

Analysing L2 Swedish word-final stops

Maryann Tan¹, Xin Xie², T. Florian Jaeger²

¹Centre for Research on Bilingualism, University of Stockholm, Sweden

²Brain & Cognitive Sciences, University of Rochester, USA

<https://doi.org/10.36505/ExLing-2019/10/0049/000411>

Abstract

We compare native (L1) and non-native (L2) word-final plosive voicing in Swedish. The L1 of the L2 speaker (Flemish) does not have word-final plosive voicing contrasts. In order to assess the effectiveness of a common approach to L2 instruction, L2 speech was elicited under two conditions: either unassisted or by playing an example L1 production and asking the L2 speaker to mimic it. Three cues to voicing—vowel, closure, and burst durations—were measured. L2 productions relied on different cues for voicing than L1 production. Mimicking reduced the difference between L1 and L2 speech.

Key words: speech, L2, Swedish, mimicking, word-final stop voicing

Introduction

Acquiring native-like pronunciation is one of the most difficult feats for L2 learners. Non-native accents are pervasive even after years of L2 learning, with potential social consequences (Fuertes et al., 2009). Formal instruction is not necessarily a significant predictor of accentedness, partly because pronunciation is not prioritised in foreign language teaching (Piske et al., 2001).

In a single-talker case study, we assess the effectiveness of a common instructional approach to L2 accent reduction—repeating (“mimicking”) words immediately after hearing the native pronunciations. The data were generated as stimuli for an ongoing project on native perception of L2 accented speech. Specifically, we compare native and non-native realizations of word-final stop voicing in Swedish, when the L1 of the L2 speaker (Flemish) has no such voicing contrast.

Materials and method

Speech stimuli comprised 32 minimal pairs of monosyllabic Swedish words (e.g. röt-röd), differing only in final stop voicing. As vowel length is phonemic in Swedish, the words of a minimal pair were always matched in phonological vowel length.

Recordings were made of a 35-year old, female native speaker. The speaker was from the north of Sweden but had lived in Stockholm for 10 years prior to the recordings. Her accent was judged as close to Central Standard Swedish (CS) by native speakers of CS. The L2 speaker was a 25-year old female native

speaker of the Central Flandrian dialect of North Belgium, with only level A1 (CEFR) knowledge of Swedish.

Recordings were made in a sound-attenuated room at the Stockholm University Multilingualism Lab. Written forms of the words were presented on a computer screen placed within a comfortable viewing distance from the speaker. Words were spoken into an Audio-Technica AT3035 microphone, placed directly in front of the speaker. Recordings were sampled at 44.1kHz. The experimenter controlled the presentation of each word.

In sessions with the native speaker, the 64 words were presented twice in random order. The speaker was prompted to produce the word when it was displayed onscreen. The two recording conditions (unassisted vs. mimicked) with the L2 speaker are described next. In both conditions, voiced and voiceless words were recorded separately in order to ease the production task for the L2 speaker.

Unassisted speech

The procedure was identical to the recordings of the native speaker, except that the 64 target words were presented three times in randomised order. This decision was made because the speaker was expected to produce a higher rate of disfluent productions, which cannot be reliably annotated.

Mimicking

Recordings were made in a session one week after the unassisted recordings. The procedure was identical to that used in the unassisted condition, with one exception: at the beginning of each trial, a recording of the native speaker producing the target word was played to the L2 speaker over Sony MDR-7506 headphones at a comfortable volume. Simultaneously and throughout the trial, the target word was displayed onscreen. An audible beep was played after 2 seconds from trial onset (after the native recording had finished playing) to cue production of the target.

Annotation

Here we analyse the first 64 tokens from each of the three conditions (L1, mimicked L2, unassisted L2). Tokens were annotated for their duration of vowel, closure, and burst. Annotations were completed in Praat (Boersma, 2001) using visual examination of spectrograms, and listening judgments. Cue boundaries were marked following conventions (Flege, Munro, & Skelton, 1992). Vowel duration was measured from the beginning of the first periodic portion of each waveform to the zero-crossing where the amplitude decreased abruptly and the waveform became sinusoidal. Burst was measured from stop release to the first zero crossing point where the amplitude became near zero. Closure was measured as the time between vowel offset and burst onset (for stops following nasals, closure onset was marked by an abrupt decline in amplitude of the nasal). Four tokens in the unassisted condition had

mispronounced vowels. These were replaced by one of the other two recordings of the same word (recall that we recorded each token three times for the non-native speaker).

Results

We compared unassisted against native speech, and against mimicked speech. Separate analyses were conducted for the three measures. All analyses used linear mixed-effects regression with maximal random effect structures (random intercepts and slopes for elicitation condition by minimal pair). Voicing was sum-coded (/d/=1 vs. /t/= -1) and elicitation treatment-coded (baseline: unassisted condition). We focus on interactions of voicing and elicitation, as this captures differences in the realization of voicing.

Native vs unassisted L2 speech

All three measures exhibited a significant interaction ($ps < .006$). Critically, the pattern of the interaction differed across the three measures. Simple effect analysis found that vowels were shorter for /d/- than for /t/-final words only in L2 speech ($p < .03$), but not L1 speech ($p > .11$). Both closure and burst were significantly shorter for /d/ than /t/ both for unassisted and mimicked L2 speech ($ps < .006$), but this difference between /d/ and /t/ was 3-times larger for native speech for closure, and 3-times larger for *unassisted* speech for bursts. The L2 speaker is thus non-native both in terms of what cues she relied on, and how they are used. Specifically, she primarily relied on a cue secondary to native speakers (burst), instead of the primary native cue (closure).

Unassisted L2 speech vs. mimicked L2 speech

All measures exhibited a significant interaction ($ps < .0002$). Simple effect analysis found that vowels were longer for /d/ than /t/ in the mimicked condition ($p < .0001$), but shorter in the unassisted condition ($p < .03$). Both closure and burst were significantly shorter for /d/ than /t/ both for unassisted and mimicked speech ($ps < .001$). But this effect was larger for mimicked speech for closure, and larger for unassisted speech for bursts. Overall, mimicked speech more closely resembled native speech (separate analyses found that mimicked speech still significantly differed from L1 speech, except for closure durations).

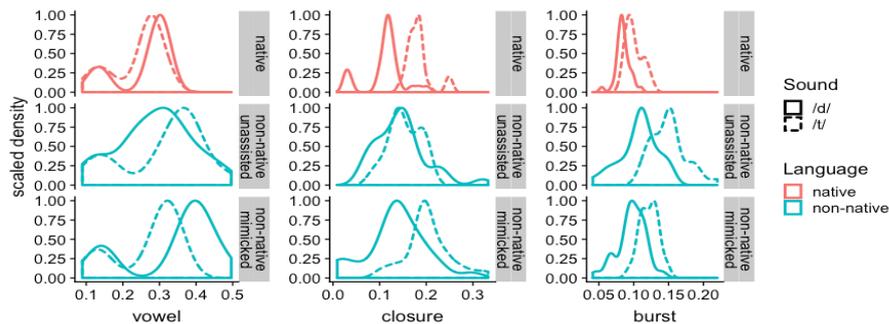


Figure 1. Density of cue durations in voiced and voiceless word pairs for native (top row), unassisted non-native (middle) and mimicked non-native speech (bottom).

Discussion and conclusion

This presents the first (small-scale) quantitative analysis of L2 word-final stops in Swedish (previous work has used phonological judgments, Hammarberg, 1997). We find that immediate repetition substantially reduces the non-nativeness of pronunciations, though our results leave open how long lasting this effect is. The specific non-native pattern we find for L2 Swedish pronunciations closely resembles the pattern found for L2 English final stop voicing by L1 Mandarin speakers (Xie et al., 2017). Both Flemish and Mandarin have stop voicing only in initial positions. The similarity in the L2 speech across these two studies might suggest that Flemish and Mandarin L2 learners transfer L1 knowledge about initial stop voicing (which relies on burst/voice onset time) to the production of L2 final stops.

References

- Fuertes, J.N., Gottdiener, W.H., Martin, H., Gilbert, T.C., Giles, H. 2012. A meta-analysis of the effects of speakers' accent on interpersonal evaluations. *Eur. J. Soc. Psychol.* 42, 12–133.
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott International*, 5 (9/10) (2001)
- Flege, J., Munro, M., Skelton, L. 1992. Production of the word-final /t/–/d/ contrast by native speakers of English, Mandarin, and Spanish. *JASA*, 92, 128–143.
- Hammarberg, B. 1997. Conditions on transfer in phonology. In A. James and J. Leather (Eds.), *Second-language speech* (161–180). Berlin: Mouton de Gruyter.
- Piske, T., MacKay, I.R.A., Flege, J.E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29 (2), 191–215.
- Xie, X., Theodore, R.M., Myers, E.B. 2017. More than a boundary shift. *JEP:HPP*, 43(1), 206–217.