The interrelation between the stimulus range and the number of response categories in vowel categorization

Benders, T.; Escudero, P.

Publication date
2010

Document Version
Author accepted manuscript

Published in
Interspeech

The interrelation between the stimulus range and the number of response categories in vowel categorization

Titia Benders, Paola Escudero

Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands
titia.benders@uva.nl, escudero@uva.nl

Abstract

We investigate the influence of the stimulus range and the number of response categories on the location of perceptual boundaries. The F1 continuum between Spanish /i/ and /e/ was presented to Peruvian listeners in three ranges. Half of the listeners could classify the tokens as /i/ and /e/, the other half chose from the five Spanish vowels. A boundary shift between /i/ and /e/ was observed as a function of the stimulus range, which was larger when listeners were given only two response categories. These results are interpreted as an effect of listeners' category expectations on speech perception.

Index Terms: perceptual category boundaries, stimulus range, response categories

1. Introduction

One aspect of speech sound categorization is the location of the perceptual boundary between two categories along an acoustic continuum. In speech perception studies, this boundary location is computed as the value along the acoustic continuum where the probability that a listener perceives either category is 50%. The present study investigates whether the values of all presented stimuli, the number of response categories available to listeners and the interrelation between those factors influence the location of perceptual boundaries.

Regarding the influence of the stimulus values on perceptual boundaries, it has been shown that vowel categorization depends on the formant values of the vowels in the preceding carrier sentence [1, 2]. Specifically, the same target stimulus was classified as /bɛt/ when preceded by a carrier phrase with a relatively low average F1 but as /bɪt/ when preceded by the same carrier phrase with a relatively high average F1 [1]. Perceptual boundaries can also be affected by the properties of a single sound. When a target sound appears after multiple repetitions [3] or after a single presentation [4] of a preceding sound, listeners tend to classify the target sound into a different category than that of the preceding sound. Thus, a stimulus is not only classified according to its own acoustic properties but also relative to the properties of the preceding sound context, which leads to changes in perceptual category boundaries.

Researchers have also examined the influence of the entire set of stimuli presented to listeners on the location of perceptual boundaries. In these studies, stimuli are divided in ranges which span a part of the whole acoustic continuum between two sound categories. Stimuli from these ranges are presented in separate blocks. Results show, for instance, that American-English listeners shift their perceptual boundary between /d/ and /t/ towards a shorter VOT when they are presented with sounds from the shorter VOT range. Similarly, when presented with the longer VOT range, they shift their perceptual boundary towards a longer VOT [5]. The effect of the VOT range has been replicated for Polish listeners but not for American-English listeners [6]. Such boundary shifts can be stronger for non-native than native listeners as shown by [7] for the F3 continuum between English /r/ and /l/.

The influence of the sound context, i.e., preceding sentence, preceding isolated sound, and the range of the entire set of stimuli, on the location of perceptual category boundaries can be interpreted as a consequence of *perceptual contrast* [8]. Researchers that demonstrate the effect of sound context on speech sound categorization generally state that observers tend to categorize sensory stimuli relative to the stimulus or stimuli with which they are paired [1, 9, 10, 11]. The present study first aims at showing that listeners' perceptual boundary location between two vowels shifts depending on the specific range of F1 values of the stimuli.

The second effect on perceptual boundaries considered in the present study is the number of response categories available to listeners. Upon hearing an initial consonant produced with acoustic values intermediate between /t/ and /d/ combined with the rhyme "ask", listeners classified the initial consonant as /t/ [12]. The fact that "task" but not "dask" is a word in English influences listeners' sound categorization. A more direct demonstration of the effect of lexical knowledge on boundary locations in speech perception is provided by [13]. Listeners were presented with a fricative consonant in between /s/ and /f/ in lexical contexts that were congruent with only one of these fricatives. The listeners subsequently categorized the ambiguous fricative sound as a possible realization of the fricative contained in the existing words. It can thus be concluded that listeners' lexical expectations influence their perceptual boundary locations. Which sound categories listeners expect to hear in a perception study may thus have an effect on perceptual boundary locations as well, which is explored in the current study.

The third aim of the present study is to investigate the interrelation between the effects of the sound context and the categories listeners expect to hear on perceptual boundary locations.

To these ends, we designed a perception experiment that manipulated both the stimulus values and the number of possible response categories available to listeners. The F1 continuum between Peruvian Spanish[1] /i/ and /e/ was divided in three ranges and presented to listeners in separate blocks. We expected the different stimulus ranges to lead to different boundary locations between the vowels /i/ and /e/. Half of the

[1] The Peruvian Spanish vowel system is similar to the Iberian Spanish vowel system in that it contains the vowels /i/, /e/, /a/, /o/ and /u/. As compared to their Iberian counterparts, the Peruvian /a/ has a lower F1, the Peruvian /e/ has a higher F2, and the Peruvian /o/ has a lower F2 [cf. 14]. The F1 values of /i/ and /e/ are comparable between the variants.

monolingual Peruvian Spanish listeners who participated in the study could classify the tokens as the vowels /i/ and /e/, while the other half could classify them as all five Spanish vowels. We predicted that the number of response categories would influence the extent to which listeners' perceptual boundary locations shift from one stimulus range to the other.

A previous study [15] used a similar experimental design to demonstrate the effect of the number of response categories on perceptual boundary shifts. In one condition, only stimuli from the continuum from /i/ to /ɪ/ were presented, while in a second condition, the same stimuli were presented interspersed with tokens of /ɛ/. Half of the listeners could categorize the tokens as /i/, /ɪ/ and /ɛ/, while the other half only had /i/ and /ɪ/ as response categories. When /ɛ/ tokens were also presented, both groups had a boundary shift between /i/ and /ɪ/ towards higher F1 values, which is in line with the perceptual contrast effect. Interestingly, the boundary shift between /i/ and /ɪ/ was larger in the two-category than in the three-category group.

In [15], listeners with two response categories had fewer response options than were actually present in the stimuli, which may have forced the listeners to adopt a response strategy. In line with this argument, the authors interpret their results as support for the *range frequency theory* which states that boundary shifts result from participants' strategy to evenly distribute their responses over the available categories [16].

We hypothesize that the number of response categories available to listeners can change their perception of stimuli and not only how they respond to them. Thus, in the current study, we aim at demonstrating the interrelation between the stimulus range and the number of response categories on boundary shifts by presenting different stimulus ranges that fall within the natural distribution of the response categories available to all listeners.

2. Methodology

2.1. 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 to 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.

2.2. Stimuli

The stimuli were isolated synthetic vowels, which were created using a simplified version of the Klatt synthesizer [17].

The 13 F1 values of the stimuli ranged from 281 Hz to 553 Hz and the steps between stimuli were approximately equal on the Erb scale (0.31 Erb). The F2 values ranged from 1893 Hz to 2557 Hz and were inversely proportional to F1 in Hz. The 7 durational values of the stimuli ranged from 80 ms to 175 ms and the steps were approximately equal when measured in natural logarithms (0.12-0.15). Combining the 13 spectral values with the 7 durational values leads to 91 unique stimuli. The values of the 91 stimuli are displayed in Figure 1.

All stimuli had a fundamental frequency that decreased linearly from 180 Hz to 140 Hz. The formant frequencies were steady throughout the vowel.

The stimuli were divided in three F1-ranges, as shown in Figure 1: A phonetically *low* range (the 49 stimuli with the F1 values from 281 Hz to 410 Hz), a phonetically *intermediate* range (the 49 stimuli with F1 values from 344 Hz to 480 Hz), and a phonetically *high* range (the 49 stimuli with the F1 values from 410 Hz to 553 Hz).

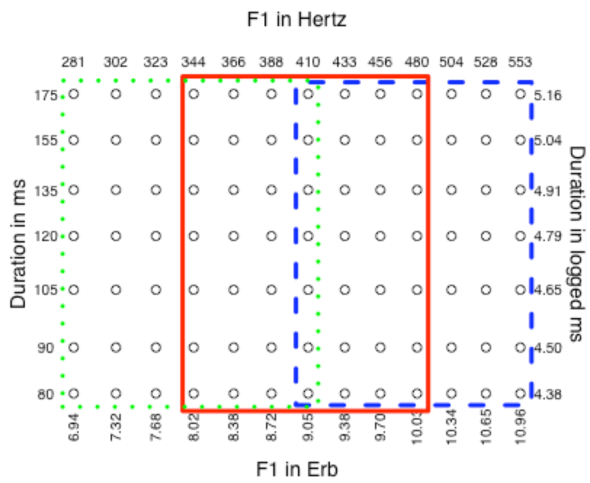

Figure 1: *F1 and duration values divided in the three stimulus ranges: low (left, dotted green box), intermediate (middle, solid red box), and high (right, dashed blue box).*

All F1 values can be considered possible productions of Peruvian Spanish /i/ or /e/, but the intermediate range contains F1 values that best match their average productions [14].

2.3. Procedure

The perception experiment was run on a PC laptop computer using the Praat program [18]. 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 response categories, i.e., "i" and "e", while the other half could choose from the five Spanish vowels, i.e., "i, e, a, o, 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, the order of which was counterbalanced across participants. For each range, a randomized block of the 49 stimuli was presented three times, leading to 147 trials per range and a total of 441 trials. Prior to the first block, participants received a practice session of 10 tokens and they could take a short break after each block.

2.4. Analysis

The data were analysed using hierarchical logistic regression analysis. Logistic regression models the influence of independent variables on a binary dependent variable. Hierarchical modeling is a compromise between pooling the data of all participants in one analysis and applying a separate model on each participant's data [19]. The coefficients for some variables in the analysis can vary per participant but are constrained by the fact that they must form a group distribution. We fit the data with the model in (1):

$$\mathrm{P}(/e/)_i=\mathrm{logit}^{-1}(\alpha + \beta_{F1\_p[i]}*F1_i + \beta_{NRC}*NRC + \beta_{LowR\_p[i]}*LowR_{\_i} + \beta_{HighR\_p[i]}*HighR_{\_i} + \beta_{LowR*NRC}*(LowR*NRC) + \beta_{HighR*NRC}*(HighR*NRC))$$

$$\beta_{F1\_p[i]} \sim \mathrm{N}(\mu_{\mathrm{F1\_p}}, \sigma^2_{\mathrm{F1\_p}}),\ \beta_{LowR\_p[i]} \sim \mathrm{N}(\mu_{\mathrm{LowR\_p}}, \sigma^2_{\mathrm{LowR\_p}}),\ \beta_{HighR\_p[i]} \sim \mathrm{N}(\mu_{\mathrm{HighR\_p}}, \sigma^2_{\mathrm{HighR\_p}}) \quad (1)$$

The responses /i/ and /e/ were coded as 0 and 1, respectively. Thus, positive β-coefficients indicate that an increase in the value of the variable increases the likelihood that a listener has responded /e/ rather than /i/. The variable *F1* refers to the vowel's F1 value expressed in Erb. Because /e/ has higher F1 values than /i/, β_{F1} was expected to be positive.

LowR is a dummy variable representing the contrast between the intermediate range condition (coded as 0) and the low range condition (coded as 1). *HighR* is a dummy variable that represents the contrast between the intermediate range (coded as 0) and the high range (coded as 1). These two variables address the first aim of the present study, as they test whether listeners shift their boundaries when the stimulus range is shifted as compared to the intermediate range. We compare the low and the high ranges to the intermediate range because, as mentioned above, the values of the latter are most comparable to Peruvian Spanish productions. A positive β_{LowR} would indicate that listeners are more likely to give an /e/ response in the low range than in the intermediate range condition and thus that their perceptual boundary between /i/ and /e/ in the low range condition is shifted towards lower F1 values than in the intermediate range. A negative β_{HighR} would indicate that listeners' boundary between /i/ and /e/ is shifted towards higher F1 values in the high range.

NRC is a dummy variable representing the number of response categories with two levels, i.e., five (coded as 0) and two categories (coded as 1). This variable addresses the second aim of the study, as it tests whether listeners' boundary location between /i/ and /e/ is dependent on the number of response categories. More specifically, a positive β_{NRC} would indicate that listeners with two response categories are more likely to give an /e/ response than listeners with five response categories and thus that their boundary location between /i/ and /e/ is on a lower F1 value. A negative β_{NRC}, on the other hand, would indicate that the perceptual boundary of listeners with two response categories is on a higher F1 value.

The interaction terms *LowR*NRC* and *HighR*NRC* are important for the third aim of the present study: They represent the effect of the number of response categories on the boundary shift from the intermediate to the low and high stimulus ranges. Because the two-category group is coded as 1 in the variable *NRC*, a positive $\beta_{LowR*NRC}$ in addition to a positive β_{LowR} would indicate that the boundary shifts towards even lower F1 values in the low stimulus range when listeners are presented with two response categories than when they are presented with five. Similarly, a negative $\beta_{HighR*NRC}$ in addition to a negative β_{HighR} would indicate that the boundary shifts towards even higher F1 values in the high stimulus range for listeners with two response categories.

In the analysis, the intercept, α, is identical for all participants because they are drawn from the same population. The β-coefficients for *F1* and for the dummy variables *NRC*, *LowR*, and *HighR* vary between participants in a standard normal distribution. The interaction terms do not vary between participants. The index of the participant is *i* and $\beta_{F1\ p[i]}$ is the β-coefficient for *F1* of the *i*th participant in the sample. This participants' β-coefficients form a normal distribution with mean $\mu_{F1\ p}$ and variance $\sigma^2_{F1\ p}$, which are estimated from the data as well. In the results, these means and variances will be reported for the coefficients that varied across participants.

The responses that were not /i/ and /e/ were excluded from the analysis because a binary logistic regression analysis can only take two response categories and because the number of /u/, /o/ and /a/ responses was too low to warrant a multinomial analysis. In total, 650 responses were removed. Of the removed responses, 36 were /u/, 110 were /o/, and 504 were /a/. After excluding the responses to other vowels than /i/ and /e/, a total of 27574 data points were included in the analysis.

An alpha level of 0.05 was adopted for all tests. All statistical analyses were conducted with the lme4 package in the open-source statistical software R [20].

3. Results

Table I shows the results of the analysis, and Figure 2 displays the modeled average regression curves in the three ranges, for the listeners with five (Fig. 2a) and two categories (Fig. 2b). As expected, β_{F1} is significantly positive, which confirms that listeners choose /e/ when a stimulus has a higher F1.

The first aim of this study was to show that the boundary location between /i/ and /e/ shifts with the stimulus range. The positive and significant β_{LowR} shows that the boundary location between /i/ and /e/ shifts towards lower F1 values in the low stimulus range as compared to the intermediate stimulus range. The negative and significant β_{HighR} shows that the boundary location shifts towards higher F1 values in the high stimulus range as compared to the intermediate stimulus range. We thus observe a boundary shift in two directions.

The second aim was to examine whether the number of response categories available to listeners affect their boundary location between /i/ and /e/. The negative and significant β_{NRC} indicates that the boundary location between /i/ and /e/ is on a higher F1 value for listeners with two response categories than for listeners with five response categories. This result suggests that in vowel perception listeners are affected by the number of available responses.

The third question in this study was to explore the interrelation between the effect of the stimulus range and the number of response categories on boundary shift. It can be observed in Figure 2 that the boundary shifts for the two-category group (panel b) are larger than those for the five-category group (panel a) in both the low and the high ranges. This is confirmed by a significantly negative $\beta_{HighR*NRC}$. $\beta_{LowR*NRC}$ is positive, but not significant.

	est	sd	se	*z*	*p*
α	-45.88		0.820	55.98	<0.01
β_{F1}	5.62	1.303	0.179	31.51	<0.01
β_{NRC}	-5.16	11.731	1.610	3.20	<0.01
β_{LowR}	3.35	1.786	0.317	10.57	<0.01
β_{HighR}	-1.71	2.011	0.356	4.81	<0.01
$\beta_{LowR * NRC}$	0.82		0.425	1.93	0.05
$\beta_{HighR * NRC}$	-2.41		0.477	5.04	0.01

Table 1. *β-coefficient estimates (est), standard deviations of the β- coefficients that were drawn from a group distribution (sd), β-coefficient standard errors (se), and z-, and p-values from the logistic regression analysis.*

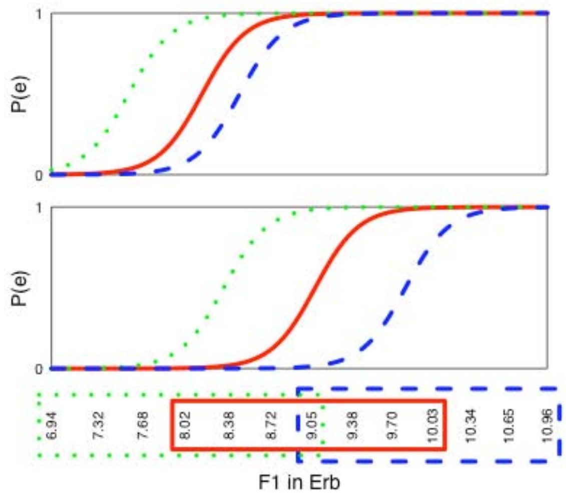

Figure 2: *Boundary locations for listeners with five (panel a) and two (panel b) response categories in the low (left, dotted green line), intermediate (middle, solid red line) and high (left, dashed blue line) stimulus ranges.*

4. Discussion

The results of the present study show, as was expected based on previous findings, that listeners shift their perceptual boundary between /i/ and /e/ as a function of the stimulus range. It is also shown that boundary locations are dependent on the number of response categories available to listeners and that the boundary shifts due to stimulus range shifts are larger when listeners can choose from only two response categories.

As mentioned in the introduction, listeners' expectations guide their speech perception. In the present study, listeners who saw "i" and "e" on a screen expected to hear only those two vowels, while listeners who saw "i, e, a, o, u" may have expected to hear five vowels. This can explain why the boundary between /i/ and /e/ was on a lower F1 value for listeners with five response options than for listeners with two response options. The expectation of hearing /a/ may have resulted in the boundary between /i/ and /e/ being located at low F1 values, thereby saving some perceptual space for /a/.

We hypothesize as well that the listeners with five response categories compared each stimulus to their stored representations of the five vowels when selecting their response. With five such anchors, listeners may be less sensitive to perceptual contrast effects and thus display smaller boundary shifts than listeners who compare the same stimuli to only two stored representations.

The present results show that when listeners are not required to respond with unlikely categories, as in [15], a reliable effect of the number of response categories on the perceptual boundary shifts can still be observed. The present results also demonstrate that the interrelation between the number of response categories and perceptual boundary shifts cannot be ascribed to listeners' strategy to divide their response equally over all available categories, as was suggested by [15]. This is because listeners with five response categories used the less likely categories, i.e., /a, o, u/, to a negligible extent.

According to our interpretation of the results, the observed boundary shifts and the interaction with the number of response categories are a true speech perception effect. However, further research is still needed to differentiate between a purely perceptual account and a response strategy.

Finally, our findings have a clear implication for speech perception studies that investigate the location of perceptual boundaries. That is, it seems more accurate to present listeners with more than two response categories and, preferably, with all possible categories that are relevant for a specific stimulus set. This methodology may lead to a more reliable estimation of perceptual boundary locations.

5. Acknowledgements

This research was funded by a grant no. 021.002.095 to TB, and grant no. 275.75.005 to PE, both from the Netherlands Organization for Scientific Research (NWO). Many thanks to Professor Jorge Iván Perez Silva and Mr. Jorge Acurio for assistance in data collection.

6. References

[1] Ladefoged, P., and Broadbent, D.E., "Information conveyed by vowels", J. Acoustic. Soc. Am., 27:98-104, 1957.
[2] Mitterer, H., "Is vowel normalization independent of lexical processing?", Phonetica, 63:209-229, 2006.
[3] Eimas, P.D., and Corbit, J.D., "Selective adaptation of linguistic feature detectors", Perception & Psychophysics, 4:99-109, 1973.
[4] Fry, D., Abramson, A., Eimas, P.D., & Liberman, A.M., "The identification and discrimination of synthetic vowels", Lang. Speech, 5:171-189, 1962.
[5] Brady, S.A., and Darwin, C.J. "Range effect in the perception of voicing", J. Acoustic. Soc. Am., 63(5):1556-1558, 1978.
[6] Keating, P.A., Mikos, M.J., and Ganong, W.F., "A cross-language study of range of voice onset time in the perception of initial stop voicing", J. Acoust. Soc. Am., 70(5):1261-1271, 1981.
[7] Yamada, R.A. and Tohkura, Y., "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners", Perception & Psychophysics, 52(4):376-392, 1992.
[8] Helson, H., "Adaptation Level Theory", Harper and Row, 1964.
[9] Simon, H.J., and Studdert-Kennedy, M. "Selective adaptation and anchoring of phonetic and nonphonetic continua", J. Acoust. Soc. Am., 43:1338-1357, 1978.
[10] Repp, B.H., and Liberman, A.M., "Phonetic category boundaries are flexible", In: S. Harnad [Ed.] Categorical perception: The groundwork of cognition, 89-112, Cambridge University Press, 1987.
[11] Holt, L.L., "The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization", J. Acoust. Soc. Am., 120(5):2801-2817, 2006.
[12] Ganong, W.F., "Phonetic categorization in auditory word perception", J. Exp. Psychol. hum. Perception Performance, 6:110-125, 1980.
[13] Norris, D., Cutler, A., and McQueen, J.M., "Perceptual learning in speech", Cognitive Psychol., 47:204-238, 2003.
[14] Chládková, K., Escudero, P., and Boersma, P. "An acoustic analysis of Peruvian and Iberian Spanish vowels: investigating the effect of phonetic context on dialectal differences in vowels", Manuscript in preparation, University of Amsterdam.
[15] Sawusch, J.R., and Nusbaum, H.C., "Contextual effects in vowel perception I: Anchor-induced contrast effects", Percept. Psychophys. 34:560-568, 1983.
[16] Parducci, A., "Category judgment: a range-frequency model", Psychological Review, 72(6):407-418, 1965.
[17] http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html
[18] Boersma, P., and Weenink, D. "Praat: doing phonetics by computer", [computer program], retrieved from www.praat.org.
[19] Gelman, A., and Hill, J., "Data analysis using regression and multilevel hierarchical models", Cambridge University Press, 2007.
[20] R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.