

# The interrelation between acoustic context effects and available response categories in speech sound categorization<sup>a)</sup>

Titia Benders<sup>b)</sup>

*Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 210,  
1012 VT Amsterdam, The Netherlands*

Paola Escudero

*MARCS Laboratories, University of Western Sydney, Locked Bag 1797, Penrith, NSW, 2751, Australia*

Matthias J. Sjerps

*Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands*

(Received 18 July 2011; revised 30 January 2012; accepted 2 February 2012)

In an investigation of contextual influences on sound categorization, 64 Peruvian Spanish listeners categorized vowels on an /i/ to /e/ continuum. First, to measure the influence of the stimulus range (broad acoustic context) and the preceding stimuli (local acoustic context), listeners were presented with different subsets of the Spanish /i/-/e/ continuum in separate blocks. Second, the influence of the number of response categories was measured by presenting half of the participants with /i/ and /e/ as responses, and the other half with /i/, /e/, /a/, /o/, and /u/. The results showed that the perceptual category boundary between /i/ and /e/ shifted depending on the stimulus range and that the formant values of locally preceding items had a contrastive influence. Categorization was less susceptible to broad and local acoustic context effects, however, when listeners were presented with five rather than two response options. Vowel categorization depends not only on the acoustic properties of the target stimulus, but also on its broad and local acoustic context. The influence of such context is in turn affected by the number of internal referents that are available to the listener in a task. © 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.3688512>]

PACS number(s): 43.71.Es, 43.71.Sy, 43.66.Lj [JMH]

Pages: 3079–3087

## I. INTRODUCTION

Speech sound categorization tasks are commonly administered to establish where on a continuum two or more categories are situated. This can be observed from the location of the classification boundary between the categories. The repeated presentation of stimuli and the limited number of response options are mostly considered to be unfortunate prerequisites for making reliable observations about category boundaries. However, at least three factors affect the classification of a stimulus: the stimulus itself, but also the surrounding stimuli and the internal referents to which listeners consider to map the acoustic input (Sawusch and Nusbaum, 1979; Restle, 1987). The current paper investigates the influences of each of these three factors and their interrelations on speech sound categorization. While the influence of stimulus properties, acoustic context, and considered internal referents has been well documented and several explanations for each of them have been put forward (see our review below), we argue that neither of those factors can be fully understood when studied in isolation. Studying the relation

between these three factors is therefore an important objective in the present study.

The influence of surrounding stimuli, or sensory context, in categorization has been documented for all sensory modalities (Kluender *et al.*, 2003, and references therein). In a simple categorization task, the influence of broad sensory context can be observed as the dependence of target perception on the range of values of the test set. The same stop-consonant stimulus is, for instance, more likely to be perceived as voiceless, i.e., having a long Voice Onset Time (VOT), when presented as part of a short-lag VOT continuum ranging from 15–35 ms, than when presented as part of a long-lag 25–45 ms range (Brady and Darwin, 1978). And when one endpoint sound, i.e., the anchor, of an /i/-/ɪ/ continuum is presented four times as frequently as the other sounds from the continuum, listeners categorize more sounds on the continuum as the non-anchor category (Sawusch and Nusbaum, 1979). This effect is not restricted to the anchor's own category boundary (Morse *et al.*, 1976) and is maintained when listeners are explicitly told that one sound will occur more often (Sawusch and Nusbaum, 1979). These findings show that the perception of any single stimulus depends on the other constituents of the stimulus set. The first aim of the current study is to replicate this influence of the acoustic values of the stimulus set, i.e., the stimulus range, on category boundary locations.

It has also been shown that the acoustic properties of the stimuli immediately surrounding a target exert a strong influence on categorization. The average formant frequency of a

---

<sup>a)</sup>Portions of this work were presented as “The interrelation between the stimulus range and the number of response categories in vowel categorization,” Proceedings of Interspeech 2010, Makuhari, Japan, September 2010, and “The interrelation between the stimulus range and the number of response categories in vowel categorization,” 2nd Pan-American/Iberian Meeting on Acoustics, Cancun, Mexico, November 2010.

<sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: [titia.benders@uva.nl](mailto:titia.benders@uva.nl)

precursor sentence influences target categorization in that a target sound from an /i/-/ɛ/ continuum is more likely to be perceived as a vowel with a relatively low *F1* (i.e., /i/ or /ɛ/) after a sentence with a generally high *F1* and as a vowel with a relatively high *F1* (i.e., /æ/ or /ʌ/) when embedded in a sentence with a generally low *F1* (Ladefoged and Broadbent, 1957; Broadbent and Ladefoged, 1960; Broadbent *et al.*, 1956; cf. Watkins, 1991; Watkins and Makin, 1994; 1996; Sjerps *et al.*, 2011, for similar effects). In addition, when multiple target stimuli are presented in a single trial, listeners tend to categorize a given stimulus into a different category than the stimuli that are played before or after (Fry *et al.*, 1962; Eimas, 1963; Thompson and Hollien, 1970; Lotto *et al.*, 1998). In the current study, we also analyze how the acoustic values of the sounds preceding the target stimulus affect its categorization.

Thus, from previous studies it appears that the acoustic context influences speech sound categorization on both broad and local scales. Because the broad acoustic context of a stimulus logically restricts its local acoustic context, it has been suggested that broad context effects are nothing more than the result of local context effects (Repp and Liberman, 1987). Kluender and Kiefte (2006) have argued that context effects can be related to the fact that listeners are sensitive to acoustic change rather than to the stable aspects in the signal. This contrast mechanism is likely to be based for an important part on general auditory processing, as speech can induce effects in the categorization of other auditory stimuli and vice versa (Holt, 2005; Stilp *et al.*, 2010). The effects of stimulus context measured on both broad and local scales can indeed be related to this single auditory-based principle of contrast as expressed by Kiefte and Kluender (2008) and Kluender and Kiefte (2006). Therefore, the third aim of the present study was to explore whether “local” contrast effects can indeed be found over a longer distance from the stimulus that is to be categorized, which is expected if broad context effects arise from a build-up of local context effects.

An alternative explanation for category boundary shifts when one anchor is presented more often or when the entire stimulus range is shifted, is that listeners have a response bias when performing a forced-choice identification task (Parducci, 1965). More specifically, it is thought that listeners in a forced-choice task will follow the (implicit) strategy to divide their responses equally over the available response options. However, boundary shifts still occur when listeners are made aware of the fact that one sound occurs more often, which is not expected if boundary shifts only result from a task strategy (Sawusch and Nusbaum, 1979). Moreover, a disadvantage of the response-bias account of boundary shifts is that it fails to capture the effect of non-speech stimuli on speech categorization and vice versa. While response biases inevitably play some role in speech categorization behavior, these observations suggest that shifts in categorization are at least for an important part the result of the more general, auditory-based, principle of contrast (Kiefte and Kluender, 2008; Kluender and Kiefte, 2006).

Sawusch and Nusbaum (1979), however, reported an interesting additional influence on speech perception that

cannot be explained with reference to auditory contrast. In their second experiment, they presented listeners with sounds on an /i/ to /ɛ/ continuum, but now the critical anchor (the stimulus presented four times more frequently than the other stimuli) was the vowel /ɛ/, which was not part of the target stimulus range and lies outside the /i/ to /ɛ/ continuum in the vowel space. A first group of listeners were not informed about the added /ɛ/ and could only respond /i/ or /ɛ/ to stimuli they would normally classify as /i/, /ɛ/, and /ɛ/. A second group received an answer sheet containing the response options /i/, /ɛ/, and /ɛ/, where /ɛ/ was already filled in for the /ɛ/-tokens, such that listeners were *de facto* classifying vowels along an /i/-/ɛ/ continuum. The second, informed, group showed a shift in their categorization responses due to the /ɛ/ anchor. The first, uninformed, group showed a bigger shift in categorization responses, so as to include the /ɛ/ anchors into their /i/ category and then shift their /i/ category toward the values normally associated with /ɛ/. Thus, it was shown that listeners' responses are not only influenced by acoustic context but also by the number of internal referents listeners are expected to consider when categorizing the acoustic input.

This finding, however, appeals to some task strategy for explaining boundary shifts, which is how Sawusch and Nusbaum (1979) interpret their result. Specifically, because listeners with two response categories had *fewer* response options than were actually present in the stimuli, the task design may have forced the listeners to deliberately divide their responses equally over the available categories, even if that led to unlikely responses. Such a task strategy on the part of the listeners is not necessary when listeners are presented with *more* response options than are present in the stimuli. If extra response options have an effect on listeners' categorization, this would thus be evidence that the considered response categories truly affect the perception of the stimuli rather than only induce a task strategy. Moreover, the aforementioned acoustic contrast mechanism may interact with the process underlying the effect of the number of response categories.

The goal of the current investigation was to comprehensively examine the contributions of three main influences on a vowel categorization decision, namely the acoustic properties of the target stimuli, the acoustic stimulus context in which the stimuli are presented (on both a broad and a local scale), and the number of response categories or internal referents that are considered during the task. We will investigate whether the distinction between context effects on a local and broad scale can be maintained, and by investigating the interrelation between the three factors, we will shed light on the mechanisms underlying boundary shifts in speech categorization tasks.

Effects of the broad stimulus context were investigated by dividing the *F1* continuum between Spanish /i/ and /ɛ/ in three sub-ranges, namely a low (lower *F1* values), an intermediate, and a high range (higher *F1* values), and presenting them to listeners in separate blocks. Over the whole range of values used, more /i/ responses were expected in the low range, and more /ɛ/ responses were expected in the high range. Due to contrast effects, however, listeners were

expected to more often categorize any individual sound as /e/ (which has a high  $F1$ ) in the context of a low range stimulus set, and as /i/ (which has a low  $F1$ ) in a high range stimulus set. Consequently, the perceptual boundaries are expected to shift toward lower  $F1$  values in the low stimulus range, and toward higher  $F1$  values in the high stimulus range.

Stimulus context effects on a local scale were investigated by examining whether the difference between the  $F1$  of the target and the preceding stimulus had an influence on the categorization of the target stimulus. An effect similar to that for broad stimulus context was expected, such that a target stimulus would be more likely to be categorized as /e/ when preceded by a stimulus that had a lower  $F1$ , and as /i/ when preceded by a stimulus that had a higher  $F1$ . Additionally, however, it was investigated whether the difference between the  $F1$  of the target stimulus and the  $F1$  of stimuli that had been presented several trials before also had an influence on the categorization of the target stimulus. This investigation could thus reveal to what extent local context effects can be dissociated from global context effects.

Finally, the influence of the internal referents that are considered as response categories was investigated by presenting listeners with different numbers of response options. One group was asked to categorize the target sounds as either /i/ or /e/, while the other group categorized the same target sounds as one of the five Spanish monophthongs /i/, /e/, /a/, /o/, or /u/. Based on the results from Sawusch and Nusbaum (1979), it was expected that listeners with two response categories would be more inclined to shift their category boundaries than the listeners with multiple response categories.

In order to maximize the size of the acoustic context effects, the experiment used a vowel categorization task, because vowels might show stronger context dependence than consonants (Cooper, 1974). We investigated these effects in Spanish, a language that has a less densely sampled vowel space than English. Spanish has only five vowel monophthongs (/i/, /e/, /a/, /o/, and /u/), which creates relatively large distances between adjacent vowels in the vowel space. Previous work by Keating *et al.* (1981) showed that VOT range effects are stronger in Polish than in English listeners. Those authors suggested that this difference may be ascribed to the larger VOT-distinction between voiced and voiceless stops in Polish than in English and the much broader (less sharply defined) voiced category in Polish. Additionally, a larger distance between categories may result in more within-category variation in the natural productions.<sup>1</sup> As vowels furthermore seem to have less sharply defined boundaries than consonants, the categorization of /i/ and /e/ by Spanish listeners was expected to be subject to relatively large acoustic context effects.

In summary, the current study tested broad and local acoustic stimulus context effects, the justification of separating broad and local context effects, and the influence of the number of response categories. Importantly, the combination of the three different factors, namely stimulus properties, acoustic stimulus context, and response categories, allowed for the examination of their interrelation in sound categorization, which will in turn deepen our understanding of the mechanisms underlying each of these factors.

## II. EXPERIMENT

### A. Participants

Sixty-four monolingual speakers of Peruvian Spanish (32 females) were tested. They were born and had spent all their lives in Lima, rated their speaking and listening abilities in English as no higher than 2 on a scale from 0–7 (0=no knowledge, 7=native speaker), and reported no knowledge of any other language than Spanish and English. They were between 18 and 28 years old and were university students at the Pontificia Universidad Católica del Perú in Lima.

### B. Materials

The stimuli were isolated synthetic vowels, which were created using a simplified version of the Klatt synthesizer (Klatt, 1980).<sup>2</sup> The 13  $F1$  values of the stimuli ranged from 281–553 Hz and the steps between stimuli were approximately equal on the Erb scale (0.31 Erb). The  $F2$  values ranged from 1893–2557 Hz and were inversely proportional to  $F1$  in hertz. Stimuli were created with 7 durational values ranging from 80–175 ms. This duration manipulation will not be addressed in the present study. Combining the 13 spectral values with the 7 durational values leads to 91 unique stimuli. The spectral values of the stimuli are displayed in Fig. 1. All stimuli had a fundamental frequency that decreased linearly from 180–140 Hz. The formant frequencies were steady throughout the vowel.

In order to examine broad scale stimulus context effects, the stimuli were divided in three  $F1$  ranges, as shown in Fig. 1: A phonetically low  $F1$  range (the 49 stimuli with  $F1$  values from 281–410 Hz), a phonetically intermediate  $F1$  range (the 49 stimuli with  $F1$  values from 344–480 Hz), and a phonetically high  $F1$  range (the 49 stimuli with  $F1$  values from 410–553 Hz). These  $F1$  values range from possible productions of /i/ in Peruvian Spanish, in the low  $F1$  range, to possible productions of /e/, in the high  $F1$  range, and the intermediate range contains  $F1$  values that best match the average productions of both vowels (Chládková *et al.*, 2011).

### C. Design and procedure

Participants listened to the stimuli over headphones and saw the orthographic representation of the response categories on a computer screen. Half of the participants had two

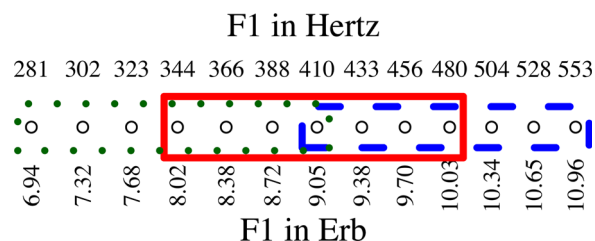


FIG. 1. (Color online) The  $F1$  values of the stimuli and ranges as used in the present experiment. The rectangles indicate the  $F1$  values that were used in each of the three range conditions: Low (dotted line), intermediate (solid line), and high (dashed line) range.

response categories, i.e., /i/ and /e/, while the other half could choose from the five Spanish vowels, i.e., /i/, /e/, /a/, /o/, and /u/. On each trial, an isolated stimulus was played once and participants had to click with the mouse on the vowel they thought they heard. All participants performed the categorization task in the three range conditions, and the six possible orders were counterbalanced across participants. For each range condition, stimuli were presented in blocks of 49 stimuli (the 7 spectral values combined with the 7 durational values) and each block was presented three times with the stimuli randomized per block. This led to 147 trials per range condition and a total of 441 trials. Prior to the first block, participants received a practice session of ten tokens and after each block they could take a short break.

The perception experiment was run on a PC laptop computer using the Praat program (Boersma and Weenink, 2011).<sup>3</sup> The experimenter was the second author, a native speaker of Peruvian Spanish.

#### D. Analysis

We examined the effect of the stimulus range, preceding stimuli, and number of response categories on listeners' categorization of sounds as either /i/ or /e/ by means of a hierarchical logistic regression analysis. Statistical analyses were conducted using the lme4 package in the open-source statistical software R (R Development Core Team, 2008). An alpha level of 0.05 was adopted for all tests.

Logistic regression models the influence of independent variables on a binary dependent variable and has been used in recent studies to analyze similar cases of speech sound categorization in native and non-native listeners (Morrison, 2007; Morrison and Kondaurova, 2009; Escudero *et al.*, 2009). Hierarchical modeling, as used in the present study, is a compromise between pooling the data of all participants in one analysis and applying a separate model on each participant's data (Gelman and Hill, 2007). In the analysis, the researcher can choose to allow some predictor coefficients to vary between participants within a pre-specified distribution. Varying the predictor coefficients over participants allows the model to capture individual differences between participants, whereas restricting the variation to a distribution captures the fact that participants are members of the same group.

In all analyses, the dependent variable was the listeners' response, for which /i/ and /e/ responses were coded as 0 and 1, respectively. The analysis tested the effects of the acoustic properties of the stimulus, the stimulus range (broad acoustic context), the preceding stimulus (local acoustic context), and the number of response categories on listeners' categorizations. Moreover, it was tested whether the number of response categories affected the effect of the broad and local acoustic context on listeners' responses. To this end, the following five main factors were included in the analysis:  $F1$  (the vowels'  $F1$  values expressed in Erb. The values for this factor were centered around 0 within each participant);  $DifferenceF1$  [the difference (in Erb) between the  $F1$  value of the target stimulus and the  $F1$  value of the preceding stimulus];  $LowR$  (the comparison between the low  $F1$  range and

the other  $F1$  ranges.  $LowR$  was coded as 1 for the low  $F1$  range and 0 for the other  $F1$  ranges.  $HighR$  was coded as 1 for the high  $F1$  range and 0 for the others);  $NRC$  (the number of response categories with two levels, coded as  $-1$  for listeners with five categories and as 1 for listeners with two categories).

The analysis also included four interactions to address the interrelation between the number of response categories available to the listeners on the one hand and the stimulus properties and the local and broad acoustic context on the other hand. These interactions were  $F1*NRC$ ,  $DifferenceF1*NRC$ ,  $LowR*NRC$ , and  $HighR*NRC$ .

The influence of each factor on the participant's response is modeled by means of a  $\beta$ -coefficient. In the hierarchical model, the coefficients of the factors that varied within participants, namely  $F1$ ,  $DifferenceF1$ ,  $LowR$ , and  $HighR$ , were allowed to vary between participants. The intercept of a (logistic) regression analysis expresses a bias for either of the two outcomes. Also the intercept was allowed to vary between participants. The coefficients for each factor must together form a normal distribution, its mean and standard deviation being estimated from the data. Together, these estimates form the first, within-subject, level of the analysis, which will be referred to as the participant level.

The factor  $NRC$  varied between participants. Therefore, the coefficient for this factor and for the interactions with this factor were fixed across participants. These factors form a second, between-subjects, level in the analysis and will be referred to as the  $NRC$  level.

The  $\beta$ -coefficient for  $F1$  ( $\beta_{F1}$ ) was expected to be positive because /e/ has higher  $F1$  values than /i/. A positive  $\beta_{DifferenceF1}$  was expected as well, since according to the contrast principle, listeners categorize a sound partially based on the difference from the preceding stimulus. The more a sound has a higher  $F1$  than the preceding sound, the more likely it is to be categorized as /e/.  $\beta_{LowR}$  was expected to be positive, as this would indicate that the perceptual boundary between /i/ and /e/ in the low range condition is placed on lower  $F1$  values than in the intermediate range.  $\beta_{HighR}$  was expected to be negative, as this would indicate that the perceptual boundary between /i/ and /e/ is placed on higher  $F1$  values in the high  $F1$  range than in the intermediate range. Regarding the interactions, a positive  $\beta_{F1*NRC}$  and a positive  $\beta_{DifferenceF1*NRC}$  would show that listeners are more sensitive to the values of the target stimulus and the difference from the preceding stimulus, respectively, if they can choose from two response categories. A positive  $\beta$ -coefficient for the interaction  $LowR*NRC$  in addition to a positive  $\beta_{LowR}$  would indicate that the perceptual boundary between /i/ and /e/ is on a lower  $F1$  value in the low range for listeners with two than for listeners with five response categories. Similarly, a negative  $\beta$ -coefficient for the interaction  $HighR*NRC$  in addition to a negative  $\beta_{HighR}$  would indicate that the listeners with two response categories have their perceptual boundary on higher  $F1$  values in the high stimulus range than listener with five response categories.

For the dependent variable, the analysis only included the responses /i/ and /e/ because the number of responses given to other vowels was too low to warrant a multinomial

analysis (/u/ = 36, /o/ = 110, and /a/ = 504). A total of 27 574 /i/ and /e/ responses were given, of which 27 510 responses were entered in the analysis. Each participant's first trial was excluded from the analysis because there was no preceding trial.

### III. RESULTS

Table I shows the results of the logistic regression models for the analysis described in Sec. II D. Below we present the results for each of the three effects considered in the aims of the present study, namely broad acoustic context (stimulus range), local acoustic context (difference from preceding stimulus), and the number of response categories.

#### A. Stimulus range effects (broad acoustic context)

As expected, the results from the analysis show that listeners choose /e/ when a stimulus has a higher  $F1$ . This is shown by a significantly positive  $\beta_{F1}$ . The results also confirm that listeners are sensitive to the broad stimulus context when categorizing stimuli: their boundary between /i/ and /e/ is on lower  $F1$  values in the low  $F1$  range and on higher  $F1$  values in the high  $F1$  range, as compared to the intermediate  $F1$  range. This is indicated by a significantly positive  $\beta_{LowR}$  and a significantly negative  $\beta_{HighR}$ , respectively. This result is visualized in Fig. 2, where the proportion of /e/ responses in the three range conditions is plotted.

To further illustrate the effect of broad acoustic context, we compared listeners' categorization of tokens with an  $F1$  value of 410 Hz (9.05 Erb) across the three stimulus range conditions. This  $F1$  value, which occurred in all three ranges, was combined with 7 durational values presented 3 times each, for a total of 21 tokens per condition. Table II gives the mean percentage of /i/ and /e/ responses per listener in the three range conditions. As can be observed, the average percentage of /i/ and /e/ responses to the tokens with an  $F1$  value of 410 Hz is approximately equal in the intermediate stimulus range condition, whereas the same stimulus is

TABLE I. Results from the main analysis, as described in Sec. II D.  $\beta$  = coefficients of the factors. Coefficient estimates (est), between-subjects variance of the coefficients varied over participants (var), coefficient standard errors (se), and  $z$ - and  $p$ -values from the logistic regression analysis are given.

	est	var	se	$z$	$p$
Level 1: participant					
intercept	-0.176	2.205	0.191	-0.921	0.357
$\beta_{F1}$	6.097	3.693	0.257	23.703	<0.0001
$\beta_{DifferenceF1}$	0.239	0.171	0.062	3.839	0.0001
$\beta_{LowR}$	4.067	4.647	0.284	14.296	<0.0001
$\beta_{HighR}$	-2.957	4.834	0.289	-10.235	<0.0001
Level 2: NRC					
$\beta_{NRC}$	0.096		0.191	0.504	0.614
$\beta_{F1*NRC}$	0.200		0.257	0.777	0.437
$\beta_{DifferenceF1*NRC}$	0.172		0.062	2.765	0.0057
$\beta_{LowR*NRC}$	0.351		0.284	1.232	0.218
$\beta_{HighR*NRC}$	-1.473		0.289	-5.097	<0.0001

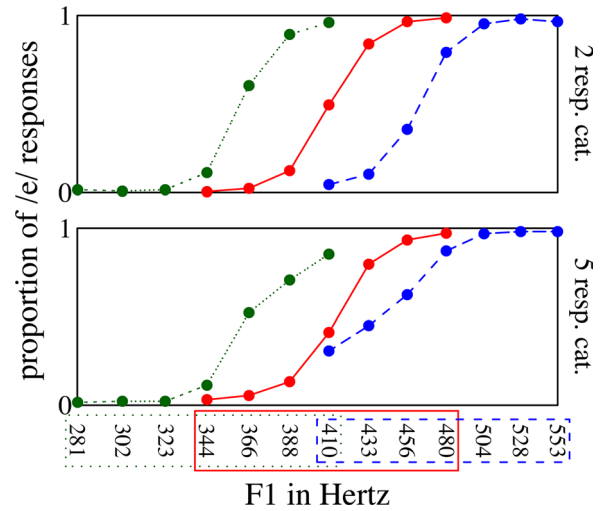


FIG. 2. (Color online) The proportion of /e/ responses (of the sum of /i/ and /e/ responses), for each of the  $F1$ -values in the low (dotted), intermediate (solid), and high (dashed) stimulus range by listeners with 2 (top) and 5 (bottom) response categories.

overwhelmingly categorized as /e/ in the low  $F1$  range and as /i/ in the high  $F1$  range.

The first aim was to examine whether listeners are sensitive to broad acoustic stimulus context in classifying isolated vowels. The results show that Peruvian Spanish listeners are sensitive to the  $F1$  range when categorizing vowels as /i/ and /e/ and shift their boundaries in two directions.

#### B. Effects of difference with the preceding stimulus (local acoustic context)

As can be seen from Table I, listeners are sensitive to the difference between the target and the preceding stimulus when categorizing vowels, because they are more likely to respond /e/ when the target stimulus has a higher  $F1$  than the stimulus preceding it, which is shown by a significantly positive  $\beta_{DifferenceF1}$ .

With respect to the second aim of the present study, we find that listeners are sensitive to the preceding local stimulus context. Specifically, the results show that listeners are more likely to respond /e/ to a target that has a higher  $F1$  than the preceding stimulus.

#### C. Long-distance "local" context effects

Although for the purposes of the general analyses presented before, we have separated broad and local context effects, it has been questioned whether a true distinction

TABLE II. The mean proportion of /i/ and /e/ responses (between-subjects standard deviation in italics between parentheses) in the low, intermediate, and high range condition to the stimuli with  $F1 = 410$  Hz, the only stimuli present in all three range conditions. Numbers do not add up to 1 because the proportion of all responses was taken and participants gave a low number of /a/, /o/, and /u/ responses to these stimuli as well.

	Low	Intermediate	High
/i/	0.09 (0.21)	0.54 (0.30)	0.82 (0.30)
/e/	0.90 (0.21)	0.45 (0.30)	0.17 (0.29)

between broad and local acoustic contrast effects can be maintained (e.g., Repp and Liberman, 1987) and the explanation of both effects in terms of auditory contrast (Kluender and Kiefte, 2006) suggests that one mechanism could underlie both effects. We analyzed to what extent the influence of the local acoustic context was affected by the number of intervening stimuli between the preceding and the target stimulus. It was expected that the closest stimuli would exert the strongest influence, but if broad context effects arise from local stimulus context effects, “local” stimulus context effects must be observable over a long distance.

To this end, six new variables were defined that referred to the difference between the  $F1$  of the target stimulus and the average over ten preceding stimuli, namely Difference\_1–10, Difference\_11–20, Difference\_21–30, Difference\_31–40, Difference\_41–50, and Difference\_51–60. For instance, Difference\_11–20 was the difference between the  $F1$  of the target stimulus and the average over the 11th through 20th preceding stimuli. In the analysis, the variable  $F1$  of the previous analysis and the six new difference variables were included and were all expected to yield positive  $\beta$ -coefficients. All coefficients as well as the intercept were allowed to vary across participants in a normal distribution. Because Difference\_51–60 requires 60 preceding stimuli, this variable can only be computed starting from the 61st trial. Recall that the three range conditions were presented to the participants in immediate succession and the stimulus preceding a target stimulus can come from a different range condition, potentially leading to a relatively large difference between the target and the preceding stimulus. Because we were interested in the local context effects within a stimulus range, only the first range condition per participant was analyzed. Therefore, the analysis only included the 61st through to the last (147th) trial of the first range condition per participant, for a total of 5513 analyzed responses.

The results showed a significant intercept (est =  $-0.6732$ , var =  $2.547$ , se =  $0.232$ ,  $z = -2.905$ ,  $p = 0.004$ ) and a significantly positive  $\beta_{F1}$  (est =  $1.501$ , var =  $0.987$ , se =  $0.279$ ,  $z = 5.375$ ,  $p < 0.0001$ ). Figure 3 displays the  $\beta$ -coefficients and the corresponding  $z$ - and  $p$ -values for the six difference variables. As can be seen, the difference from the ten stimuli preced-

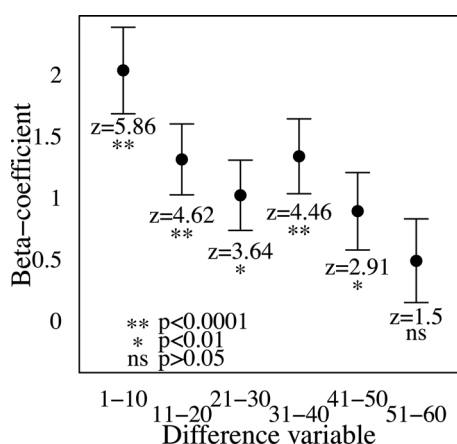


FIG. 3. The  $\beta$ -coefficients of the six difference variables. Error bars represent the standard errors of the coefficients. Below each coefficient, the  $z$ -values are given, as well as an indication of the magnitude of the  $p$ -value.

ing the target stimulus has the largest effect on listeners’ categorization, and the effect becomes smaller with more intervening stimuli between the target and the stimuli it is contrasted with. Yet, even after 40 intervening stimuli, the 41st through 50th preceding stimulus from the same range condition still have a significant effect on categorization.

The third aim of the present study was to explore whether the distinction between broad and local acoustic contrast effects can be maintained. Since the local context effects can be traced back over 40 intervening stimuli, these results support the idea that broad range effects result from a buildup of local context effects over a relatively large distance.

## D. The effect of the number of response categories

Figure 2 displays the proportion of /e/ responses given to the individual stimuli in the three  $F1$  ranges, for listeners with five and two categories separately. The results in Table I give no indication that the number of response categories has a direct effect on listeners’ categorization because  $\beta_{NRC}$  is not significant. Also, the lack of a significant interaction  $F1*NRC$  gives us no indication that the boundary steepness changes when listeners can choose from only two, rather than five, response categories. Importantly, the boundary shifts in the high stimulus range are larger when listeners can choose from only two response categories. This is shown by a significantly negative  $\beta_{HighR*NRC}$  in addition to the negative  $\beta_{HighR}$ . We did not observe that the boundary shifts in the low stimulus range are larger when listeners have only two response options, although the non-significant positive  $\beta$ -weight for the interaction  $LowR*NRC$  in addition to the positive  $\beta_{LowR}$  is in the expected direction. The analysis also shows a significantly positive  $\beta_{DifferenceF1*NRC}$ , which indicates that listeners are more sensitive to the difference between the target  $F1$  and the  $F1$  of the preceding stimulus if they have only two response categories.

The fourth aim of this study was to examine whether the number of response categories presented to listeners has an effect on their categorization behavior. The results show that if listeners can choose from only two instead of five response categories, they are more sensitive to both broad and local acoustic contexts.

## IV. GENERAL DISCUSSION

The present study was conducted to examine the contribution of three main influences on vowel categorization. The first influence was the acoustic properties of the target stimulus. The second influence was the acoustic context of the stimuli. This was instantiated by the  $F1$  range of the stimuli for the broad acoustic context and by the preceding stimuli for the local acoustic context. The third influence was the internal referents that listeners considered for the categorization task, for which two groups of listeners, who were presented with the same stimulus sets, had either two (/i/, /e/) or five (/i/, /e/, /a/, /o/, /u/) response options.

The results from this study indicate that listeners are sensitive to the broad acoustic context when categorizing vowel stimuli, as has been shown by others (e.g., Brady and Darwin, 1978; Keating *et al.*, 1981). Listeners shift their

category boundary between /i/ and /e/ with the range of the  $F1$  values of the stimuli. Such influences were also observed on a more local scale, as more /e/ responses are given when the preceding stimulus has a lower  $F1$  than the target stimulus.

Importantly, it has been questioned whether a true distinction between broad and local acoustic contrast effects can be maintained (e.g., Repp and Liberman, 1987) and the explanation of both effects in terms of auditory contrast (Kluender and Kiefte, 2006) suggests indeed that one mechanism underlies both effects. The results presented in this study show that local context could be traced back over 40 intervening stimuli, with stronger influences for more recent stimuli. This implies that the acoustic context has an effect that lasts longer than the 250 ms observed in electrophysiological studies (Liederman *et al.*, 2005), or the 13 stimuli observed in previous behavioral tasks (Holt, 2005). These findings question whether the true distinction between broad and local context effects can be maintained (cf. Repp and Liberman, 1987). Speech sound perception is continually updated with new information, while the older information still carries through. The distinction between broad and local context effects appears only a matter of experimental operationalization, rather than a true distinction in the listeners' processing.

We also investigated the interplay between perceptual contrast and the role of higher level influences. We found that contrast effects are stronger when listeners are presented with only two response categories than when they are presented with five. This was apparent when the influence of the preceding stimulus was considered and in the high stimulus range. The number of response categories available to listeners thus constrains their sensitivity to acoustic context effects. The finding that both local acoustic context and broad acoustic context were similarly influenced by the number of response categories also argues against a theoretical distinction between purely local versus global acoustic context effects.

The proposed buildup of contrast effects over multiple trials suggests that, especially at the start of the experiment, the responses of the participants in the different groups should be increasingly influenced by the range condition. Importantly, this divergence between the range conditions should differ between participants in the two, versus the five response category groups, as a smaller number of response categories makes listeners more susceptible to acoustic context effects. The top panels of Fig. 4 display the boundary locations in the first three blocks of the experiment for listeners with two response categories. It displays the development of the stimulus range effect from the first to the third block of 49 stimuli, and thus includes only each participant's *first* range condition.

A gradual buildup of contrast effects predicts that differences in the boundary location between the intermediate  $F1$  range and the low and high  $F1$  ranges become more pronounced in the second and third blocks than in the first. As can be seen from Fig. 4, already in the first range condition, listeners' boundaries between /i/ and /e/ are on lower  $F1$  values in the low stimulus range and on higher  $F1$  values in the high stimulus range, as compared to the intermediate

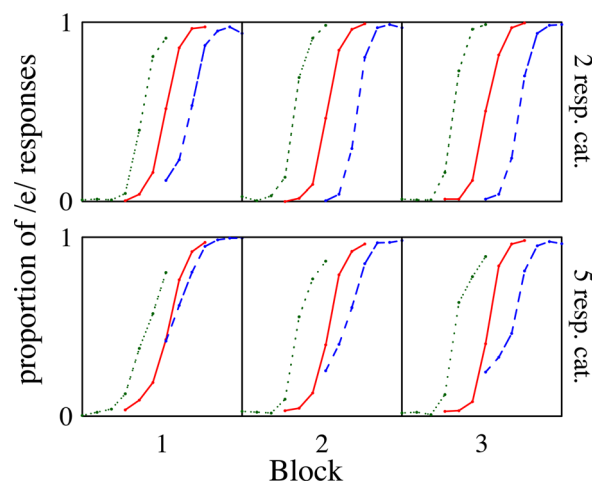


FIG. 4. (Color online) The proportion of /e/ responses (of the sum of /i/ and /e/ responses) for each of the  $F1$ -values in the low (dotted), intermediate (solid), and high (dashed) stimulus range by listeners with 2 (top) and 5 (bottom) response categories. From left-to-right, panels display data for the three consecutive blocks of the participants' first range condition.

stimulus range. A supplementary analysis, reported in the Appendix, confirmed this observation. Importantly, however, there is an increase in the extent of the boundary shifts over the course of the first range condition, showing that listeners adjust their boundaries more to the stimulus range as the range progresses. This is in line with the conclusion that broad context effects result from local contrast effects that build up over the course of an experiment. The longer a range condition lasts, the more acoustic context has been provided and the stronger the context effects will be.

The buildup of context effects across blocks allows us to straightforwardly address the effect of the number of response categories on context effects. The bottom panels of Fig. 4 display the boundary locations in the first three blocks of the experiment for listeners with five response categories. A comparison between the two groups reveals that the development over blocks in the high stimulus range depends on the number of response categories: listeners with two response categories shift their boundary early, while listeners with five categories require more time for their boundary shifts (again, this observation was confirmed by the additional analyses reported in the Appendix). This delay in the boundary shift in the high stimulus range, we argue, was caused by the availability of the response category /a/. The vowel /a/ has the highest  $F1$  value of all Spanish vowels and is the next vowel category if the /i/ to /e/ continuum employed in the present study were continued toward even higher  $F1$ -values. If the availability of the response category /a/ leads listeners to believe that they may hear this vowel at some point during the task, they may implicitly preserve a part of the vowel space for /a/. We indeed observe that the listeners who may be expecting to hear /a/ are reluctant to move their boundary between /i/ and /e/ toward the higher  $F1$  values. The boundaries between /i/ and /e/ can move more freely toward higher  $F1$  values, however, once listeners realize that /a/ is irrelevant because it is never the category they perceive. If listeners have never considered /a/ as a

TABLE III. Results from the analysis on the development of the range effect, as described in Sec. IV.  $\beta$  = coefficients of the factors. Coefficient estimates (est), between-subjects variance of the intercept and coefficients varying over participants (var), coefficient standard errors (se), and  $z$ - and  $p$ -values from the logistic regression analysis are given.

	est	var	se	$z$	$p$
Level 1: participant					
intercept	-0.659	5.512	0.498	-1.323	0.186
$\beta_{F1}$	7.905	8.799	0.421	18.760	<0.0001
$\beta_{Block}$	-0.172	0.316	0.148	-1.164	0.244
Level 2: Range and NRC					
$\beta_{LowR}$	7.275		0.756	9.617	<0.0001
$\beta_{HighR}$	-2.942		0.753	-3.909	<0.0001
$\beta_{NRC}$	0.072		0.498	0.144	0.886
$\beta_{LowR*Block}$	0.528		0.215	2.448	0.014
$\beta_{High*Block}$	-0.609		0.219	-2.778	0.005
$\beta_{LowR*NRC}$	0.585		0.752	0.779	0.436
$\beta_{HighR*NRC}$	-1.788		0.752	-2.377	0.017
$\beta_{Block*NRC}$	0.119		0.148	0.807	0.420
$\beta_{LowR*Block*NRC}$	0.161		0.215	0.749	0.454
$\beta_{HighR*Block*NRC}$	0.441		0.219	2.010	0.044

relevant category, because it was not a response option, they show no hesitation in moving the boundary between /i/ and /e/.

This additional analysis is in line with the results from the main analysis, in which we found that listeners with five response options shifted their boundary between /i/ and /e/ less than the listeners with two response categories in the high stimulus range, whereas no such difference was apparent in the low stimulus range. The perceptual contrast mechanism thus affects a listener's division of the vowel space, but the effect of this mechanism is reduced by the listeners' expectation to hear multiple relevant categories with similar acoustic properties.

For the current study we chose Peruvian Spanish, with only five vowel monophthongs, so as to maximize context effects (based on Keating *et al.*, 1981). Future research could investigate whether the strength of context effects is further reduced when participants have more response categories. Such research could, for instance, be performed with speakers of a language with more vowels, like English or Dutch.

What is the mechanism underlying the observed acoustic stimulus context effects? It has been argued that listeners are highly sensitive to acoustic contrast (Kluender and Kiefte, 2006), as well as to contrast in every sensory domain (cf. Kluender *et al.*, 2003). A general mechanism of perceptual contrast could indeed explain the observed effects. The long-distance local context effects suggest that both long- and short-term acoustic context contribute to the overall acoustic context in which a categorization decision is made.

In addition, however, the present study demonstrates that having more referents available in a sound categorization task reduces the strength of context effects in speech perception. This suggests that the effect of the general perceptual mechanism of contrast can be attenuated by higher-level influences.

The majority of perceptual context effects observed in the current study are probably due to the automatic and sensory-general mechanism of perceptual contrast. Yet, the availability of relevant perceptual categories was shown to

limit acoustic context effects. To better understand speech perception we thus need to understand the interrelation between general auditory mechanisms and higher level expectations. The current study is a step toward unraveling these intricate phenomena.

## ACKNOWLEDGMENTS

This research was funded by a travel grant awarded to P.E., and T.B.'s work was supported by Grant No. 021.002.095, both from the Netherlands Organization for Scientific Research (NWO). Many thanks to Professor Jorge Iván Perez Silva and Jorge Acurio for assistance in data collection. We would also like to thank Bart de Boer, Rob van Son, Paul Boersma, Sophie ter Schure, and an anonymous reviewer for their comments on earlier versions of this paper.

## APPENDIX

Here we report on the additional analysis that was mentioned in Sec. IV. The corresponding data are displayed in Fig. 4. In the analysis, the variables  $F1$ , NRC, LowR, and HighR from the main analysis were entered. To test the development of the broad acoustic context effects over time, an additional variable in the analysis was Block, coded as -1, 0, and 1 for the first, second, and third block, respectively. In addition to the main effects, the interactions LowR\*Block, HighR\*Block, Block\*NRC, and the three-way interactions were entered in the analysis. The intercept and the coefficients for  $F1$  and Block were allowed to vary between participants in the multilevel model. The coefficients for Range, NRC, and the interactions with these factors were fixed across participants. Only the first three blocks of 49 stimuli of each participants were entered in the analysis, that is, only the first range condition. The total number of analyzed responses was 9337. Table III shows the results for all main effects and interactions that were included in the model.



<sup>1</sup>Thanks to an anonymous reviewer for pointing us to this alternative explanation.

<sup>2</sup>The simplified synthesizer that was used has been created by H. Timothy Bunnell and can be found under [www.asel.udel.edu/speech/tutorials/synthesis/vowels.html](http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html).

<sup>3</sup>The experiment was run in an earlier Praat version.

- Boersma, P., and Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Retrieved from <http://www.praat.org/> (date last viewed 07/15/11).
- Brady, S. A., and Darwin, C. J. (1978). "Range effects in perception of voicing," *J. Acoust. Soc. Am.* **63**, 1556–1558.
- Broadbent, D. E., and Ladefoegd, P. (1960). "Vowel judgements and adaptation level," *Proc. R. Soc. London, Ser. B* **151**, 384–399.
- Broadbent, D. E., Ladefoegd, P., and Lawrence, W. (1956). "Vowel sounds and perceptual constancy," *Nature* **178**, 815–816.
- Chládková, K., Escudero, P., and Boersma, P. (2011). "Context-specific acoustic differences between Peruvian and Iberian Spanish vowels," *J. Acoust. Soc. Am.* **130**, 416–428.
- Cooper, W. E. (1974). "Adaptation of phonetic feature analyzers for place of articulation," *J. Acoust. Soc. Am.* **56**, 617–627.
- Eimas, P. D. (1963). "The relation between identification and discrimination along speech and non-speech continua," *Lang. Speech* **6**, 206–217.
- Escudero, P., Benders, T., and Lipski, S. C. (2009). "Differences in the perceptual cue weighting of Dutch vowels by Dutch, German and Spanish natives," *J. Phonetics* **37**, 452–465.
- Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). "The identification and discrimination of synthetic vowels," *Lang. Speech* **5**, 171–189.
- Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, London), pp. 1–625.
- Holt, L. L. (2005). "Temporally nonadjacent nonlinguistic sounds affect speech categorization," *Psychol. Sci.* **16**, 305–312.
- Keating, P. A., Mikos, M. J., and Ganong, W. F. (1981). "A cross-language study of range of voice onset time in the perception of initial stop voicing," *J. Acoust. Soc. Am.* **70**, 1261–1271.
- Kiefte, M., and Kluender, K. R. (2008). "Absorption of reliable spectral characteristics in auditory perception," *J. Acoust. Soc. Am.* **123**, 366–376.
- Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.
- Kluender, K. R., Coady, J. A., and Kiefte, M. (2003). "Sensitivity to change in perception of speech," *Speech Commun.* **41**, 59–69.
- Kluender, K. R., and Kiefte, M. J. (2006). "Speech perception within a biologically realistic information-theoretic framework," in *Handbook of Psycholinguistics*, edited by M. A. Gernsbacher and M. Traxler (Elsevier, London), pp. 153–199.
- Ladefoegd, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.
- Liederman, J., Frye, R., Fisher, J. M., Greenwood, K., and Alexander, R. (2005). "A temporally dynamic context effect that disrupts voice onset time discrimination of rapidly successive stimuli," *Psychon. Bull. Rev.* **12**, 380–386.
- Lotto, A. J., Kluender, K. R., and Holt, L. L. (1998). "Depolarizing the perceptual magnet effect," *J. Acoust. Soc. Am.* **103**, 3648–3655.
- Morrison, G. S. (2007). "Logistic regression modelling for first- and second-language perception data," in *Segmental and Prosodic Issues in Romance Phonology*, edited by M. J. Solé, P. Prieto, and J. Mascaró (John Benjamins, Amsterdam), pp. 119–236.
- Morrison, G. S., and Kondaurova, M. V. (2009). "Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis," *J. Acoust. Soc. Am.* **126**, 2159–2162.
- Morse, P. A., Kass, J. E., and Turkienicz, R. (1976). "Selective adaptation of vowels," *Percept. Psychophys.* **19**, 137–143.
- Parducci, A. (1965). "Category judgment: a range-frequency model," *Psychol. Rev.* **72**, 407–418.
- R Development Core Team. (2008). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, <http://www.R-project.org> (date last viewed 04/05/11).
- Repp, B. H., and Liberman, A. M. (1987). "Phonetic categories are flexible," in *Categorical Perception*, edited by S. Harnad (Cambridge University Press, Cambridge), pp. 89–112.
- Restle, F. (1987). "Assimilation predicted by adaptation-level theory with variable weights," in *Cognitive Theory*, edited by N. J. Castellan and F. Restle (Erlbaum, Hillsdale), pp. 75–91.
- Sawusch, J. R., and Nusbaum, H. C. (1979). "Contextual effects in vowel perception 1 anchor- induced contrast effects," *Percept. Psychophys.* **25**, 292–302.
- Sjerps, M. J., Mitterer, H., and McQueen, J. M. (2011). "Constraints on the processes responsible for the extrinsic normalization of vowels," *Attention, Percept., Psychophys.* **73**, 1195–1215.
- Stilp, C. E., Alexander, J. M., Kiefte, M., and Kluender, K. R. (2010). "Auditory color constancy: calibration to reliable spectral properties across nonspeech contexts and targets," *Attention, Percept., Psychophys.* **72**, 470–480.
- Thompson, C. L., and Hollien, H. (1970). "Some contextual effects on perception of synthetic vowels," *Lang. Speech* **13**, 1–13.
- Watkins, A. J. (1991). "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion," *J. Acoust. Soc. Am.* **90**, 2942–2955.
- Watkins, A. J., and Makin, S. J. (1994). "Perceptual compensation for speaker differences and for spectral-envelope distortion," *J. Acoust. Soc. Am.* **96**, 1263–1282.
- Watkins, A. J., and Makin, S. J. (1996). "Effects of spectral contrast on perceptual compensation for spectral-envelope distortion," *J. Acoust. Soc. Am.* **99**, 3749–3757.