Perceived quality of Swedish with a foreign accent – Comparison of speech with different temporal organization

Bosse Thorén
Department of Language Studies University of Umeå

Abstract

In an attempt to study variations in perceived quality as an effect of temporally differing Swedish with a foreign accent, recordings of one Spanish and one Estonian speaker were presented to groups of native Swedish listeners. Both inter-speaker differences as well as intra-speaker differences between intact and manipulated versions were studied. Temporal manipulations refer only to the durations of phonologically long segments. Segments were lengthened in the case of the Spanish speaker and shortened in the case of the Estonian speaker. Listeners tended to rate the pronunciation of the Estonian speaker higher than that of the Spanish speaker regardless of intact or manipulated version, whereas listeners who compared intact version to manipulated version tended to rate versions with "long sounds" higher than versions with "short sounds". A substantial part of the native Swedish listeners rated intact and manipulated versions equally, but if they rated one over the other, the "long sound" version was always rated as best pronunciation with respect to three quality-related variables.

Introduction

Adults learning a second language tend to retain a non-native accent, even after many years of experience with the target language. Accordingly, teachers of second languages can thus not expect that all students acquire a native-like pronunciation, but teachers can probably make a difference in helping students to achieve a pronunciation nearer the perfect/native pronunciation (e.g. Moyer 1999). For Swedish as a second language a strategy focusing on temporal prosody has been partially evaluated by Thorén (2008). The strategy called Basic Prosody is tested for its compliance with available findings in the field of Swedish prosody, but it still requires testing with respect to how the implemented features of Basic Prosody are perceived by native users of Swedish; do they appreciate a pronunciation that may contain many phonetic features from the first language, but follows a native Swedish temporal pattern?

Basic Prosody

The term Basic Prosody refers to a simplified description of Swedish prosody for L2-pedagogical purposes. Focusing on the lengthening of phonologically long segments is assumed to promote the signaling of both stress and quantity. Phonologically long segments are defined as either the vowel in a stressed syllable or the immediately following consonant. Spectral and tonal correlates of stress are seen as secondary to temporal correlates, and also more depending on regional variety. Similarly spectral correlates of quantity are seen as secondary and depending on regional variety. This priority is originally the result of teaching experience and complies with findings of e.g. Elert (1964), Hadding-Koch & Abramson (1964), Thorén (2003) and Fant & Kruckenberg (1994). The strategy rests on many teachers' and other listeners' appreciation of increased duration of phonologically long segments in L2-Swedish. Positive reactions to improved speech rhythm have been strong compared to improvements in the spectral realizations of phonemes.

The present study

Does an L2-speaker of Swedish sound better in native Swedish ears if the durations of phonologically long sounds are more in agreement with the Swedish length pattern, than "too short"? The question is posed both in the case of two speakers with different durational patterns and also in comparing two versions of the same speaker; one intact version and one with manipulated durations.

Thorén (2010) compared durations in phonologically long segments in the Swedish of a native Estonian speaker, a native Spanish speaker and a native Swedish speaker. It appeared that the Estonian speaker exaggerated the durations of phonologically long segments compared to the native Swedish speaker and

that the native Spanish speaker had durations substantially shorter than the native Swedish speaker. The same speakers are used in the present study where native Swedish listeners rate the spoken Swedish of the Estonian and the Spanish speaker with respect to "closeness to native Swedish pronunciation", intelligibility and "listener friendliness". Comparing the original speech of two L2-speakers of course involves many different speech parameters in addition to the duration of segments, but comparing two versions of the same speaker allows for control of the acoustic parameters. The digitally made manipulation is the only difference between the versions.

Assuming that correct timing (i.e. durational contrasts between stressed and unstressed syllables including durational contrasts between phonologically long and short segments), is essential with respect to Swedish pronunciation, enhancing this variable is expected to have a positive effect on the way native Swedish listeners perceive the quality of the L2-Swedish pronunciation.

Method

Speakers

Three speakers of Swedish as a second language were recorded when telling a short story inspired by a strip of drawn pictures showing a boy playing football in the road, being hit by a car, going by ambulance to the hospital and then coming back home with crutches and his leg in plaster, and in this condition playing with his ball again. The speakers had Spanish, Estonian and Arabic respectively as their first language. The native speaker of Arabic was used only in his intact version to check inter-rater consistency among the native Swedish listeners. One native speaker of Swedish was recorded as a reference.

Manipulations

From the Spanish and Estonian speakers two versions for each speaker were created; one intact and one with manipulated durations of phonologically long segments. The Spanish speaker had phonologically long segments lengthened towards values same or close to those of the native Swedish speaker recorded under the same circumstances. The Estonian speaker had phonologically long segments shortened towards the values of the Spanish speaker. Manipulations were only applied to phonologically long segments, with one exception: When the speakers included the unstressed – but temporally stretched - words *och* 'and' and *men* 'but' in a hesitation pause, the durations were changed as if the words had been stressed, since durational proportions became very prominent although the words were unstressed.

The cartoon-based speech sequences resulted in many common words used by all speakers, which guided comparisons and degrees of manipulations. Durational changes were guided mainly by these measurements but had in some cases to be made "by ear". Duration changes were of the magnitude 200% (Spanish speaker) and 50% (Estonian speaker).

Presentation

Intact versions of all three speakers were presented to 28 native Swedish listeners. They rated the L2-speech with respect to "closeness to native Swedish pronunciation", "intelligibility" and "listener friendliness". The speakers were rated on a scale 1-10 where 10 is always "the best Swedish pronunciation".

29 other native Swedish listeners rated the manipulated version of the Spanish and the Estonian, plus the intact version of the Arabic speaker, along the same variables as the first group of listeners.

Finally another 30 native Swedish listeners compared intact (A) and manipulated (B) versions of the Estonian and Spanish speakers in a direct comparison task, answering the questions: Is version A or B more close to native Swedish pronunciation or is there no difference? The same question was posed with respect to intelligibility and listener friendliness.

Result

Perceived pronunciation quality

The most obvious difference is between speakers as can be seen in figure 1 a-c. The female Estonian speaker was rated as "best pronunciation" in all 3 variables, and the male Arabic control speaker vas rated as second best, whereas the male Spanish speaker received the lowest scores.

The ranking between speakers in the manipulated version was the same as for the intact version and the differences between intact and manipulated versions were minimal, as shown in figure 1 a-c. As can be seen in the figures the

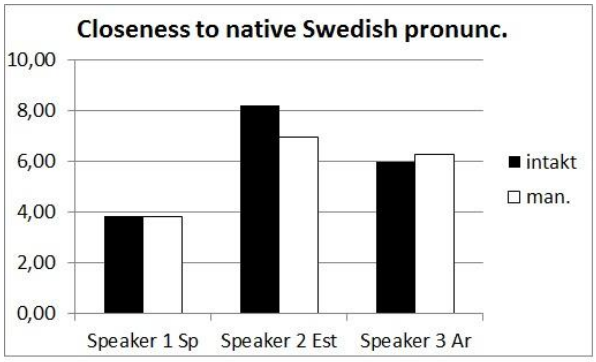

Figure 1 a. Average rating of intact and manipulated versions for the three speakers, with respect to general degree of foreign accent. Speaker 1Sp is the Spanish speaker, Speaker 2 Est is the Estonian speaker and Speaker 2 Ar is the Arabic control speaker.

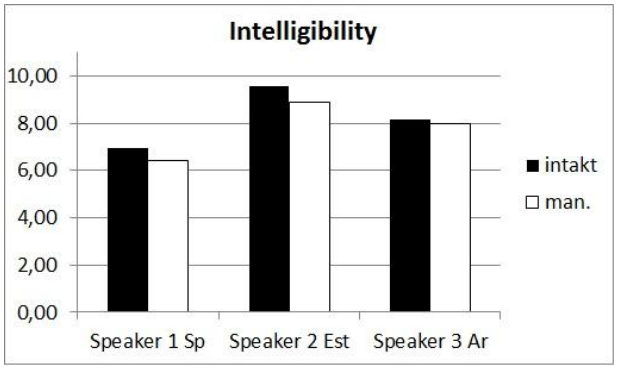

Figure 1 b. Average rating of intact and manipulated versions for the three speakers, with respect to intelligibility.

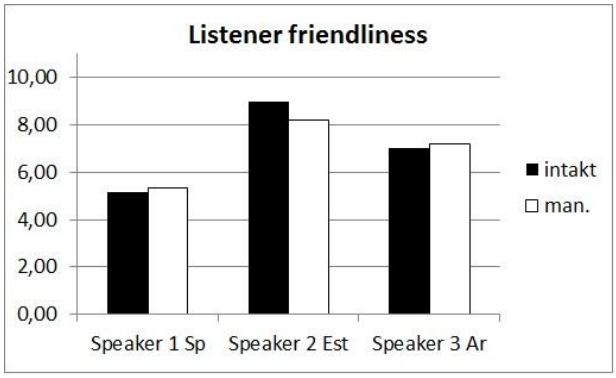

Figure 1 c. Average rating of intact and manipulated versions for the three speakers, with respect to listener friendliness.

differences in rating between the intact and manipulated version of the Spanish speaker is on the one hand inconsistent, i.e. the manipulated version is rated slightly lower with respect to intelligibility but slightly higher with respect to listener friendliness, and on the other hand of the same magnitude as the differences in rating the Arabic speaker.

A consistent tendency is however that the Estonian speaker is rated as "worse" in the version with shortened durations of phonologically long segments. This difference is significant for all quality variables ($p < 0.02$).

The two listener groups did not rate the Arabic speaker significantly different, which indicates good inter-rater agreement.

Comparison of intact and manipulated versions

Figure 2 a and b show whether the native Swedish listeners judged either of the intact or the manipulated versions as "closer to native Swedish pronunciation", "more intelligible", "more listener friendly" or equal in those respects. The solid black columns represent intact (preserved long durations) versions in the case of the Estonian speaker and manipulated (incerased durations) in the case of the Spanish speaker.

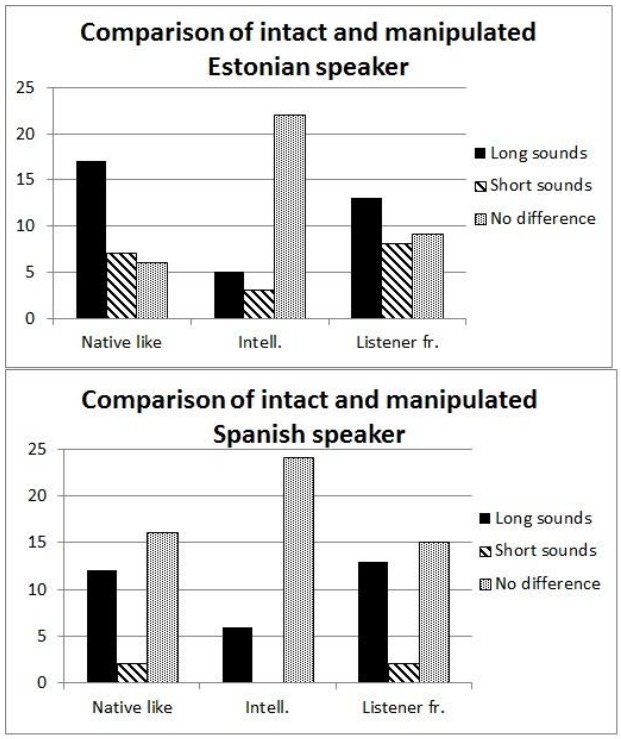

Figure 2 a (upper panel) and b (lower panel). Comparison of intact and manipulated versions. Black columns represent preference of long segments as "best", stripes rated short segments as "best" and gray columns rated the two versions as equal in quality.

As shown in figures 2 a and b, the proportion of "no difference between intact and manipulated versions" is overall high (in total 51% of all responses), which indicates that manipulations of segment durations is not perceived as changing the quality of the pronunciation to any substan-

tial degree. What is also clear is that among the responses preferring either "long sounds" or "short sounds" (in total 49% of all responses) a vast majority preferred "long sounds" (75% of those who choose either long or short sound versions). The variable of "intelligibility" receives relatively great proportions of "no difference" responses for both speakers, and the same variable gets relatively few "short sound"-responses. The general tendency seems to be: If listeners rate on version over the other, they rate the "long sound"-version highest.

Discussion

Since intact and manipulated versions were perceived differently only in the case of the Estonian speaker, when rated by independent listener groups, the result was only partly as expected. The enhanced durations made to the Spanish speaker did not receive higher scores than the intact "short sound"-version. One can speculate whether manipulations per se can reduce the "naturalness" of the speech sample and thus account for the lower ranking of the Estonian speaker version with reduced durations. Assuming that, the non-effect of the enhanced durations in the case of the Spanish speaker could be interpreted as resistance to reduced naturalness due to improved temporal pattern.

The direct comparison of intact and manipulated versions, gives a more consistent picture. Although a majority of the responses indicated no perceived difference in pronunciation quality, especially with respect to intelligibility, there were more preference for "long sounds" when judging "closeness to native Swedish pronunciation" and "listener friendliness". This indicates that both speech samples showed a high degree of intelligibility in their intact versions, but that perceived "closeness to native pronunciation" and "listener friendliness" were more affected by the manipulations.

Since "no difference" responses made up a great proportion in the case of all variables the conclusion is that the manipulations did not contribute much to how the speech samples were perceived. Therefore we must also conclude that linguistic factors other than temporal prosody influence how the L2-speech is perceived. It is obvious for a trained listener that the Estonian speaker is somewhat more advanced than the Spanish speaker with respect to grammar and vocabulary. The Spanish speaker also tends to pronounce Swedish voiced non-initial stops as fricatives. The Estonian speaker showed a generally higher level of native-like realizations of Swedish phonemes.

Teachers of Swedish as a L2 tend to react more consistently to temporal changes in a L2 pedagogical context compared to the result of the present study, which may indicate that teachers of Swedish as a L2 have trained their perception with respect to small differences in foreign accented Swedish. A small portion of the listeners in the present study belong to the mentioned teacher category.

A similar experiment could be carried out with speakers who are less intelligible in their intact speech. A task involving poetry-reading or stimuli presented together with noise could also be suitable to test whether temporal enhancement would make a bigger difference when the speech/listening task is more demanding.

References

Elert C-C (1964) *Phonological Studies of Quantity in Swedish*. Uppsala: Almqvist & Wiksell.

Fant G & Kruckenberg A (1994) Notes on stress and word accent in Swedish *STL-QPSR* 2-3/1994

Hadding-Koch, K. & Abramson, A. (1964) Duration versus spectrum in Swedish vowels: Some perceptual experiments. In *Studia Linguistica 18*, 94-107.

Moyer, A. (1999) Ultimate attainment in L2 phonology. The critical factors of age, motivation and instruction. S*tudies in second language Acquisition*, 21, 81-108.

Thorén, B. (2003) Can V/C-ratio alone be sufficient for discrimination of V:C/VC: in Swedish? A perception test with manipulated durations. *Phonum* (Department of Phonetics, Umeå University) 9, 49-52.

Thorén B. (2008) *The priority of temporal aspects of L2-Swedish prosody – Studies in perception and production.* Doctoral dissertation. Department of Linguistics, Stockholm University.

Thorén, B. (2010) Durations of phonologically long segments in native and foreign accented Swedish. Proceedings from *Fonetik 2010* Lund, June 2–4, 2010, 103-106.