condition. As for the error rates, there were significant effects of Group

($F1(1,57) = 21.2$, $p < .001$; $F2(1,78) = 44.9$, $p < .001$) and Lexical Status ($F1(1,57) = 22.6$, $p < .001$; $F2(1,78) = 11.2$, $p < .001$), as well as an interaction between these two factors ($F1(1,57) = 7.0$, $p < .02$; $F2(1,78) = 9.1$, $p < .003$). An analysis restricted to the late learners revealed a significant effect of Practice ($F1(2,36) = 3.7$, $p < .04$; $F2(2,156) = 8.3$, $p < .001$), with the beginners making more errors than the advanced learners (14.7 vs. 5.9, $p < .02$), the intermediate learners falling in-between. ANOVAs on the reaction times likewise revealed significant effects of Group ($F1(1,57) = 39.8$, $p < .001$; $F2(1,78) = 454.7$, $p < .001$) and Lexical Status ($F1(1,57) = 206.4$, $p < .001$; $F2(1,78) = 101.8$, $p < .001$), but no interaction between the two ($F < 1$). The effect of Practice on reaction time did not reach significance ($p > .1$), though. Thirty-six percent of the late learners fell within two standard deviations of the control Spanish sample on the error rate, and 20% (all advanced learners) fell within two standard deviations regarding both the error rate and the reaction times.

5.2.3. Analysis

The late learners made more errors on the words in the test condition, where nonwords differed from real words in the position of stress, than in the control condition, where nonwords differed from real words in a phoneme (24% vs. 6%). This difference might be due to the fact that the words in the test condition were less frequent than those in the control condition. Participants simply might not have known certain low-frequency items and therefore shown increasing error rates for both words and nonwords. We hence need to establish that the effects we observed in the test condition were due to the difficulty in processing stress rather than to limited lexical knowledge. In order to do this, we conducted a reanalysis separating the test items into four frequency bands containing 28 items each: High, Medhigh, Medlow, and Low, with mean log frequencies of 2.5, 1.7, 1.2, and 0.5, respectively. (The mean log frequency of the control items was 3.1.) In particular, we classified each response as either a hit (real word, response ‘yes’), a false alarm (nonword, response ‘yes’), a miss (real word, response ‘no’), or a correct rejection (nonword, response ‘no’), and calculated A' scores for each participant across the test items in each frequency band, as well as across the control items. The average A' scores in the three groups of late learners and the control groups are shown in Fig. 2.

An ANOVA restricted to the late learners on the test items with the between-participant factor Practice (Beginner vs. Intermediate vs. Advanced) and the within-participant factor Frequency (High vs. Medium High vs. Medium Low vs. Low) showed a marginal effect of Practice ($F(2,36) = 3.0$, $p < .07$), a significant effect of Frequency ($F(3,108) = 3.6$, $p < .02$), and no interaction between these two factors ($F(6,108) = 1.6$, $p > .1$). An ANOVA restricted to the Spanish controls showed a significant effect of Frequency only ($F(3,57) = 7.6$, $p < .0001$).

5.3. Discussion

In this experiment, late learners and Spanish controls performed a lexical decision task on word–nonword minimal pairs that differ either in the position of stress (test

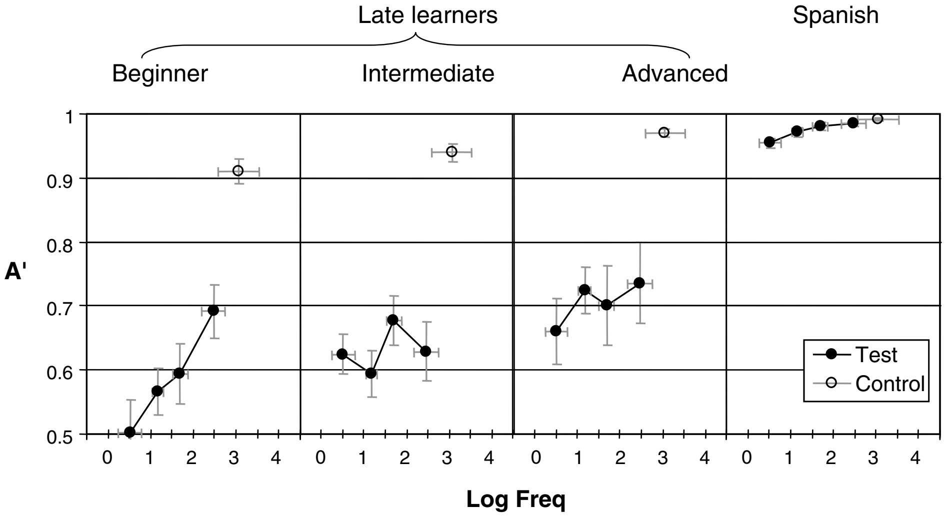

Fig. 2. Mean A′ scores for three groups of French late learners of Spanish as well as Spanish controls in the test and control conditions as a function of $\log_{10}$ Frequency in Experiment 2. Frequency is calculated as number of occurrences per 1.25 million (Sebastián-Gallés et al., 2000). Horizontal error bars represent the standard deviation of the frequencies in the group of words; vertical error bars indicate one standard error of the A′ score around the mean.

condition) or in a single phoneme (control condition). In the test condition, the late learners' error patterns for words and nonwords were very different. For words, the error rate was moderate (24%), and it diminished with degree of practice with Spanish, as well as with word frequency (with additive effects of these two variables). For nonwords, the error rate was very high (58%), meaning that participants predominantly classified the nonwords as words. This error rate was not influenced by degree of practice with Spanish, and increased with word frequency. Contrary to the late learners, the Spanish control participants made very few errors overall.

The late learners' pattern of results obtained in the control condition is very different from that in the test condition. In the control condition, which contained high-frequency items only (mean log frequency: 3.1), the error rate for the words was indeed similar to that for the high-frequency words in the test condition (6% vs. 12%, respectively), but the error rate for the nonwords was much lower than that for the high-frequency nonwords in the test condition (15% vs. 67%, respectively). Both the error rates for the words and the nonwords decreased as a function of subjects' level of practice.

These results demonstrate that the stress 'deafness' effect observed with the sequence recall task is not limited to the encoding of stress in short-term memory but extends to lexical access. In other words, the suprasegmental distinction that is pertinent to the encoding of words in Spanish barely seems to be available to French learners of Spanish, even to those who have quite a good mastery of the Spanish lexicon. This is consistent with the hypothesis that the encoding of stress remains by and large unavailable to French learners of Spanish, at least as far as usage in on-line tasks is concerned.

6. General discussion

The main finding of this study is that French late learners of Spanish have much difficulty in the perception of stress, as assessed in a sequence recall task and a lexical decision task. In the sequence recall experiment, the late learners were undistinguishable from the French monolingual controls and they did not show an effect of practice: those learners who on average had spent three years in a Spanish-speaking country performed as badly as beginners with only a few months of practice with Spanish. Likewise, in the on-line lexical decision experiment, the late learners' error rates on word–nonword minimal pairs that differ on stress were very high, irrespective of practice with Spanish; in contrast, on control word–nonword minimal pairs involving a phonemic distinction, advanced learners made less errors than beginners, showing their improvement in the construction of a Spanish lexicon.

The present results allow us to rule out the possibility that the stress 'deafness' displayed by native speakers of French is due to the absence of a metalinguistic representation of stress. Indeed, the late learners in the present study did have metalinguistic awareness of stress, since they had been taught explicitly about contrastive stress as part of their Spanish language training, but they did not perform differently from the control French monolinguals. Hence, the deficit in the perception of stress found in French listeners is a bona fide and persistent perceptual/processing limitation. Consequently, we conclude that the good performance of French listeners in some of the experiments in Dupoux et al. (1997, 2001) is due to the fact that in these experiments, subjects could use an acoustic representation rather than an abstract phonological one.

Our results are important for both methodological and theoretical reasons. On the methodological side, the case of French stress 'deafness' shows that standard discrimination tasks can underestimate processing difficulties in a foreign language. Stress has such large acoustic correlates that French participants can resort to acoustic strategies in certain tasks. However, this intact psychoacoustic ability[3] does not translate into good performances in other, more demanding, tasks. This result, then, highlights the importance of testing participants with a range of tasks, going from discrimination tasks to tasks that measure lexical activation (see also Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés & Baus, 2005).

On the theoretical side, our results allow us to draw some conclusions regarding the appropriateness of the different types of speech perception models for the case of suprasegmentals. First, models that are cast in terms of the 'warping' of acoustic space (Francis & Nusbaum, 2002; Iverson et al., 2003) do not necessarily predict per-

[3] A study using near-threshold stimuli might actually show that French speakers have a subtle psychoacoustic deficit to differences in *F*0, duration, and/or energy (see, for instance, Hallé et al., 2004).

Please cite this article in press as: Dupoux, E. et al., Persistent stress 'deafness': The case of French ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.001

ceptual problems with suprasegmentals in French. Indeed, although French has no contrastive stress, it makes use of prosodic features at several processing levels, for instance to mark prosodic boundaries, grammatical focus, and allophonic variation. Regarding the first issue, Christophe, Peperkamp, Pallier, Block, and Mehler (2004) found that in French, phonological phrase boundaries are marked by small differences in *F*0, duration, and energy of the segments surrounding the boundary. Furthermore, French listeners are able to exploit these cues on-line for the purposes of word segmentation. Interestingly, the differences in the acoustic parameters in Christophe et al. (2004) are of the same order of magnitude as or even smaller than the ones introduced in our Experiment 1 (vowel duration difference: 30–40%; *F*0 difference: 1.8–2.8 semitones; energy difference: 1.3 dB) (see also Millotte, Wales, & Christophe, in press). This suggests that French participants have a rather good sensitivity to the suprasegmental correlates of stress, and hence their processing difficulties with stress are unexpected under this account.

Second, models that appeal to the existence of native categories or prototypes predict no particular problems with contrastive stress. Indeed, the prosodic characteristics of stress minimal pairs cannot be mapped to any existing prosodic prototype in French. As a consequence, according to models such as the ones proposed by Best (1994) and Flege (1995), it should be relatively easy to add prototypes for contrastive stress contours. However, this does not seem to be the case for late learners, at least not for those with the amount of exposure to Spanish that we tested.

Third, models that are formulated in terms of the influence of native phonological features (Lado, 1957; Brown, 2000) correctly predict an impairment in stress processing in French. Indeed, French has no contrastive suprasegmental feature at all: French has no contrastive stress, no pitch accent, no lexical tones, and no long vowels. As a result, the minimal pairs that we used in our two experiments cannot be differentiated linguistically with the use of an existing L1 feature. Interestingly, the impairment that results appears to be located at a rather abstract phonological level, instead of at the psychoacoustic level, and is only revealed with tasks that make a direct use of a phonological representation. This meshes well with findings concerning other suprasegmentals 'deafnesses', such as the vowel length 'deafness' that has been documented in both monolingual French listeners (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999) and bilingual Spanish listeners (McAllister et al., 2002).[4]

In sum, among the three types of speech perception models, the ones that are cast in terms of perceptual warping and prototype formation clearly fail to account for the present results, whereas the ones that highlight the influence of the native phonological features fare much better. This is somewhat surprising, since in the realm of segmental 'deafnesses', it is the former types of models that have received a large

[4] Note that although the absence of contrastive phonological features may be sufficient to yield a phonological 'deafness', their presence does not imply that a non-native contrast can be learned easily. For instance, in the domain of segmental contrasts, Flege and Port (1981) found that although Arabic speakers make contrastive use of both voicing and labiality, they fail to adequately produce the English contrast /p-b/ contrast. Similarly, Peperkamp and Dupoux (2002) reported stress 'deafness' in native speakers of Finnish and Hungarian, two languages with fixed stress in which vowel length is contrastive.

amount of empirical support (Flege, 1995; Guion et al., 2000; Kuhl, 2000; Best et al., 2001; Kingston, 2003). One may speculate that this is not due to a linguistic difference between segments and suprasegmentals, but rather to the way in which suprasegmentals are processed in the brain. Suprasegmentals have slow and large-scale acoustic correlates, which makes them likely candidates for right hemisphere processing. Indeed, contrary to non-native segments, non-native suprasegmentals tend to activate regions in the right hemisphere that are homologous to language areas (Gandour et al., 1998; Klein et al., 2001). It could be that during early language acquisition, only those suprasegmentals that are contrastive in the native language are captured by the left hemisphere. In adulthood, non-native suprasegmentals are processed in the right hemisphere, and are hence not available for phonological representations. More research using both brain imagery and early bilinguals is needed to test this hypothesis.

Appendix A. Sequences used in the sequence recall task (Experiment 1)

two-item sequences: 11, 12, 21, 22

four-item sequences:
used twice: 1211, 1121, 2122, 2212, 2111, 1112, 2221, 1222, 1212, 2121, 1122, 2211
used once: 1221, 2112, 1221, 2112

Appendix B. Items used in the lexical decision task (Experiment 2)

Bisyllabic test items (only the words are indicated, the nonwords being derived from them by changing the position of stress, indicated here by an acute accent)

ágil	dátil	láser	perdón	sofá
alúd	débil	licór	pintór	tabú
atún	debú	líder	pólen	tapíz
balcón	dedál	limón	postál	tótem
bidé	dócil	lóro	póster	túnel
borrón	dólmen	mantél	póte	tupé
buzón	fácil	mármol	pulmón	tutór
café	fémur	marrón	puñál	útil
cáliz	fértil	melón	puré	vádo
calór	fósil	menú	rímel	vejéz
cáncer	fusíl	metál	rincón	verbál
cánon	gabán	mítin	rópa	víga
cárcel	górro	móvil	rubí	vírgen
carríl	gurú	mudéz	salúd	volcán
cénit	hábil	nácar	sartén	vudú

Please cite this article in press as: Dupoux, E. et al., Persistent stress 'deafness': The case of French ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.001

Appendix B (*continued*)

césped	jamón	néctar	séso	zúlo
chófer	járro	niñéz	séto	
chúrro	jóven	paréd	sifón	
cómic	láca	paté	sillín	
córtex	lápiz	pedál	símil	

Trisyllabic test items (word–nonword minimal pairs; stress is indicated by an acute accent)

cadáver–*cadavér	humedád–*humédad	tenedór–*tenédor
caníbal–*canibál	nadadór–*nadádor	vegetál–*vegétal
comedór–*comédor	recitál–*recítal	volátil–*volatíl
corazón–*corázon	régimen–*regimén	volúmen–*volumén
déficit–*deficít	resúmen–*resumén	
difícil–*dificíl	soledád–*solédad	

Control items (word–nonword minimal pairs)

aire–*airo	experiencia–*exporiéncia	noche–*soche
año–*añi	familia–*femilia	nunca–*nenca
ayer–*ayem	fuerte–*fuerbe	palabra–*pelabra
blanco–*blanto	futuro–*feturo	paz–*maz
cabeza–*tabeza	grupo–*brupo	presidente–*presodente
cara–*cora	hora–*hore	razón–*ralon
cielo–*cieto	importante–*impertante	señor–*veñor
clase–*blase	interior–*inderior	social–*sicial
contra–*cuntra	lado–*ladi	tiempo–*tiempe
cuatro–*muatro	luz–*suz	tres–*trel
difícil–*daficil	mañana–*mañano	vida–*vidi
dinero–*dijero	mar–*mur	voz–*vuz
edad–*emad	miedo–*tiedo	
epoca–*étoca	momento–*momenco	

References

Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of the Acoustical Society of America, 109*, 775–794.

Best, C. T., McRoberts, G. W., & Sithole, M. N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345–360.

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.

Casagrande, J. (1984). *The sound system of french*. Washington, DC: Georgetown University Press.

Chen, Q. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association, 32*, 21–39.

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.

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 method to study stress 'deafness'. *Journal of the Acoustical Society of America, 110*, 1606–1618.

Flege, J. E. (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.

Flege, J. E., Frieda, A. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*, 169–186.

Flege, J. E., & Mackay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, 1–34.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*, 2973–2978.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*, 125–146.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experiment Psychology: Human Perception and Performance, 28*, 349–366.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, D. (2000). A cross-linguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12*, 207–222.

Gandour, J., Wong, D., & Hutchins, D. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport, 9*, 2115–2119.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'r' and 'l'. *Neuropsychologia, 9*, 317–323.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577–588.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (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). Categorical perception of Taiwan Mandarin Chinese tones by Chinese versus French native speakers. *Journal of Phonetics, 32*, 395–421.

Iverson, P., Kuhl, P. K., 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. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kingston, J. (2003). learning foreign vowels. *Language and Speech, 46*, 295–349.

Kiriloff, C. (1969). On the auditory discrimination of tones in Mandarin. *Phonetica, 20*, 63–67.

Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage, 13*, 646–653.

Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science, 97*, 11850–11857.

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.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/ . II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

MacKay, I. R., Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58*, 103–125.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229–258.

Millotte, S., Wales, R., & Christophe, A. (in press). Phrasal prosody disambiguates syntax. Language and Cognitive Processes.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1981). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics, 18*, 331–340.

Morton, J., Crowder, R., & Prussin, H. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology, 91*, 169–190.

Morton, J., Marcus, S., & Ottley, P. (1981). The acoustic correlates of "speechlike": A use of the suffix effect. *Journal of Experimental Psychology: General, 110*, 568–593.

Nosofsky, R. N. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

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). Somerville, MA: Cascadilla Press.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*, B9–B17.

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.

Piske, T., Flege, J. E., MacKay, I. R., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 59*, 49–71.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague, 4*, 79–96.

Rossi, M. (1980). Le français, langue sans accent? [French: A non-stressed language?]. In I. F'onagy & P. Léon (Eds.), *L'accent en français contemporain [Stress in modern French]* (pp. 13–51). Paris: Didier.

Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.), *Twenty-first century psycholinguistics* (pp. 279–292). Mahwah, NJ: Lawrence Erlbaum Associates.

Sebastián-Gallés, N., Cuetos, F., Carreiras, M., & Martí-Antonin, M. (2000). *LEXESP. Léxico informatizado del español*. Barcelona: Publicacions de la Universitat de Barcelona.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240–255.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition, 72*, 111–123.

Shen, X. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association, 24*, 27–47.

Stagray, J., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics, 21*, 143–163.

Takagi, N. (2002). The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *Journal of the Acoustical Society of America, 111*, 2887–2896.

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.

Trubetzkoy, N. S. (1939/1969). *Principles of phonology translated by C.A. Baltaxe*. Berkeley, CA: University of California.

Vaissière, J. (1991) Rhythm, accentuation and final lengthening in French. In: J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (108–120). Wenner–Gren International Symposium Serie, vol. 59.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649–3658.

White, C. (1981). Tonal perception errors and interference from English intonation. *Journal of the Chinese Language Teachers Association, 16*, 27–56.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America, 120*, 1063–1074.

Please cite this article in press as: Dupoux, E. et al., Persistent stress 'deafness': The case of French ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.001