



Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access[☆]

Begoña Díaz^{a,*}, Holger Mitterer^b, Mirjam Broersma^b, Núria Sebastián-Gallés^c

^a Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany

^b Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

^c Brain and Cognition Unit, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain

ARTICLE INFO

Article history:

Received 14 June 2011

Received in revised form 14 February 2012

Accepted 1 May 2012

Keywords:

Bilingualism

Phonology

Language learning

Speech perception

ABSTRACT

The extent to which the phonetic system of a second language is mastered varies across individuals. The present study evaluates the pattern of individual differences in late bilinguals across different phonological processes. Fifty-five late Dutch-English bilinguals were tested on their ability to perceive a difficult L2 speech contrast (the English /æ/-/ɛ/ contrast) in three different tasks: A categorization task, a word identification task and a lexical decision task. As a group, L2 listeners were less accurate than native listeners. However, at the individual level, almost half of the L2 listeners scored within the native range in the categorization task whereas a small percentage scored within the native range in the identification and lexical decision tasks. These results show that L2 listeners' performance crucially depends on the nature of the task, with higher L2 listener accuracy on an acoustic-phonetic analysis task than on tasks involving lexical processes. These findings parallel previous results for early bilinguals, where the pattern of performance was consistent with the processing hierarchy proposed by different models of speech perception. The results indicate that the analysis of patterns of non-native performance can provide important insights concerning the architecture of the speech perception system and the issue of language learnability.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

A general observation about learning a second language (L2) is that the later a language is learned, the poorer the final proficiency attained (DeKeyser, 2000; Flege, Yeni-Komshian, & Liu, 1999; Johnson & Newport, 1989; Weber-Fox & Neville, 1996). Although this is true of most learners, some individuals fall outside this general tendency (Bongaerts, 1999; Rossi, Gugler, Friederici, & Hahne, 2006). Certainly, some late L2 learners demonstrate excellent performance while some early learners are particularly impervious to experience. One of the language domains where these individual differences are most evident is phonology. Within phonology, however, there are distinguishable processes (Pisoni & Luce, 1987), which may not be equally easy to master. For example, recognizing speech sounds requires a different skill than recognizing words containing these sounds, and L2 learners who are proficient at the former may not necessarily be proficient at the latter. Thus, different perceptual tasks that are meant to measure phonological processing might show very different outcomes depending on which phonological process they tap into. This might affect the outcome of any study interested in speech sound perception.

Sebastián-Gallés and Baus (2005) demonstrated that highly proficient early bilinguals differed in their mastery of the L2 phonology and that such individual variability was reflected in the pattern of performance across phonological processes (i.e., categorization of isolated vowels, identification of vowels within words, detection of vowel mispronunciations within words). More bilinguals were able to perform within the limits of native listener performance on a phonological categorization task than on tasks tapping into processes of lexical access and selection. These results demonstrate the existence of distinguishable phonological processes (Pisoni & Luce, 1987) also in non-native listeners. However, the bilinguals tested in Sebastián-Gallés and Baus (2005) had acquired the second language early in life and were highly proficient. A question that remains open is whether late (and less proficient) bilinguals may show different patterns of individual differences than were observed in early bilinguals. It may be the case that late bilinguals' non-native phonological processes are less efficient. If so, the patterns reported by Sebastián-Gallés and Baus (2005) reflecting different degrees of difficulties in the processes of lexical access must not be observed in late bilinguals.

In the present study, we address the question of whether the same pattern of individual differences in L2 phonetic mastery is found in late bilinguals. This question is approached by evaluating L2 phonological processes of late bilinguals across similar domains as tested by Sebastián-Gallés and Baus (2005), ranging from acoustic-phonetic analysis to lexical access. The results of this study will

[☆] The authors want to thank Xavier Mayoral for his technical support.

* Corresponding author. Tel.: +49 34199402480; fax: +49 3419940104.

E-mail address: diaz@cbs.mpg.de (B. Díaz).

shed important light on our understanding of the mechanisms underlying non-native language learning, as well as help to characterize individual differences.

Language-specific phonological representations are very rapidly established in early development. Infants develop perceptual sensitivities to the phonemes of the native language before the end of the first year of life (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). At the same time, the capacity to perceive many non-native phoneme contrasts also declines in this period (Werker & Tees, 1984; Kuhl et al., 1992; Polka & Werker, 1994; Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009; Albareda-Castellot, Pons, & Sebastián-Gallés, 2011; for exceptions to this pattern see Best, McRoberts, & Sithole, 1988; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Best & McRoberts, 2003). This age effect on the acquisition of the phonology, as well as syntax, of an L2 has been claimed to stem from a reduction in brain plasticity during development (Johnson & Newport, 1989; Pulvermüller & Schumann, 1994). However, even bilinguals who acquired the L2 early in life still experience difficulties in the perception of some L2 phonetic contrasts (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Pallier, Bosch, & Sebastián-Gallés, 1997; Pallier, Colomé, & Sebastián-Gallés, 2001; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer, & Díaz, 2006; Williams, 1980).

Nevertheless, studies on L2 phonetic learning have shown that not all L2 phonemes are equally difficult to learn (Best, 1995; Best, McRoberts, & Sithole, 1988; Flege, 1995). The ease or difficulty of learning new speech sounds is determined by the relation between the established L1 categories and the new L2 categories. For instance, the learning of an L2 phonetic contrast is especially difficult when both elements of the contrasts are assimilated to the same L1 category (e.g., the English /r/-l/ speech contrast maps onto the Japanese category /l/).

Another relevant factor that should be considered when assessing L2 phonetic command is the type of L2 phonological process involved in the experimental task. Sebastián-Gallés and Baus (2005) assessed the learning of a difficult L2 contrast in a large population of very early and highly proficient Spanish-Catalan bilinguals. The participants were raised in monolingual Spanish families, thus Spanish was their only language in infancy and early childhood. From the age of four (at the latest), they were continuously exposed to Catalan. For this population, previous studies showed that the L2 Catalan-specific contrast /ɛ/-/e/ is very difficult to master (Bosch, Costa, & Sebastián-Gallés, 2000; Pallier et al., 1997; Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés et al., 2005, 2006). Participants in the study by Sebastián-Gallés and Baus (2005) were compared across three different L2 auditory tasks involving the Catalan contrast /ɛ/-/e/: a categorization task, a gating task and a lexical decision task. Each task was intended to assess different phonological processes. First, acoustic-phonetic analyses were evaluated by means of a categorization task. Second, a modification of the gating task with a two-forced-choice between a minimal pair visually presented was used to assess the time course of lexical activation. Finally, a lexical decision task was conducted to measure the degree of phonetic detail encoded during lexical access. When compared with the performance of a group of native Catalan listeners, the degree of early bilinguals' proficiency gradually decreased across tasks. The results of the categorization task showed that a high percentage of participants (68.3%) were able to categorize the stimuli within the range of natives. In the gating task, a lower percentage of participants (46.6%) were able to choose the correct response on the basis of the same amount of phonetic information that the Catalan natives required to choose the correct response. Finally, only a few participants (18.3%) showed a native-like level of performance in the lexical decision task.

Taken together, these results suggest that the choice of a particular task substantially modulates the results of any study on non-native phonetic mastery. With regards to individual differences in Sebastián-Gallés and Baus (2005) study, 23% of the participants scored below native level in all three tasks while a small minority of the participants

(12%) performed consistently within the native range in all three tasks (the rest of the participants fell within the range of natives for some, but not all, tasks).

Interestingly, these individual differences in the mastery of the L2 phonology mirrored the discrimination accuracy of native phonemes. Díaz, Baus, Escera, Costa, and Sebastián-Gallés (2008) compared the Mismatch Negativity (MMN), an electrophysiological brain response, of two groups of bilinguals, selected from the participants in the study by Sebastián-Gallés and Baus (2005), who performed either below or at native-like levels on all three L2 behavioral tasks. Therefore, only participants representing extreme cases of non-native speech perception abilities were selected. The two groups differed in their brain responses to an unknown foreign speech contrast and also to a native phonetic contrast, while no differences between the groups were found in the discrimination of tones that varied in several basic acoustic parameters (i.e., frequency, duration, and order of presentation). The more proficient group had larger MMNs than the less proficient group, indicating more precise phoneme representations in both the first and the unknown language. This result highlights the relevance of exploring individual differences in L2 for a better comprehension of general language mechanisms.

As mentioned, the population studied by Sebastián-Gallés and Baus (2005) and Díaz et al. (2008) consisted of early bilinguals who were exposed to the second language very early in life (and very intensively). It remains to be determined whether the same pattern of individual differences in L2 phonetic mastery is found in late and less proficient bilinguals. To address this issue, we tested 55 Dutch listeners who learned a second language, English, and achieved high levels of proficiency while growing up in a monolingual Dutch environment. All participants started to learn the L2, English, at the age of 11–12 years and had comparable English education and experience backgrounds. Note that “perfect” acquisition of L2 phonology is already compromised as early as 4 (Pallier et al., 1997) or 7 years of age (Caramazza et al., 1973); hence, in terms of phonological learning, the present sample of participants can be labeled as late bilinguals.

These late bilinguals were evaluated in their phonological processing of the English /æ/-/ɛ/ vowel contrast. Previous studies have shown that Dutch native listeners experience significant difficulties in distinguishing this contrast in phonetic categorization and identification studies (Broersma, 2005a; Cutler, Weber, Smits, & Cooper, 2004; Schouten, 1975) as well as word recognition studies (Broersma & Cutler, 2011; Cutler & Broersma, 2005; Weber & Cutler, 2004). Participants were tested in three behavioral tasks involving the critical English /æ/-/ɛ/ contrast: a phoneme categorization task, a lexical decision task, and a word identification task. Sebastián-Gallés and Baus (2005) used a categorization task, a gating task, and a lexical decision task to explore distinct phonological processes. The present study also includes a categorization and a lexical decision task. However, the gating task is replaced by a word identification task. The word identification task, as the gating, measures phonological representations in lexical forms by asking participants to match a spoken word between two orthographically presented words that form a minimal pair (words that differ in only one phoneme). However, in the gating task, participants are presented sequentially with segments of the critical word that increase in duration. In contrast, the word identification task presents the critical word only one time, which can potentially increase the difficulty of the task.

The categorization task was intended to assess acoustic-phonological analysis of the difficult L2 phonetic contrast. The English /æ/ and /ɛ/ vowels differ in spectral and temporal dimensions, and both dimensions are exploited by native English listeners for distinguishing English vowels (Cebrian, 2006; Flege, Bohn, & Jang, 1997). In addition, previous evidence has shown that Dutch listeners are able to exploit English vowel duration as a cue for final fricative voicing (Broersma, 2008, 2010). Whether Dutch-English late bilinguals also use spectral and temporal information for vowel categorization will be assessed by experimentally manipulating

frequency and duration. Thus, for the present categorization task, a 7-step continuum of seven synthesized vowels was created ranging from /æ/ to /ɛ/. In addition, for each stimulus, two instances were created with the prototypical duration of either /æ/ or /ɛ/. Stimuli were presented one at a time as participants categorized them as belonging to either /æ/ or /ɛ/ categories.

The lexical decision and the word identification tasks evaluated the extent to which the phonetic parameters that distinguish /æ/ and /ɛ/ are used in lexical access and selection. In the lexical decision task, participants were presented with words containing either /æ/ or /ɛ/. Non-words were created by replacing the critical English vowel by its counterpart (e.g. æ-word: lamp, æ-non-word: lemp). Participants had to decide if the stimulus was an existing English word or not. In the word identification task, English minimal pairs were presented. Each minimal pair differed in the critical English contrast /æ/-/ɛ/ (e.g. æ-word: cattle, ε-word: kettle). Words were presented in neutral carrier sentences realized by a native English speaker (i.e., "Now I say..."). Each sentence presentation was accompanied by two pictures representing the two words that comprised the minimal pair (e.g., a picture of 'cattle' and one of a 'kettle'). Participants were asked to indicate which picture matched the word they heard. To determine native performance level, a group of English native listeners was also tested. Based on their scores, appropriate thresholds were established for each task.

Overall performance of Dutch participants was expected to be less accurate than that of native English listeners in the three tasks. Moreover, based on the previous work of Sebastián-Gallés and Baus (2005), a decline in Dutch participants' performance was expected as lexical processes become necessary in task performance. That is, Dutch participants should be more proficient in the categorization task than in the lexical decision and word identification tasks. Crucially, we assessed individual variability by evaluating the distribution of L2 listeners' performance patterns over the three experimental tasks.

2. Methods

2.1. Participants

Fifty-five native speakers of Dutch (42 females; age = 21.16, *sd* = 2.47) and twelve native speakers of English (7 females; age = 28.25, *sd* = 3.76) took part in the experiment. Dutch participants were recruited from the Max Planck Institute participant pool and were fluent speakers of English. All participants had received at least 5 years (mean = 7, *sd* = 2.0) of English instruction in primary and secondary education, starting, on average, when they were 11 year-olds (*sd* = 1.01). All except four of the Dutch native speakers had lived in the Netherlands all their lives (two participants had lived in Canada for four and eight months, respectively, and the other two had spent six months and 1 year, respectively, in South Africa). Dutch participants were university students at the undergraduate or graduate level. Participants self-reported that they were highly proficient in English (students attending Dutch universities follow lectures in English). They were paid for their participation. None of them reported having any hearing or language-learning difficulty.

In addition, twelve native British English speakers participated in the study (given the reduced variability in native performance in these types of experimental situations, it is general practice to test a reduced number of native participants: Munro, Flege, & MacKay, 1996; Flege, MacKay, & Meador, 1999; Sebastián-Gallés & Baus, 2005). All British English listeners were graduate or undergraduate students. Four English speakers were exchange students at the Radboud University Nijmegen. They reported a low or non-existing knowledge of Dutch. They were paid for their participation. The other eight native English speakers performed an online version of the three tasks posted on the Max Planck Institute for Psycholinguistics web site. None of them had any knowledge of Dutch. Out of these eight participants, four were exchange students at the University of Barcelona. They were also paid

for their participation. The other four native English speakers were recruited from different UK universities. They were all in the UK at the time of testing and no financial compensation was given for their participation in the study.

2.2. Stimuli

In all the tasks speech stimuli were presented. A male British English speaker recorded the (models of the) materials for the categorization (Task 1), the lexical decision (Task 2), and the word identification (Task 3) tasks.

2.2.1. Task 1: categorization task

For the categorization task, seven synthesized vowels along the /ɛ/-/æ/ continuum were created using the source-filter synthesis module of the PRAAT software (Boersma, 2001). The two endpoint stimuli of the continuum (*s1* = /ɛ/ and *s7* = /æ/) were created based on the formant frequencies measured in the native British English speaker.

The two endpoints of the continuum differed in *F1* and *F2* (/ɛ/: *F1* = 600 Hz, *F2* = 1800 Hz; /æ/: *F1* = 740 Hz, *F2* = 1630 Hz) while the other formants were identical (*F3* = 2750 Hz, *F4* = 3400 Hz and *F5* = 4500 Hz). From the endpoint stimuli, five stimuli were created by varying *F1* and *F2* in steps of 23.33 Hz and 28.33 Hz, respectively. In addition, the English vowels /ɛ/ and /æ/ are also different in duration. As duration is a contrastive feature in the Dutch phonetic system, the continuum stimuli were susceptible to being categorized based on duration cues. For that reason, two tokens of each synthesized stimulus were created. Each stimulus had one token with the duration of /ɛ/ (115 ms) and one with the duration of /æ/ (165 ms) as measured in the native English speaker productions.

2.2.2. Task 2: lexical decision task

Similar to in Sebastián-Gallés and Baus (2005) and Sebastián-Gallés et al. (2005), a lexical decision task was employed. This task was a direct replication of the lexical decision experiment reported in Broersma and Cutler (2008, 2011). Sixteen monosyllabic English words containing the vowel /ɛ/ and sixteen monosyllabic English words containing the vowel /æ/ were selected. Logarithmic lemma frequencies per million of the experimental words were calculated with the CELEX lexical database of British English (Baayen, Piepenbrock, & Gulikers, 1995). No frequency differences were observed between æ-words (2.10) and ε-words (2.00) ($t_{(15)} < 1$). For each word, a list of non-words was created by exchanging the vowels /æ/ and /ɛ/. For instance, from the ε-word "desk" the corresponding non-word "dask" was created, and from the æ-word "lamp" the corresponding non-word "lemp" was created. The former were labeled ε-type stimuli (i.e., ε-word: desk, ε-non-word: dask) and the latter æ-type stimuli (i.e., æ-word: lamp, æ-non-word: lemp). In addition, 84 English words and 84 non-words, all monosyllabic, were selected as filler items. Filler non-words were created by replacing either a vowel or the final consonant of real English words without violating the phonotactic constraints of English. Replacements made for the filler non-words were expected to be relatively easy for Dutch listeners to perceive, as they involved phoneme contrasts that are also present in the Dutch phonology. None of the stimuli were homophonous to any existing Dutch words (for more details, see Broersma & Cutler, 2011).

Two stimulus lists containing the same number of experimental words and non-words of each vowel type were constructed in such a way that any word and its corresponding non-word counterpart did not appear in the same list. In addition, both lists included all filler words and non-words. In total, each list was made up of 100 words and 100 non-words. The materials were recorded by the British English speaker while he read the items one by one, separated by a pause, in a clear citation style. The recording was made in a soundproof booth using a high quality microphone onto digital audiotape and down-sampled to 16 kHz during transfer to a computer.

2.2.3. Task 3: word identification task

For the word identification task, 51 English minimal pairs were selected. Each pair of words differed in the vocalic contrast /ɛ/-/æ/ (for instance, kettle /kɛtəl/ and cattle /kætəl/). Logarithmic lemma frequencies per million of the experimental words were calculated with the CELEX lexical database of British English (Baayen et al., 1995). No frequency differences were observed between the words containing /æ/ (1.16) and /ɛ/ (1.44) ($t_{(50)} = 1.44, p > 0.05$). Words were spoken by the male native English speaker in the neutral carrier sentence “Now I say...”. The speaker read the carrier sentences with the experimental words one by one, separated by a pause, in a clear citation style. The recording was made in a soundproof booth using a high quality Sennheiser microphone, recorded onto a computer at a 44.1 kHz sampling rate. In addition to the auditory utterances, one picture of each word was selected. To ensure that the pictures were correctly identified, the printed word was added to the picture file below the original picture.

2.3. Procedure

All participants first performed the lexical decision task, then the word identification task, and finally the categorization task, in a single test session. This order was fixed for all participants to prevent them from being aware of the experimental phonetic contrast manipulated in the lexical decision task. However we report the procedures and results in ascending order of difficulty and lexical involvement, for ease of discussion, i.e. we address the categorization task first, followed by the lexical decision task, and then the word identification task. Dutch participants were tested one at a time in a sound-attenuated booth at the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands). They were seated facing a monitor and a response box with two buttons. All tasks were controlled with NESU (Nijmegen Experiment Set-Up) experimental software. In all three tasks, participants received written instructions in their native language. Stimuli were presented binaurally over closed headphones at a comfortable listening level. The four native English participants recruited in the Netherlands were tested in the same controlled testing conditions. The eight English speakers who participated in the online version of the experiment performed the task in their own homes. Before participating, they were informed about the auditory nature of the study. They were explicitly asked to perform the tasks in a quiet environment without any potential distractors. In addition, they were asked to set the volume of their own computers at a comfortable level and to use headphones.

2.3.1. Task 1: categorization task

For the categorization task, participants were instructed to press the left button when hearing a vowel similar to the one contained in the English word “Ed”, and the right button when hearing a vowel similar to the one in the English word “ad”. Instructions stressed both accuracy and speed of responses. The task began with 14 practice trials, one for each stimulus, presented in random order. Each practice and experimental trial started with the presentation of an asterisk in the center of the screen for 500 ms, directly after which the stimulus was played. Participants had 2 seconds to respond before the next trial began. Fifty-five different lists were created. Each list included 12 blocks, and each block presented the 14 synthesized stimuli in semi-random order. The randomization prevented the same stimulus from being presented twice in succession. Participants could rest after every 35 trials, when a pause message was presented on the screen. The task resumed when participants pressed a response button.

2.3.2. Task 2: lexical decision task

For the lexical decision task, participants were asked, by means of written instructions, to press a green response button with their dominant hand when they heard an English word, and a red button with their non-dominant hand when they heard a non-word. Participants were asked to respond as quickly and as accurately as possible.

The task started with ten practice trials. Items were presented in a semi-random order, with the restriction that at least three filler items appeared between two experimental items. No time limit was imposed for the responses. The next item was presented 350 ms after each button press. Half of the participants were tested with one of the lists and the other half with the other.

2.3.3. Task 3: word identification task

Before the word identification task began, participants read the instructions. The task consisted of two blocks separated by a pause. The 102 words were randomly assigned to the blocks with the constraint that only one of the counterparts of each minimal pair was presented in a given block of 51 stimuli. The presentation of the two blocks was randomized. In total, 50 different lists were constructed.

Each trial started with the simultaneous presentation of two pictures (and their orthographical form) each referring to one counterpart of a minimal /æ/-/ɛ/ pair. After 450 ms, the English sentence was played. Participants had 3 seconds to respond before the next trial started. Participants were asked to push the button corresponding to the spatial location of the matching picture on the screen (either left or right).

3. Results

3.1. Task 1: categorization task

As participants' responses for the categorization task on each trial were discrete (i.e. either “ɛ” or “æ” responses), a log odds ratio transform¹ was applied on the average proportion of “æ” responses for each participant (Dixon, 2008; Jaeger, 2008). Higher log odds indicate more “æ” responses. The log odds were submitted to a repeated measures ANOVA, with the within-participants factors “Step” (each of the 7 continuum steps), “Duration” (ɛ-type and æ-type) and the between-participants factor “Language group” (Dutch and English). As displayed in Fig. 1, there was a main effect of groups of participants ($F_{(1,65)} = 24.71; p < 0.001$), with the Dutch listeners overall giving -0.17 /ɛ/-responses and native listeners -0.84 . There was also the expected effect of “Step” ($F_{(6,390)} = 112.91; p < 0.001$), with more /ɛ/-responses the more /ɛ/-like the stimulus. Crucially, the “Language group” and “Step” factors interacted significantly ($F_{(6,390)} = 21.62; p < 0.001$). This reflects the steeper categorization slopes for the English than for the Dutch listeners. T-tests comparing the two language groups for each continuum step revealed significant differences on all stimuli (all p -values < 0.01) except on s5 ($t_{(65)} < 1$). Nevertheless, Dutch participants showed sensitivity to the difference between the continuum-extreme stimuli (average log odds of “æ” responses for s1 and s2 compared to s6 and s7: $t_{(54)} = 7.51, p < 0.001$). Not surprisingly, English participants

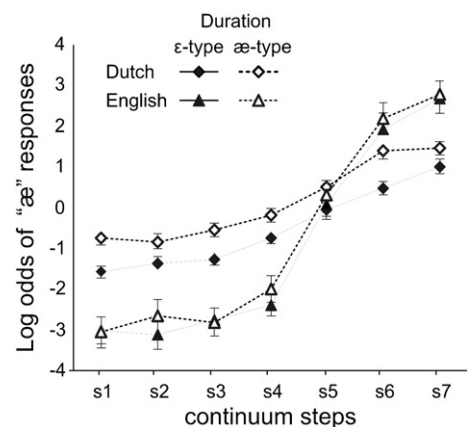


Fig. 1. Log odds of mean “æ” responses for each of the 7-step stimuli in the /ɛ/-/æ/ continuum for Dutch and English listener groups separately for each duration type. Higher log odds indicate more “æ” responses. Bars depict standard error.

also categorized the continuum-extreme stimuli differently ($t_{(11)} = 21.45, p < 0.001$). The factor “Duration” was also significant ($F_{(1,65)} = 7.69; p < 0.01$) and did not interact with any other factor. Participants gave more æ-responses when the tokens had the prototypical duration of the English vowel /æ/ when compared to the tokens with the prototypical duration of the English vowel /ε/.

To assess individual differences in non-native language perception, a performance score was calculated for each participant following Sebastián-Gallés and Baus (2005). The categorization score was computed by averaging the log odds of æ-responses for the continuum steps 1 and 2, on the one hand, and for the continuum steps 6 and 7, on the other, pooling over the two durations. The average log odds of steps 1 and 2 was subtracted from the average of steps 6 and 7. Note that this score can be interpreted as a perceptual distance between the endpoints in terms of signal detection theory (d', Macmillan & Creelman, 1991). High positive scores reflect a good separation of /ε/ and /æ/. Scores close to zero reflect that participants did not respond differently to steps 1 and 2 than to steps 6 and 7, while negative scores indicate that participants' responses showed a reverse pattern. The results showed a great variability among participants (Fig. 2). In order to evaluate individual L2 performance, native-like performance based on the scores of the group of English native listeners was established. In previous studies, the native performance range was calculated by subtracting two or three standard deviations from the natives' mean (Flege, MacKay et al., 1999; Munro et al., 1996; Sebastián-Gallés & Baus, 2005). In the present study, as in Sebastián-Gallés and Baus (2005), given the difficulty of the tasks for non-native participants, it was decided to use three standard deviations below the natives' mean. This calculation yielded a cut-off point of 2.76 (natives' mean: 5.36, sd = 0.86). Twenty-four Dutch participants (43.63% of the total population) scored above this value.

3.2. Task 2: lexical decision task

Prior to analysis of the responses for the lexical decision task, outlier items were determined based on the results of the English native listeners by subtracting three standard deviations (0.21) from the mean of the natives' proportion of correct responses (0.88, note that this value is comparable to other studies with similar phonetic manipulations in lexical decision tasks: Sebastián-Gallés et al., 2005; Broersma & Cutler, 2008, 2011; Cutler & Broersma, 2005). This calculation yielded a value of 0.24. One filler nonword, two æ-nonwords and three ε-nonwords did not reach that value of correct responses and, therefore, they and also their word counterparts were not included in subsequent analyses. First, proportions of correct responses to filler items (i.e., control) were submitted to analysis (see Fig. 3). An ANOVA including the factors “Lexicality”

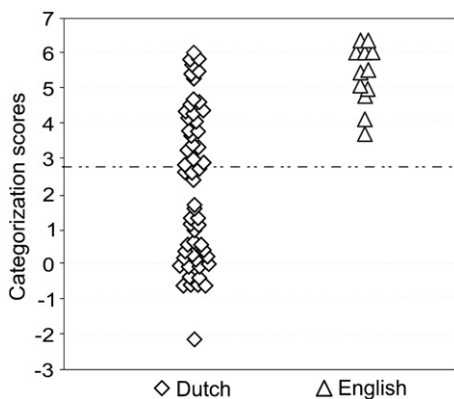


Fig. 2. Individual scores of Dutch and English participants in the categorization task. The slashed-dotted line represents the native threshold (3 sd below mean of native scores).

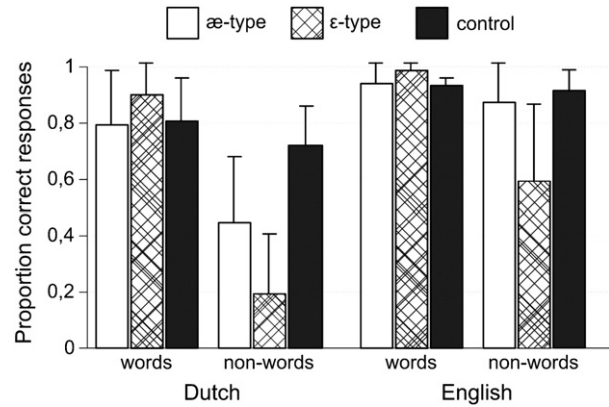


Fig. 3. Proportion of correct responses as a function of language group for the different stimulus types in the lexical decision task. Bars depict standard deviations.

(word and non-word) and “Language group” (Dutch and English) showed differences between the groups of participants ($F_{(1,65)} = 10.99, p < 0.001; F_{2(1,133)} = 92.78, p < 0.001$). English participants were more accurate than Dutch participants in judging the lexical status of the filler items (English: 0.92, Dutch: 0.79). Neither the factor “Lexicality” nor the interaction between the two factors was significant (all $F > 1$). Importantly, Dutch participants' responses were above chance level (0.5) for both filler words ($t_{(54)} = 13.18, p < 0.001$) and filler non-words ($t_{(54)} = 13.40, p < 0.001$).

Second, for the experimental items, Dutch participants accepted the majority (70%) of the experimental non-words as real words (see Fig. 3). Following previous studies (Sebastián-Gallés, Vera-Constan, Larsson, Costa, & Deco, 2009; Sebastián-Gallés et al., 2005, 2006), accuracy analyses were carried out by using the A' statistic. The A' statistic is a non-parametric unbiased index of sensitivity with 0.5 corresponding to chance performance and 1.0 to perfect discrimination (McNichol, 1972)—in this case of experimental words and non-words. Table 1 shows mean A' scores for each language group.

Separate repeated-measures ANOVAs by subjects and by items were carried out on A' scores with the factors “Stimulus type” (æ-type and ε-type) and “Language group” (Dutch and English). The analyses revealed significant differences between the two groups of participants ($F_{(1,65)} = 72.89, p < 0.001; F_{2(1,25)} = 47.16, p < 0.001$). Again, Dutch performance was poor compared to the native English listeners (see Table 2). The “Stimulus type” factor was also significant ($F_{(1,65)} = 5.52, p > 0.05; F_{2(1,25)} = 26.34, p > 0.001$), revealing higher accuracy in responses to æ-type than to ε-type items (see Table 1). The interaction between the two factors did not reach significance (both F_1 and $F_2 < 1$).

To evaluate individual performance, each Dutch and English participant's A' scores for æ-type and ε-type stimuli were considered (see Fig. 4). Again, the pattern of native performance was calculated based on the scores of English listeners. A cut-off point of three standard deviations below the mean (see Table 1) was calculated for each stimulus type. This calculation yielded an A' value of 0.80 for the æ-type stimuli and of 0.71 for the ε-type stimuli. When taking into account the participants' performance for the two vowels, only seven Dutch listeners (12.72% of the participants) scored within the native range for both stimulus types.

Table 1 Mean A' for each group of participants and each stimulus type in the lexical decision task (standard errors in parentheses).

	Dutch	English
æ-type	0.74 (0.11)	0.95 (0.05)
ε-type	0.66 (0.14)	0.89 (0.06)

Table 2
Proportion of correct responses for each group of participants and each stimulus type in the word identification task (standard errors in parentheses).

	Dutch	English
æ-word	0.61 (0.01)	0.80 (0.01)
ε-word	0.87 (0.03)	0.99 (0.02)

3.3. Task 3: word identification task

Participants responded to almost all of the trials in the word identification task. Missing responses (0.49%) were considered incorrect responses. Items were discarded from statistical analysis if the rate of correct identification by the native English listeners was three standard deviations (0.17) below the mean (0.91). From the 102 items, four æ-words (alimentary, expand, marry, cattle) were discarded. No ε-words were discarded, except those whose /æ/ counterparts scored below the limit described above. These items were excluded from the analyses for both groups of participants. Table 1 shows the percentage of correct responses for each group and word type.

For statistical comparisons, the proportion of correct responses was transformed to log odds¹ (Dixon, 2008; Jaeger, 2008). Separate ANOVAs by subjects (F_1) and items (F_2) were carried out with the factors “Word type” (æ-word and ε-word) and “Language group” (Dutch and English). As shown in Fig. 5, English listeners performed the task more accurately than Dutch listeners ($F_{1(1,65)} = 126.10, p < 0.001$; $F_{2(1,92)} = 368.24, p < 0.001$). In addition, there was a main effect of the factor “Word type” ($F_{1(1,65)} = 109.34, p < 0.001$; $F_{2(1,92)} = 93.05, p < 0.001$); ε-words were better identified than æ-words. Moreover, there was a significant interaction between the factors “Word type” and “Language group” ($F_{1(1,65)} = 10.26, p < 0.05$; $F_{2(1,92)} = 6.66, p < 0.05$). Post-hoc comparisons showed that both groups of participants (i) performed more accurately on ε-words than on æ-words (Dutch: $t_{(54)} = 9.11, p < 0.001$, English: $t_{(11)} = 6.00, p < 0.001$) and (ii) differed from one another in identifying both ε- and æ-words ($t_{(65)} = 10.05, p < 0.001$ and $t_{(65)} = 7.83, p < 0.001$ respectively). The interaction between the factors “Word type” and “Language group” can hence be attributed to the stronger asymmetric pattern for English than for Dutch participants in the identification of each word type (Fig. 5)².

Individual performance was evaluated by averaging the log odds of correct responses, for each word type separately. Again, the performance pattern of Dutch participants was characterized by great variability (see Fig. 6). The native range for each word type was determined by subtracting three standard deviations from the mean. For the æ-words (mean: 2.30, sd: 1.03), the cut-off point was -0.78 . For the ε-words (mean: 4.40, sd: 0.34), the native threshold was established at 3.36. All Dutch participants performed the task within the native range for the æ-words. From the 55 Dutch participants, 5 (9.09%) performed the task within the native range for the ε-words as well.

4. Discussion

We tested how well Dutch late L2 learners of English perceive and make use of the English contrast between the vowels /æ/ and /ε/

across three tasks: a categorization task, a lexical decision task, and a word identification task. In all tasks, Dutch listeners' performance was below that of English native listeners and was characterized by high variability.

For the categorization task, Dutch listeners (as a group) did not categorize the stimuli as accurately as the native English listener group did, as shown by the Dutch listeners' shallower categorization curves (Fig. 1). This result confirms the difficulties that Dutch listeners experience in perceiving the English contrast /æ/-/ε/ (Schouten, 1975). Nonetheless, Dutch participants were sensitive to the phonetic differences between the unambiguous stimuli (s1–s2 and s6–s7). Interestingly, Dutch listeners (as well as native listeners) exploited both frequency and duration cues for categorizing the stimuli. Previous findings have shown that Dutch listeners use vowel duration as a cue to categorize final fricative voicing in English (Broersma, 2008, 2010). The present data extends previous findings by showing that Dutch listeners exploit duration cues also during discriminating difficult non-native vowel contrasts, as /æ/-/ε/. For the lexical decision task, Dutch participants performed poorly compared to native English listeners. Nevertheless, Dutch participants demonstrated a fairly good lexical knowledge, as assessed by their performance on the filler stimuli. For the word identification task, Dutch participants were less accurate than the native English listeners in identifying words with the respective vowel sounds.

In addition, the results of the two tasks involving lexical access, lexical decision and word identification tasks, were both characterized by asymmetric patterns of performance for both Dutch and English listeners. For the lexical decision task, participants performed better on the æ-type stimuli than on the ε-type stimuli: the non-words with /ε/ embedded (as in the æ-non-word “lemp”), were less times misidentified as real English words than the non-words with /æ/ (as in the ε-non-word “dask”). For the word identification task, participants showed a better identification rate of words containing the phoneme /ε/ than of words containing /æ/ suggesting that ε-words tend to win the lexical competition process from minimally different æ-words. Thus, in this task, hearing either ‘cattle’ or ‘kettle’ led relatively often to a ‘kettle’ interpretation. The asymmetries observed in both tasks indicate that words containing the English vowel /ε/ are relatively strongly activated during lexical retrieval. This suggests that the asymmetry occurs at later stages of lexical processing.

The same asymmetry in lexical access of English words containing either the vowel /æ/ or /ε/ has been reported previously for both Dutch listeners (Escudero, Hayes-Harb, & Mitterer, 2008; Weber & Cutler, 2004) and native English listeners (Broersma, 2005b). Weber and Cutler (2004) proposed that the existence of a native Dutch category /ε/ (similar to the English vowel /ε/) prevents Dutch listeners from misidentifying this vowel as the L2 category /æ/. However, the L2 category /æ/ has no homologue category in the participants' L1 and, therefore, it is easily misidentified as an /ε/. This claim is supported by Cutler, Weber, and Otake (2006), who found the same asymmetric pattern in the lexical access of Japanese listeners for English words involving the contrast /r/-/l/. In this study, distractors caused less interference for Japanese listeners (as assessed by eye monitoring) in the recognition of target pictures that contained the L2 category /l/ (which has a homologous category in Japanese) than in the recognition of target pictures that contained an /r/. In addition, asymmetries in native language processing are frequent (Polka & Bohn, 2003). We propose that the asymmetry found here may be due to the higher frequency of /ε/ than of /æ/ in English: /æ/ is less frequent in English and it does not exist in the Dutch vowel repertoire. According to CELEX (Baayen et al., 1995), the mean lemma frequency of words containing /ε/ is higher than that of words containing /æ/ (0.34 vs. 0.26), while the number of words is similar (namely 14,006 vs. 14,230). This asymmetry in the lexical frequency of the vowels is therefore mirror in the perceptual system. Yet, a potential contribution of the speaker's idiosyncratic pronunciation of /æ/ and /ε/ cannot be completely ruled out. However, it has to be noted that in the categorization task, no asymmetry was

¹ Log odds = $\ln[p/1-p]$, p represents percentage of correct responses and \ln represents natural log.

² It is important to note that this data set shows the importance of logistic transformation of identification results in data analysis (Dixon, 2008). In the raw percentage data, the bias towards /ε/ responses seems stronger in the Dutch L2 group (87%–61% = 26%, see Table 1) than in the English L1 group (99%–80% = 19%). Given their higher accuracy, however, the English group could clearly not have produced the same /ε/-bias as the Dutch group: with 80% correct on /æ/-trials, they would have to produce 106% correct responses on /ε/-trials. Conceptually, the English group seems to indeed have the stronger bias. There were hardly any trials in which an /ε/-word was mistaken as an /æ/-word; in fact only three English participants did not have 100% correct responses on the /ε/-trials (see Fig. 4). The logistically transformed data set captures this, and shows a stronger bias toward /ε/ for the English participants.

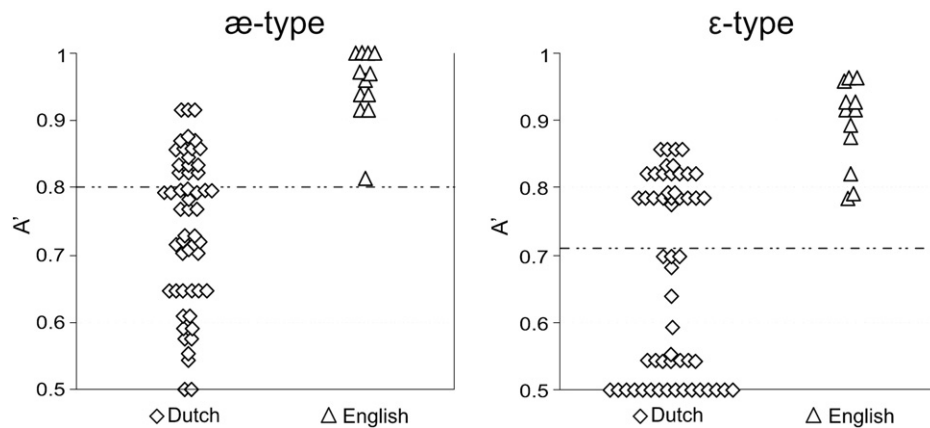


Fig. 4. Individual scores for Dutch and English participants in the lexical decision task. The slashed-dotted line represents the native threshold.

found (again, mirroring the Sebastián-Gallés & Baus, 2005, and Broersma, 2005a, results). Thus, in the absence of lexical information, participants do not perform differently across these phonemes.

As for the issue of individual differences, Dutch participants varied widely in their performance in each task. Table 3 shows the percentage of Dutch participants for each task whose scores fell within the range of 3 standard deviations below the native English listeners' average³, a generous interpretation of "nativelike" performance on English. For the categorization task, only 43.63% of the participants categorized the stimuli within the range of native listeners. This is a relatively low percentage as compared with the results obtained by Sebastián-Gallés and Baus (2005), who tested early Spanish-Catalan bilinguals in a categorization task, a gating task and a lexical decision task. In that study, approximately 68% of Spanish-Catalan bilinguals fell within the range of Catalan-natives in a categorization task, a percentage clearly above the present one. For the lexical decision task, the evaluation of individual performance again revealed a large variability in the non-native group ranging from chance (0.5) to native-like level, although few participants (12.72%) performed within the native range. This is similar to the range observed in Sebastián-Gallés and Baus (2005) study, therefore confirming the difficulty of this task for non-native listeners. For the word identification task, only a small percentage of Dutch participants succeeded in identifying the words in the accuracy range of the native listeners (9.09%). As this task was not employed in Sebastián-Gallés and Baus (2005) study, no similar results are available for comparison.

When comparing Dutch participants' performance across the three tasks, different patterns are observed. Table 4 displays the native-like pattern distributions across the three tasks. Almost half of the Dutch listeners (47.27%) failed to perform within the native range in any of the three tasks. A third of the participants (32.72%) was able to categorize the synthesized stimuli of the /æ/-/ε/ continuum within the native listeners' range, but did not manage to perform within the native listeners' range in either of the lexical tasks. Few Dutch participants scored within the native range in two of the three tasks (10.9%), and not a single Dutch listener performed within the native listeners' range in all three tasks. These results closely resemble those

reported by Sebastián-Gallés and Baus (2005). The results of Sebastián-Gallés and Baus (2005) showed a monotonic decrease in non-native performance across the three tasks as the task relied more strongly on lexical representations. A high percentage of participants (68.3%) performed the categorization task within the range of a native group. The gating task showed an intermediate level of native-like performance (46.6%). Finally, the percentage dramatically decreased for the lexical decision task (18.3%). Our current main objective was to investigate whether a similar pattern of results can be obtained with later-onset bilinguals than those who participated in the study of Sebastián-Gallés and Baus (2005). Even though the overall level of performance by the bilinguals in that study was higher than in the current one, there was a similar trend over three different tasks which evaluated the phonological processing of a difficult L2 contrast; that is, high competence of the non-native listeners in the categorization task, which decreased in the lexical decision task and the word identification task. Thus, late bilinguals' L2 perception seems to be constrained by task demands in a similar way as that of early bilinguals.

The close resemblance of the two data sets suggests that the pattern of good performance on phonetic tasks and worse performance on lexical tasks may be the general pattern for a wide variety of L2 learners. Indeed, a similar experimental strategy was used to explore the learning of suprasegment knowledge (the perception of stress patterns) in late French-Spanish bilinguals by Dupoux, Sebastián-Gallés, Navarrete, and Peperkamp (2008). In that study, the performance of late French-Spanish bilinguals also showed that all late bilinguals had great difficulties in the use of stress to access the lexicon. The convergent pattern of individual performance on L2 tasks in the present and previous L2 studies (i.e., L2 listener's higher accuracy on acoustic-phonetic analysis tasks than on tasks involving lexical processes) is consistent with the processing hierarchy proposed by several models of speech perception

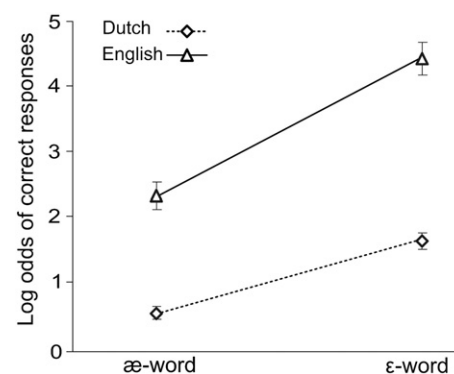


Fig. 5. Log odds of mean correct identifications in the word identification task for each word type for Dutch and English speakers. Bars depict standard errors.

³ There were no major differences in the performance of the native listeners depending on whether they were tested in England or abroad. In a repeated measures ANOVA with the factors "Task" (categorization, word identification, and lexical decision tasks), "Vowel" (æ and ε), and "Group" (living in England and living in a non-English speaking country) only the interaction "Group" × "Task" was significant ($F(2,20) = 5.82, p = 0.01$). Nevertheless, post-hoc two-sample *t*-tests only revealed a marginal difference between the two groups for the lexical decision task on ε-words ($t(10) = 2.18, p = 0.054$, for all other comparisons $p > 0.1$). In addition, no one of the participants tested at their own home, and, therefore, in potentially less favorable listening conditions, systematically performed the tasks at a lower level than the other participants.

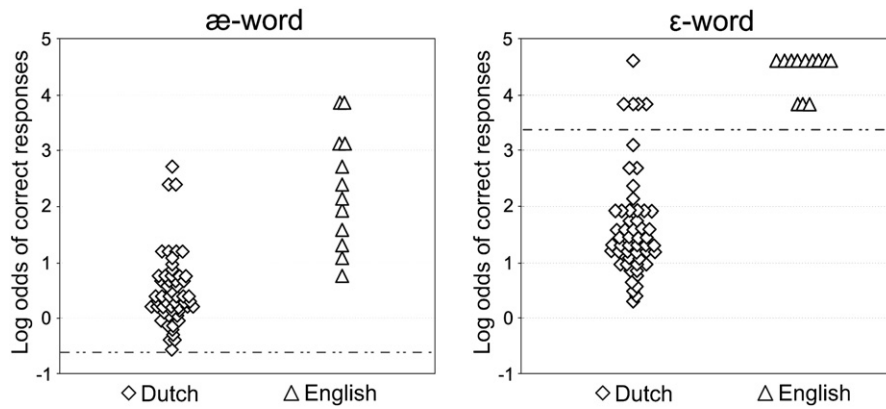


Fig. 6. Individual scores of Dutch and English participants for æ- and ε-words in the word identification task. The slashed-dotted line represents the native threshold.

(Hickok & Poeppel, 2007; McClelland & Elman, 1986, among many others). According to these models speech perception is carried out in separate and hierarchical stages: Initially phonemes are analyzed, the resulting outcome of phoneme analysis is then used to select the corresponding lexical entries, and based on the selected lexical forms the semantic and syntactic relations are analyzed to extract the linguistic meaning of the speech signal. The fact that non-native speech processing is differentially constrained at the different processing stages studied here (phonology and lexicon) suggests that the architecture of L2 perceptual processes, even in late L2 acquisition, is organized in a similar manner as native language perception processes.

In the Sebastián-Gallés and Baus (2005) study, 12% of the bilingual listeners consistently showed native-like performance across the three tasks; in the present study, we were unable to identify any such bilingual listener. This difference may have several potential sources. Whereas Sebastián-Gallés and Baus (2005) tested early bilinguals, our participants started L2 acquisition later in life (at 11 years of age, on average). Whereas their participants grew up in a bilingual society, ours lived in a monolingual society, although at the moment of the testing, the participants had significant exposure to the second language. The different results might also be due to the different tasks employed, as well as the materials and their unique acoustic-phonetic properties. Indeed, the gating task employed by Sebastián-Gallés and Baus (2005) was here replaced by a word identification task, heavily depending on the existence of appropriate lexical representations. This task proved to be even more difficult to perform than the lexical decision task. Finally, different non-native vowel contrasts and bilingual populations from different languages were tested in the present study (/æ/-/ε/ in Dutch-English bilingual) versus the Sebastián-Gallés and Baus (2005) study (/ε/-/e/ in Spanish-Catalan bilinguals). In spite of these relatively minor differences, there are important similarities between both studies. In particular, in spite of differences in testing conditions between the current study and Sebastián-Gallés and Baus (2005), native performance was very high in both cases. This result clearly indicates the robustness of native listening, as long as participants are tested in relatively quiet conditions.

The current study replicates the finding by Sebastián-Gallés and Baus (2005) that the performance of the bilinguals decreased in comparison to the native listeners as the tasks had greater lexical involvement. The current study is nevertheless more than a mere replication,

as the same result was obtained with a group of late(r) bilinguals. There are two ways to explain this apparently quite general decrease in performance over tasks. First, it may simply be the case that the lexical decision and the word identification tasks are more taxing: They employ natural speech, which is physically highly variable even within listeners, and the critical vowels were presented together with other speech sounds. Therefore, the complexity of these tasks would better reveal the difficulties of L2 learners in comparison to native listeners. However, if the difficulty of the task was the cause of the different proficiency levels, one would also expect a higher percentage of Dutch listeners performing within the native range in the word identification task as compared to the lexical decision task. In terms of difficulty, the word identification task is easier than the lexical decision task because participants are presented with the orthographic form of the words before the auditory word is presented. The spelling out of the two potential candidate words in the word identification task made participants aware of the experimental manipulation and allowed participants to determine at exactly what point within each word the target speech sound would appear.

Second, it may be the case, as we argued in Introduction, that there are different phonological processes (Pisoni & Luce, 1987). We evaluated this question in two ways. First of all, we tested whether the performance on the different tasks correlated (see Table 5). The results seem to show that pre-lexical and lexical phonological processing are separable. The two “extreme” tasks, the categorization and the lexical decision tasks, one tapping pre-lexical and the other lexical processing respectively, do not correlate with each other at all. The word identification task seems to be intermediate between lexical and pre-lexical phonological processing, as performance on this task correlates with the performance on both the categorization and the lexical decision tasks (see Table 5). This intermediate position is supported by an analysis of this task: just as in the categorization task, participants

Table 3
Percentage of Dutch speakers who scored within the range of English native listeners for each perception task.

Categorization	Lexical decision	Word identification
43.63%	12.72%	9.09%

Table 4
Native-like performance distributions of Dutch speakers across the three tasks. The different performance patterns that could be potentially observed are displayed (crosses represent failure to achieve nativelike performance, and checkmarks represent native-like performance) and the actual percentage of L2 listeners whose performance follows the given pattern.

Categorization	Lexical decision	Word identification	% of participants
X	X	X	47.27%
✓	X	X	32.72%
X	✓	X	5.45%
X	X	✓	1.81%
✓	✓	X	5.45%
✓	X	✓	5.45%
X	✓	✓	1.81%
✓	✓	✓	0%

Table 5

Correlations between the tasks for the Dutch and English speakers. For the word identification and the lexical decision tasks, participants' responses were averaged for the two different stimulus types (/ɛ/ and /æ/ sets). Statistical significant correlations are indicated by asterisk (* = $p < 0.05$, ** = $p < 0.001$).

	Lexical decision	Word identification
Categorization	Dutch: $r = -0.05$ English: $r = 0.08$	Dutch: $r = 0.42^{**}$ English: $r = -0.32$
Word identification	Dutch: $r = 0.27^*$ English: $r = 0.38$	

have to decide between two alternatives. Because both alternatives are lexical items, it requires the involvement of lexical processing, which then leads to the bias towards /ɛ/ also observed for the lexical decision task (see below for a discussion of this asymmetry). A second possibility to approach the issue of separable phonological skills is to investigate how many bilinguals performed within and below native-listener limits on the different tasks. If the lexical decision task were simply a more difficult task than the categorization task, all participants who performed well on the former should also perform well on the latter. This was, however, not the case. Of the about 12% of bilinguals who performed within the native listener limits for the lexical decision task, less than half also performed well on the categorization task. This pattern of performance suggests that it is indeed possible to distinguish phonological processing at the pre-lexical versus lexical levels.

A relevant question is why fewer late bilinguals manage to score within the performance range of native listeners in the word identification task than in the lexical decision task. In the word identification task, participants were presented with the orthographic form of the words before the auditory word was presented and they could anticipate the exact temporal point at exactly the target speech sound would appear. However, the presentation of the written words seems not to decrease the difficulty implied in lexical tasks. One potential reason is that lexical competition in the word identification task was maximal because participants had to choose one of two words (whereas in the lexical decision task the words competed with non-words). In addition, in the word identification task, the maximum competitor word of the heard word was written on the screen. This could increase the lexical activation for the competitor and, consequently, make difficult the correct selection of the lexical entry corresponding to the heard word.

The present study shows large variability between late bilinguals in their mastery of the L2 phonology. What is the origin of these individual differences? A previous study on early bilinguals has shown that exceptionally good and poor perceivers of an L2 phonological contrast differed in their discrimination of a native phonological contrast: The good perceivers showed a more accurate discrimination of the native phonemes (Díaz et al., 2008). This previous finding indicates that, when relevant factors as L2 age of acquisition or exposure are similar, individual differences in L2 proficiency are caused by a general language mechanism. Yet, no previous study has compared native phonological discrimination in late bilinguals who, as in Díaz et al. (2008), maximally differ in their L2 phonological capabilities.

This is the first study to systematically evaluate the mastery of a difficult L2 contrast in late bilinguals. We have provided evidence of i) strong individual variability in the mastery of an L2 that cannot be explained by factors such as age of acquisition or general language environment, and ii) distinct proficiency levels as a function of the phonological processes involved in the task. These results imply that language teaching programs, regardless of the age of the students, must adapt to the students' learning pace and that successful phonetic training must cover a wide range of phonological tasks.

Acknowledgments

This work was supported by a Pre-doctoral fellowship from the Spanish Government to Begoña Díaz, and a Veni grant from the

Netherlands Organisation for Scientific Research (NWO) to Mirjam Broersma. This research was supported by grants from the Spanish Ministerio de Ciencia e Innovación (SEJ2009-09072; Consolidation-Ingenio2010-CDS-2007-00012) and the Catalan Government (SGR 2009-1521). Núria Sebastián-Gallés received the prize "ICREA Acadèmia" for excellence in research, funded by the Generalitat de Catalunya.

References

- Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from a new paradigm. *Developmental Science*, 14, 395–401.
- Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The Celex lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). Baltimore: York Press.
- Best, C., & McRoberts, G. (2003). Infant perception of nonnative consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46, 183–216.
- Best, C., McRoberts, G., LaFleur, R., & Silver-Issenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behavior & Development*, 18, 339–350.
- Best, C. T., McRoberts, G., & Sithole, N. M. (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, 45–60.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Bongaerts, T. (1999). Ultimate attainment in foreign language pronunciation: The case of very advanced late foreign language learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Erlbaum.
- Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12, 189–222.
- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46, 217–243.
- Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, 117, 3890–3901.
- Broersma, M. (2005). *Phonetic and lexical processing in a second language*. MPI series in Psycholinguistics. Max Planck Institute.
- Broersma, M. (2008). Flexible cue use in nonnative phonetic categorization. *Journal of the Acoustical Society of America*, 124, 712–715.
- Broersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *Journal of the Acoustical Society of America*, 127, 1636–1644.
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System: An International Journal of Educational Technology and Applied Linguistics*, 36, 22–34.
- Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, 64, 74–95.
- Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 54, 421–428.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372–387.
- Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. Hardcastle, & J. Beck (Eds.), *A figure of speech* (pp. 63–91). Mahwah, NJ: Erlbaum.
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34, 269–284.
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116, 3668–3678.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499–533.
- Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 16083–16088.
- Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59, 447–456.
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, 106, 682–706.
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36, 345–360.
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–272). Baltimore, MD: York Press.
- Flege, J. E., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- 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, 2973–2987.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104.

- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.
- Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.
- Macmillan, N. A., & Creelman, D. (1991). *Detection theory: A user's guide*. Oxford: Blackwell Publishers.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McNichol, D. (1972). *A primer of signal detection theory*. London: Allen & Unwin.
- Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, 17, 313–334.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–B17.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445–449.
- Pisoni, D., & Luce, P. (1987). Acoustic–phonetic representations in word recognition. *Cognition*, 25, 21–25.
- Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221–231.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421–435.
- Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning*, 44, 681–734.
- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 18, 2030–2048.
- Schouten, M. E. H. (1975). Native-language interference in the perception of second-language vowels: An investigation of certain aspects of the acquisition of a second language. Unpublished doctoral dissertation. The Netherlands: Utrecht University.
- Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 279–292). New York: Erlbaum.
- Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*, 12, 874–887.
- 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., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, 18, 1277–1291.
- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72, 111–123.
- Sebastián-Gallés, N., Vera-Constán, F., Larsson, J. P., Costa, A., & Deco, G. (2009). Lexical plasticity in early bilinguals does not alter phoneme categories: II. Experimental evidence. *Journal of Cognitive Neuroscience*, 21, 2343–2357.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25.
- Weber-Fox, C. M., & Neville, H. J. (1996). Maturation constraints on functional specialization for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231–256.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, 7, 49–63.
- Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. H. Yeni-Komshian (Ed.), *Child phonology, perception* (pp. 185–215). New York: Academic Press.