THE RELATIVE CONTRIBUTIONS OF INTONATION AND DURATION TO INTELLIGIBILITY IN NORWEGIAN AS A SECOND LANGUAGE

Snefrid Holm

Norwegian University of Science and Technology
Snefrid.holm@hf.ntnu.no

ABSTRACT

An experiment is presented which investigates the relative contributions of intonation and duration to intelligibility of Norwegian as a second language (N2). Recordings of Norwegian sentences read by speakers of 7 native languages (L1s) were used. The global intonation and the phoneme durations of each N2 utterance were manipulated so as to match a native Norwegian speaker's productions of the same sentences. A perception experiment was carried out in which native Norwegian listeners wrote down the words they perceived of each N2 sentence. The results show that only intonation manipulation enhances N2 intelligibility for the English and German L1 groups and that only duration manipulation enhances N2 intelligibility for the French, Tamil and Persian L1 groups.

Keywords: foreign accent, L2, intelligibility.

1. INTRODUCTION

The goal in second language classrooms has become to enhance intelligibility rather than to eradicate foreign accent [8]. It is therefore necessary to identify which pronunciation features contribute most to intelligibility of non native speech so that teachers can focus upon these prioritized aspects. Several researchers have tried to establish such hierarchies of pronunciation deviations, among them Anderson-Hsieh, Johnson and Kohler [1], Bannert [3], Derwing and Munro [5] and Munro and Derwing [8], but the matter remains unclear. There are however several studies showing that prosodic features may be more important than segmental features (see Munro and Derwing [7] and Anderson-Hsieh and Kohler [2]).

An experiment is carried out to determine the contributions of intonation and duration to intelligibility of Norwegian as a second language.

2. METHOD

2.1. Speech material

The speech material consists of read Norwegian sentences. There were 2 speakers from each of the 7 L1s English, German, French, Tamil, Chinese, Persian and Russian. 3 different sentences from each speaker were used, yielding a total of 42 N2 sentences. The speakers were students of Norwegian as a second language at different course levels. The results will however be investigated across utterance variants from each individual speaker and not across different speakers. One native Norwegian (N1) speaker was recorded reading the same sentences. The N1 speech serves as a Norwegian template for the N2 speech. Sound files were digitized with a sampling rate of 44.1 kHz and high-pass filtered at 75 Hz.

2.2. Speech manipulations

All speech manipulations are PSOLA syntheses carried out with the Praat program [4].

2.2.1. Intonation

For each utterance, the N1 intonation contour was stylized and copied onto each corresponding N2 utterance. The intonation contour had to be corrected in the time domain because of the durational differences between the N1 and the N2 utterances. These corrections were done manually by moving the intonation points sideways in the stylized curves.

In this way a set of intonation manipulated stimuli was obtained where the N2 utterances' intonation contours had been replaced by N1 intonation contours.

With the same material, it had earlier been experienced that the manipulation methods themselves lowered intelligibility to the extent that the manipulated utterances were far less intelligible

than the original utterances. This happened both for intonation manipulation and for duration manipulation. For this reason I could not use the original N2 utterances for comparison with the manipulated N2 utterances. The intonation manipulated stimuli had in fact not only been given an N1 intonation contour, but the contour had also been stylized. For the comparison with the intonation manipulated stimuli, I wanted therefore to generate stimuli that were altered in F0 but retained the N2 intonation contour. This was done by simply stylizing the intonation contours of the original N2 utterances.

These stylized intonation stimuli serve as a baseline with which to compare the intonation manipulated stimuli. They are referred to as close-original intonation stimuli in the following.

2.2.2. Duration

For the duration manipulation, each N1 utterance and each N2 utterance were first segmented into phonemes. For each utterance, each N2 phoneme was shortened or lengthened so as to match the duration of the corresponding N1 phoneme in the N1 version of the same sentence.

In this way a set of stimuli was obtained where the phoneme durations of the N2 utterances were equal to the phoneme durations of each of the corresponding N1 utterances.

I could not use the original N2 utterances for comparison with the duration manipulated utterances for reasons explained in section 2.2.1 above. In the duration manipulated stimuli, not only the internal temporal organization of the utterances had been altered, but the duration of the whole utterances had as a side effect also been altered (thus influencing the impression of speaking rate) such that they matched the N1 utterance durations. This is because the sum of the duration of each phoneme equals the duration of the whole utterance. I therefore generated a set of stimuli where each whole N2 utterance was linearly expanded or compressed to match the utterance duration of the corresponding N1 utterance.

This set of utterance duration adjusted stimuli serves as a basis with which to compare the duration manipulated stimuli. They are referred to as close-original duration stimuli in the following.

2.3. Perception experiment

Native Norwegian informants listened to the stimuli and wrote down the words that they perceived of each sentence. Intelligibility is measured as percent correctly perceived words per sentence.

Many informants listened to two sets of stimuli whereas some listened to only one set. The learning effect arising from hearing the sentences for the second time is not discussed in this paper due to lack of space. The results will be investigated through multifactor statistical tests where learning is singled out as a factor and so does not interfere with the measurement of the factor manipulation.

3. RESULTS

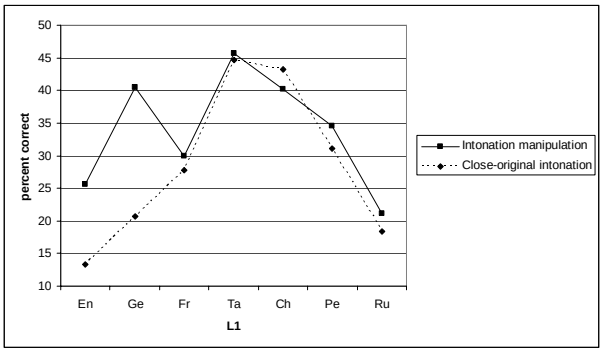
First the intonation manipulated stimuli will be compared with the close-original intonation stimuli to see whether N2 utterances become more intelligible with N1 intonation contours. Second the duration manipulated stimuli will be compared with the close-original duration stimuli to see whether N2 utterances become more intelligible with N1 phoneme durations. Thirdly the intelligibility of intonation manipulated stimuli will be compared with the intelligibility of duration manipulated stimuli to further support the findings from the two previous comparisons.

There are 38 listeners for the duration manipulated stimuli, 39 listeners for the intonation manipulated stimuli and 20 listeners for each of the close-original stimuli. The results are investigated across listeners. A subgroup of 20 listeners had listened to both the intonation manipulated and the close-original intonation stimuli, and another subgroup of 20 listeners had listened to both the duration manipulated and the close-original duration stimuli. This enables further within listener investigations by means of repeated measures tests. The within listener tests have the advantage that they eliminate the factor listener from the investigations.

3.1. Intonation

In Figure 1 the intelligibility scores of the intonation manipulated stimuli and the close-original intonation stimuli are displayed.

Figure 1: Intelligibility of the close-original intonation stimuli and the intonation manipulated stimuli.



The figure shows that the N2 utterances with manipulated intonation contours generally have higher intelligibility scores than the close-original intonation utterances. This is the case for all L1s except Chinese where there is a small opposite tendency showing higher intelligibility scores for the close-original intonation utterances.

An ANOVA for independent samples across L1s showed that intonation manipulated stimuli are significantly more intelligible than close-original intonation stimuli ($F(1, 2492) = 28.325$; $p < 0.001$)

Native speakers of English and German have the largest differences between the stimuli sets. An ANOVA for each L1 individually shows that there were only significant effects for these two L1s separately: English ($F(1, 356) = 20.233$; $p < 0.001$) and German ($F(1, 356) = 39.840$; $p < 0.001$). The differences for the rest of the L1s did not reach significance on their own.

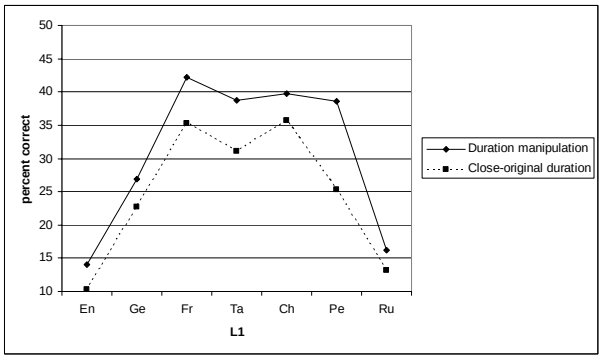
A subgroup of the listeners (n=20) had listened to both the intonation manipulated and the close-original intonation stimuli. This enabled adding investigations within listener by means of ANOVAs for repeated measures. This subgroup of listeners showed the exact same results as seen above with significant differences between the stimuli sets across L1s ($F(1, 838) = 24.891$; $p < 0.001$) and significant differences only for the L1s English ($F(1, 118) = 16.1$; $p < 0.001$) and German ($F(1, 118) = 34.543$; $p < 0.001$) separately.

Both types of comparison thus show that intonation manipulated stimuli are more intelligible than close-original intonation stimuli for the L1 groups English and German.

3.2. Duration

Next we look into the differences in intelligibility between the duration manipulated stimuli and the close-original duration stimuli. Figure 2 shows the intelligibility scores for these two sets of stimuli.

Figure 2: Intelligibility of the close-original duration stimuli and the duration manipulated stimuli.



The figure shows that the duration manipulated stimuli have higher intelligibility scores than the close-original duration stimuli. This is true for all L1s.

The differences are tested by means of an ANOVA for independent samples across L1s. This shows that the duration manipulated stimuli are more intelligible than the close-original duration stimuli. ($F(1, 2408) = 27.832$; $p < 0.001$).

An ANOVA for each L1 separately shows significant differences between the stimuli sets for French ($F(1, 344) = 5.326$; $p \leq 0.05$), Tamil ($F(1, 344) = 4.250$; $p \leq 0.05$) and Persian ($F(1, 344) = 11.295$; $p \leq 0.01$) such that duration manipulated stimuli are more intelligible than close-original duration stimuli.

For the subgroup (n=20) that listened to both the duration manipulated and the close-original duration stimulus sets, ANOVAs for repeated measures showed significantly enhanced intelligibility for the duration manipulated stimuli across L1s ($F(1, 838) = 14.332$; $p < 0.001$) and significant differences for each of the L1s French ($F(1, 118) = 4.364$; $p \leq 0.05$), Tamil ($F(1, 118) = 3.989$; $p \leq 0.05$) and Persian ($F(1, 118) = 7.613$; $p \leq 0.01$) separately such that duration manipulated stimuli are more intelligible.

Both types of comparison show that duration manipulated stimuli are more intelligible than close-original duration stimuli for native speakers of French, Tamil and Persian.

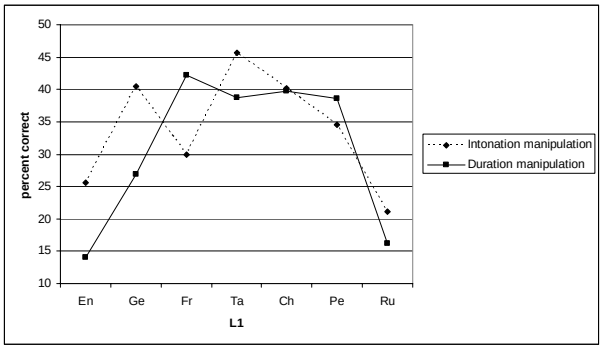
3.3. Intonation versus duration

In this section the intelligibility of the intonation manipulated stimuli and the intelligibility of the duration manipulated stimuli are compared. This

comparison is between two manipulated sets of data. They are not compared to close-original stimuli. Therefore this comparison is meant to merely support the findings in 3.1 and 3.2.

The intelligibility scores of these two sets of stimuli are displayed in Figure 3 below.

Figure 3: Intelligibility of the intonation manipulated stimuli and the duration manipulated stimuli.



In Figure 3 it is shown that for English, German, Tamil, Chinese (very small difference) and Russian speakers the intonation manipulated stimuli have the highest intelligibility scores. For the French and Persian speakers the duration manipulated stimuli have the highest scores.

The difference between the sets of stimuli is investigated through ANOVAs for independent samples. Across L1s, it is shown that the intonation manipulated stimuli are more intelligible than the duration manipulated stimuli ($F(1, 3206) = 18.041$; $p < 0.001$).

ANOVAS for each L1 separately were carried out. For English ($F(1, 458) = 28.876$; $p < 0.001$) and German ($F(1, 458) = 28.078$; $p < 0.001$) intonation manipulated utterances are more intelligible than duration manipulated utterances. This supports the results from the previous sections. The intonation manipulated stimuli are more intelligible than the duration manipulated stimuli also for the Russian ($F(1, 458) = 6.121$; $p \leq 0.05$) and Tamil ($F(1, 458) = 10.711$; $p \leq 0.01$) groups. However, no effect of intonation manipulation was found for Russian in 3.1, and the Tamil group has been shown to be affected by duration manipulation in 3.2, not by intonation manipulation. For French the duration manipulated stimuli are more intelligible than the intonation manipulated stimuli ($F(1, 458) = 16.980$; $p < 0.001$). This is in accordance with the findings in 3.2.

The results in this section cannot be easily related to the results from the previous sections, maybe because the duration manipulation automatically has affected intonation to some degree. This is because when a portion of a signal is lengthened or shortened, the slope of the intonation curve will be affected for this portion.

4. CONCLUSION

Intonation contributes more to intelligibility than duration for the English and German L1 groups, whereas duration contributes more than intonation for the French, Tamil and Persian L1 groups. An important finding is therefore that the results differ according to the speakers' native languages.

5. FURTHER WORK

In order to relate the results to deviations between each of the L1s and Norwegian, production analyses have been started, but at this stage no clear relation between deviation and effect of manipulation has been found.

The reported work is part of an ongoing larger investigation of the relative contributions of intonation and duration to both intelligibility and perceived degree of foreign accent. The investigation regarding perceived degree of foreign accent has been published [6].

6. REFERENCES

[1] Anderson-Hsieh, J., Johnson, R., Kohler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning* 42, 529-555.

[2] Anderson-Hsieh, J.R., Kohler, K. 1988. The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning* 38, 561-593.

[3] Bannert, R. 1995. Intelligibility and acceptability in foreign accented Swedish: The effects of rhythmical and tonal features. Reports from the Department of Phonetics, Umeå University. *PHONUM* 3, 7-29.

[4] Boersma, P., Weenink, D. 2006. *Praat- Doing phonetics by computer* (Version 4.4.17) [Computer program]. Retrieved April 19, 2006, from http://praat.org/.

[5] Derwing, T., Munro, M.J. 1995. Processing time, accent and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289-306.

[6] Holm, S. 2006. The relative contributions of intonation and duration to degree of foreign accent in Norwegian as a second language. *Proc. 19th FONETIK* Lund, 61-64.

[7] Munro, M.J., Derwing, T. 2005. Second language accent and pronunciation teaching: A research based approach. *TESOL Quarterly* 39(3), 379-397.

[8] Munro, M.J., Derwing, T. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 49 (Supp. 1), 285-310.