

PICKING UP THE CUES TO A NEW CONSONANT CONTRAST: DANISH LEARNERS' PRODUCTION AND PERCEPTION OF ENGLISH WORD-FINAL /s/ - /z/

Nikola Anna Eger^a, Ocke-Schwen Bohn^b

^aInstitute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität, Munich, Germany;

^bEnglish Department and ConAmore, Department of Psychology, Aarhus University, Denmark
eger@phonetik.uni-muenchen.de; ocke.bohn@dac.au.dk

ABSTRACT

The present study examined how non-native speakers produce and perceive a voicing contrast that does not exist in their native language, and which is importantly signaled by an acoustic cue which has a very different function in the native language. Danish has no fricative voicing contrasts, which in English occur in initial, medial, and final position. An important acoustic cue for the final voicing contrasts in English is the duration of the preceding vowel. In Danish, vowel duration is used to contrast vowels. In Experiment 1, Danish learners of English produced only a small and nonsignificant difference in the vowel:fricative ratio, but Experiment 2 revealed that the same learners differentiated the final voicing contrast perceptually, influenced by vowel duration much like the native speakers. The overall results imply that although L2 learners do not produce certain unfamiliar contrasts native-like, they can use L2 cues to differentiate contrasts in perception.

Keywords: L2-acquisition, cross-lingual perception, unfamiliar contrast, voicing contrast, Danish

1. INTRODUCTION

What happens when second language (L2) learners encounter a non-native contrast that does not exist in their native language (L1)? Previous studies have shown that L2 learners may be able to differentiate a new contrast of the L2 in production and in perception, but they may do this using acoustic cues that are only secondary or nonfunctional for native speakers of the L2. For example, L1 Japanese listeners use F2 frequency to perceptually differentiate English /r/-/l/, even though F3, not F2, is the major acoustic cue for this contrast [11]. Similarly, listeners with L1

Spanish have been reported to differentiate English /i/-/ɪ/ using duration, whereas L1 English listeners use spectral differences and all but ignore duration [2].

Many earlier studies examined the production and/or perception of non-native contrasts that are primarily signaled by acoustic cues that are not employed in the L1 of the L2 learners, such as vowel duration in Spanish. More recent studies have also examined whether learners can transfer their L1-based acquaintance with a contrast in one position (e.g., initial voicing contrast) to an unfamiliar non-native position (e.g., final voicing contrast, see [4]).

The present study contributes to the existing literature by examining whether an acoustic cue that serves a specific function in the L1 can be re-assigned to serve a different function in the L2. Specifically, we explored L1 Danish (DK) learners' production and perception of the English (EN) final fricative voicing contrast /s/-/z/. DK has two voiceless fricatives /f, s/, lacks voiced fricatives, and vowel duration is exclusively used to signal phonemic vowel contrasts [6]. The most important segmental function of vowel duration in English is to signal whether the following tautosyllabic consonant is phonologically voiced (long vowel) or voiceless (short vowel), see [9].

Experiment 1 compared the production of English word-final /s/-/z/ by L1 speakers of DK and EN to establish whether DK speakers can re-assign vowel duration from signaling vowel contrasts in the L1 to signaling voicing contrasts in the L2. Experiment 2 examined the perception of this word-final contrast and the influence of the duration of the preceding vowel by presenting two nine-step continua from [s] to [z], differing in vowel duration.

2. EXPERIMENT 1: PRODUCTION

2.1. Method

2.1.1. Participants

15 L1 DK speakers (4 m, 11 f, $M_{age} = 23.2$ years) and 11 L1 EN speakers (8 m, 3 f, $M_{age} = 26.7$ years) participated as unpaid volunteers. The DK participants, who were current or former students at Aarhus University, had learned English as a foreign language for a mean of 9.7 years and had contact with English on a daily basis. The L1 English speakers, who all came from the US, had varying levels of familiarity with DK and with German, neither of which has final voicing contrasts.

2.1.2. Design and Procedure

The words for the reading task were selected to examine the production of word-final /s/ and /z/. Participants produced randomized lists with six repetitions of words including the segments of interest (*peas*, *piece*, *wise*, *rice*, *buzz*, *bus*, *shoes*, *goose*), as well as ten distractor words.

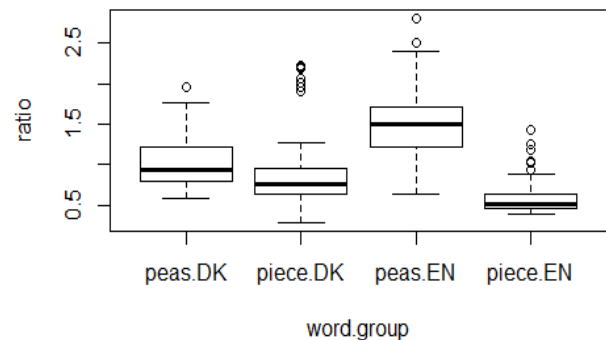
Participants were instructed to read the words one by one in a clear citation style with the numbers before each word. Speakers were recorded in a sound-treated environment using high fidelity digital equipment.

2.2. Results and Discussion

Acoustic analyses were conducted using *praat* [1]. First, final fricative productions were labeled as voiced, voiceless or devoiced (initially voiced and voiceless in the end) based on visual inspection of periodicity in the waveform and the spectrogram. Second, the durations of the fricative *fdur* and the preceding vowel *vdur* were measured. To normalize for speaking rate differences, and to account for the fact that fricative duration differs as a function of voicing [7], we based further analyses on the ratio of *vdur* to *fdur*. The present study presents the results for only one of the minimal pairs, *peas* - *piece*. The EN speakers' productions of the fricative in *peas* were devoiced or voiceless. Two DK speakers produced devoiced fricatives for the word *peas*, the remaining 13 DK speakers produced the fricative completely voiceless.

The ratios of *vdur* to *fdur* revealed that the EN speakers did not neutralize the voicing contrast, and even the DK speakers produced ratios that were, overall, higher for final /z/ than for /s/. Figure 1 shows that the EN speakers produced a clear contrast in terms of the *vdur* to *fdur* ratio for /z/ vs /s/, whereas the DK speakers produce only a small difference in the ratios for final /z/ vs /s/. An ANOVA on the ratios with language group (EN, DK) as between-subjects factor and final voicing (/z/, /s/) as within-subjects factor revealed a significant effect of final voicing ($F[1,24] = 49.8$, $p < 0.001$) but not of language group. The significant language group x final voicing interaction ($F[1,24] = 28.6$, $p < 0.001$) was explored through Bonferroni-adjusted post-hoc tests which revealed a significant difference between the *vdur* to *fdur* ratios in the productions of the EN speakers ($p < 0.001$) but not of the DK speakers. Note, however, that the nonsignificant difference produced by the DK speakers is in the expected direction, with slightly higher ratios for final /z/ than /s/. (One DK speaker produced ratios with opposite values, probably caused by confusion of the orthographic forms. This contributed to the nonsignificant results.)

Figure 1: Boxplots of the duration ratios for the two words *peas* with voiced /z/ and *piece* with the voiceless fricative /s/, paired with the groups Danish DK and native English EN speakers.



Further acoustic analyses are necessary to determine the amount of periodicity and noise in the fricatives to enable a comparison and discuss neutralization within the consonant itself.

To conclude, both the native (EN) and the non-native (DK) speakers produced the phonologically "voiced" final /z/ as a devoiced or voiceless fricative. The EN speakers signaled the final /s/-/z/ contrast through significantly different *vdur* to *fdur* ratios. The nonsignificant trend for the DK speakers in the correct direction (higher ratio for

final /z/ than /s/) raises the question of whether DK speakers display some sensitivity to the duration ratio cue in perception. The results of the study [4] lead us to expect that, if DK speakers show some sensitivity to this cue, they might not be able to use it in a native-like manner.

We addressed these issues in a perception experiment in which the same speakers as in Experiment 1 participated.

3. EXPERIMENT 2: PERCEPTION

Previous research has indicated that the differences between native speakers and L2 learners might not only be found in the accuracy of their perception, but also in the manner in which they perceive the non-familiar contrast (or familiar contrast in unfamiliar position, see [4]). To test this hypothesis, we conducted a perception experiment in which vowel duration was manipulated. It can be assumed that the EN listeners in this experiment will also rely on vowel duration, which they used in their productions to differentiate the final fricative voicing contrast. Based on the results of the production experiment we hypothesize that:

1. EN listeners perceive more /z/ when the vowel is phonetically long and more /s/ when the vowel is phonetically short.
2. EN listeners perceive the word-final /s/-/z/ contrast more categorically than DK listeners.
3. DK listeners rely even more on vowel duration than the native speakers.

H3 was motivated by the fact that the DK phoneme inventory does not have any fricative voicing contrast. Thus, DK listeners might rely more on cues other than fricative voicing, e.g., on vowel duration which exists in DK but is used exclusively to differentiate vowel contrasts.

3.1. Method

3.1.1. Participants

The same speakers as in Experiment 1 participated immediately after the reading task in the perception experiment.

3.1.2. Materials

To explore whether the two language groups (EN and DK) differ in the use of the vowel duration cue, two nine-step continua were created. The two nonce words [vi:s] and [vi:z] were recorded from a

native EN speaker. These tokens were morphed using the STRAIGHT algorithms [8] to two nine-step continua in which the fricative alters stepwise from a completely voiceless [s] to a voiced [z]. The nine steps of the fricative were shortened to the same duration (201ms). The continua differed in that the vowel (preceding the fricative) in the “long” continuum was phonetically long (292 ms), whereas the vowel in the “short” continuum was phonetically short (179 ms).

Since the duration of both the vowel (for *short* and *long*, respectively) and the nine steps of the fricatives did not vary within one continuum, the duration ratio was non-informative, such as it was static and on one end of the continua even revealed misleading information, i.e. a phonetically long vowel in combination with a voiceless fricative, and a phonetically short vowel with the voiced counterpart. Thus, the participants were presented with tokens with either a phonetically short or long vowel combined with one fricative of the nine steps from the [s-z] continuum.

The initial consonant [v] was replaced by a cluster [gw], which does not exist in Danish and only in loanwords and names occurring in English. This ensured that the stimuli do not occur as real words in English or in Danish.

3.1.3. Design and Procedure

Ten repetitions of the eighteen stimuli (9 steps x 2 vowel durations) were presented in a two-AFC perception experiment. The response options were “z” and “s”, and the order of presentation (long vowel or short vowel) was counter-balanced. Participants were offered a short break after each block of 36 presentations. The experiment was self-paced and the participants heard the next stimulus one second after entering their response.

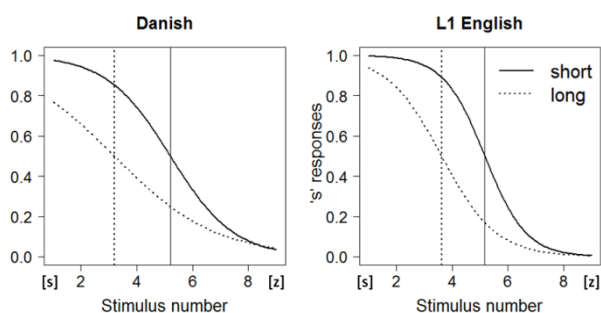
The participants were instructed that they would hear several made-up English names which began with the same but ended in two different sounds. After familiarization with natural tokens of *peas* and *piece*, during which they were instructed to only pay attention to the final consonant, participants listened and responded to the stimuli in the *praat* [1] stimulus presentation module over high-fidelity headphones in a sound-treated environment.

3.2. Results and Discussion

From the proportion of “s” responses, slopes were derived as sigmoid curves using the generalized linear mixed model function (*glmer*) [3], which calculates the proportion of the binary response (“s” or “z”) to the stimulus number (from 1 to 9), with subject as a random factor. The *glmer* function returns values for the intercept and the steepness of the slope (the larger the unsigned values, the steeper the slope, labelled as *m*), from which estimates of the category boundaries can be computed.

Figure 2 shows the psychometric functions for the two groups and the two continua. Both the EN and the DK listeners differentiated the contrast perceptually. The effects of L1 and continuum on perception are summarized as follows: First of all the L1-English listeners had steeper categorization curves than the Danes, as predicted in H1 ($\chi^2[1] = 7.2, p < 0.01$). However, listeners’ responses varied greatly, ranging from highly categorical distinction ($m = -2.44$) to very flat slopes ($m = -0.13$) for the Danish listeners. The English listeners’ responses also varied considerably, especially for the continuum *long* (m ranging from -2.26 to -0.33). As hypothesized in H2, the category boundaries depended on the vowel context, i.e. the listeners perceived more /z/ when the preceding vowel was phonetically long ($\chi^2[1] = 4.1, p < 0.05$).

Figure 2: Averaged psychometric curves showing the proportion of “s” responses for the continua and the two listener groups



H3 could not be confirmed, since the category boundaries were only shifted to a very small degree for the Danish compared to the English listeners. However, the vowel context had a significant effect on the slopes, with steeper values in continuum *short* than *long* ($\chi^2[1] = 11.0, p < 0.001$).

The results led to following conclusions: Even though the Danish listeners perceive the word-final

voicing contrast less categorically than the L1-English listeners, they seem to be able to differentiate the contrast perceptually.

4. GENERAL DISCUSSION

The present study examined how Danish learners of English as an L2 produce and perceive the unfamiliar voicing contrast /s/-/z/ in word-final position. Experiment 1 revealed that the EN speakers produced significantly different vowel:ficative ratios for the final fricative voicing contrast. The DK speakers, however, produced only a nonsignificant difference. Experiment 2 revealed that the same DK participants as in Experiment 1 differentiated the final fricative voicing contrast perceptually, albeit less categorically than the EN speakers. Both groups were similarly influenced by the vowel duration cue.

So what does this imply about second-language acquisition and the processes which learners engage in when they encounter unfamiliar contrasts? Broersma [4] argues that unfamiliar cues might never be used in a native-like manner because a sufficiently high level of accuracy can be achieved even without native-like command.

In the present study, the native Danish learners of English did not perform as well as the native English speakers, neither in production nor in perception. However, the L2 learners demonstrated some awareness of the same cue as used by the native English participants, both in production (with a nonsignificant difference in the native-like direction) and in perception (with a less categorical differentiation of the voicing contrast). Future studies are needed to determine whether the production-perception asymmetry found in the present study, with a fairly clear differentiation of the final voicing contrast only in perception, not in production, changes as a function of experience. Several previous studies of L2 speech have shown that experienced L2 learners may achieve near-native-like levels in production, whereas perception may lag behind [5, 10]. The results of these studies lead us to hypothesize that Danish speakers of English are likely to improve their use of the vowel duration cue to signal the final voicing contrast in production, whereas further improvement in the use of this cue in perception is likely to be minimal.

5. REFERENCES

- [1] Boersma, P., Weenink, D. 2014. PRAAT: doing phonetics by computer [Computer program]. Version 5.3.76, retrieved 8 May 2014 from <http://www.praat.org/>
- [2] Bohn, O.-S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium MD: York Press, 279-304.
- [3] Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., White, J. S. S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24(3), 127-135.
- [4] Broersma, M. 2005. Perception of familiar contrasts in unfamiliar positions. *J. Acoust. Soc. Am.* 117, 3890-3901.
- [5] Flege, J. E., Eefting, W. 1987. Production and perception of English stops by native Spanish speakers. *J. Phonetics* 15, 67-83.
- [6] Grønnum, N. 1998. Danish. *J. International Phonetic Association* 28(1-2), 99-105.
- [7] Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108(3), 1252-1263.
- [8] Kawahara, H., Masuda-Katsuse, I., de Cheveigné, A. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27(3), 187-207.
- [9] Port, R. F., Dalby, J. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32(2), 141-152.
- [10] Sheldon, A., Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243-261.
- [11] Yamada, R. A. 1995. Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium MD: York Press, 305-320.