The merger of short /ø/ and /ɵ/ in Eastern Swedish – a combined production and perception test

Lena Bergström
University of Uppsala, Sweden
E-mail: lena.bergstrom@nordiska.uu.se

ABSTRACT

In parts of Sweden an ongoing sound change tends to cause a merger of short /ø/ and /ɵ/ in words such as *döma* /døma/ 'judge' and *dumma* /dɵma/ 'stupid'. This leads to several homonyms in the language and, even though it seldom causes linguistic misunderstandings, it tends to be considered as a substandard way of speaking. This paper addresses the relationship between production and perception of the two sounds. On the basis of acoustical analyses, 20 informants were selected to take part in a perception test. Half of the informants produced the merger, and half of them did not. The two groups show no major differences in the perception test, which leads to the conclusion that the sound change is subconscious and is first observed in the production.

1. INTRODUCTION

In standard Swedish, short /ø/ and /ɵ/ are very close, due to the change of short /ɵ/ from a palatal to a central vowel in the 18th century. Thereby /ɵ/ became phonetically closer to /ø/ than it had been earlier. This new situation caused problems keeping the two phonemes apart [4:2]. Today there is an ongoing sound change which makes the phonemes becoming even closer, and it is not uncommon to hear people with a complete merger. In general, people are negative towards this merger, and consider it as a substandard way of speaking.

My aim is to find out if speakers who keep the two phonemes apart in their production more easily categorise the sounds /ø/ and /ɵ/ as different phonemes than the speakers who themselves have the merger. I am also interested to find out if the two groups agree on where the phoneme boundary is located.

The relationship between speech perception and production has been a long-standing issue, and concerning Swedish vowels, Tore Janson made a perception test 20 years ago to study how people from Stockholm perceive the long /a/ (see [2]).

Janson considers every speaker to have some perception rules not based on their own production, but rules are nevertheless needed in order to understand other speakers. When you hear the new pronunciation, the norms for your own production will change and the perceptual border will move in the same direction [2:28]. However, the old perceptual norms are still needed to be able to understand 'old-fashioned' speakers when your own production has changed. This is also supported by previous studies, especially within second language acquisition where it has been demonstrated that perceptual training leads to an improvement in production, which implies a close link between perception and production [1].

This leads me to formulate a null hypothesis that predicts a close agreement between production and perception, i.e., the speakers who have the merger in their speech will also be perceptually sensitive to it. In contrast, an alternative hypothesis would predict a lack of systematic agreement between the production and perception aspects of the phenomenon.

2. SUBJECTS

In order to find speakers suitable for the test, recordings were made where subjects were asked 60 informal questions. Every question was supposed to be answered by a single word containing /ø/ or /ɵ/ in different phonetic environments. On the basis of acoustical analyses, 20 subjects were then selected to take part in a perception test. Half of the informants produced the merger, and half of them did not.

3. PERCEPTION TEST

Authentic recordings of the samples /døma/ and /dɵma/ were made in an anechoic chamber. F1 and F2 for /ø/ and /ɵ/ were measured at two points: 15 ms from the beginning and 15 ms from the end of the vowel, respectively. This is shown in Table 1:

Table 1

	F1a	F1b	F2a	F2b
/døma/	465,5 Hz	520,3 Hz	1535,5 Hz	1117,2 Hz
/dɵma/	370.7 Hz	423,4 Hz	1631,2 Hz	1282,3 Hz

ISBN 1-876346-48-5

Since F1 and F2 did not vary systematically, the distance between /ø/ and /ɵ/ was calculated for each formant and divided into 22 steps. Every second stimulus was manipulated from the original /ø/ and the remaining stimuli from the original /ɵ/. No frequency step was larger than 20 Hz to make sure that there would not be any noticeable differences between two stimuli next to each other.

Following this method, two stimulus continua were created in which F1 and F2 varied between extreme values for the respective vowels. F1 and F2 were extracted from the vowels in /døma/ and /dɵma/, and the formant contours were manipulated for each vowel using the Praat program package. In the first series, formant frequencies based on /døma/ were manipulated to gradually approximate those of /dɵma/ (series a); and in the other series (series b), the synthetic continuum went in the opposite direction to approximate the formant values of /døma/. Cool Edit Pro was used to reduce noise and to add a tone before each stimulus. As a result, 22 new vowel qualities were produced. In the perception test, each vowel quality appeared six times resulting in a total of 132 stimuli.

As language users, we are not used to identify speech sounds outside a linguistic context, and for this reason I let the subjects listen to whole word to get a more natural listening situation, which would also give a more reliable result. The stimuli were presented in random order and informants were asked to identify each stimulus as either /døma/ or /dɵma/.

The manipulations were made using data in Hertz, but since it was a perception test, the data were recalculated to Bark values to make sure that the stimuli were evenly distanced perceptually.

4. ANALYSIS

Some of the informants made their identifications consistently and without any noticeable difficulties, other informants had more problems. To examine if there was a relationship between the consistency in the answers and the informants' pronunciation, I estimated the standard deviation for every informant's identification of every stimulus.

A stimulus identified as /ø/ was assigned the value 1 and a stimulus identified as /ɵ/ was assigned the value 0. The greater the standard deviation, the more consistent the informant had been during her identifications, and had consequently interpreted the six stimuli to have the same phonemic quality in each trial. The results of this test revealed that of the eight listeners who had the smallest standard deviations, six listeners displayed a merger in their production. This result could support my hypothesis which said that those with a merger are more uncertain of which phoneme they hear, but the result is more likely to have occurred by random, since the remaining four listeners with the merger had high standard deviations and were very consistent in their identifications. There is no clear evidence that the informants in one group have a better test performance than those in the other group, but it is obvious that people who keep the two phonemes apart in their production also as a group identify the stimuli in the same way, while people with a merger either have a high or a very low score on the test.

The stimuli were numbered from 1 to 22, where stimulus number 1 was a clear /ø/ and stimulus number 22 a clear /ɵ/. The stimulus displaying a 50 % identification as /ø/ could be regarded as the boundary between the two phonemes. There are of course individual differences among the listeners, but the majority in both informant groups reported a perceptual boundary at stimulus number 11. One problem with this account is that some of the informants did not have a clear transition from /ø/ to /ɵ/; one stimulus is five out of six times identified as /ɵ/, while a stimulus with formant frequencies closer to /ɵ/ is by the same listener categorised as /ø/ every time. There is also a small difference between the two informant groups. Seven informants with merger but only two informants from the other group displayed this varied pattern.

It is three times more likely that people with merger identify the five stimuli closest to /ɵ/ as /ø/, than those in the other group. In order to examine if there were any significant differences between the two informant groups regarding how a single stimulus was categorised, I performed an $\chi 2$ test. There were significant differences ($p=0.05$) for stimuli 2, 3, 19 and 20. In all cases, those who kept the phonemes apart in their own speech were also more correct in their identification of the stimuli. Also, those with the merger categorised more often stimuli 15 to 18 as /ø/, though these differences were significant only at a lower level ($p=0.1$). It is also worth noting that those who keep the phonemes apart more often classify stimuli 1 to 13 as /ø/, while those with a merger identify stimuli 14 to 22 as /ø/ more frequently. Even though the differences are small, the results are unambiguous. People with the merger in their own production, seem to have bigger problems with the perception test.

My first hypothesis is therefore falsified. The phoneme boundary is located in the same place for the two groups, and those with the merger in their production were not more liable to identify *döma* as /dɵma/. Consequently, my second hypothesis is more reliable. The margins are not big, but the results indicate that those with a merger tend to be more uncertain and their answers are more spread both on a group level and on an individual level.

To achieve even more reliable results, it is important to make a narrow phonetic analysis of the subjects' pronunciation. It is also important that different generations of people are included in the test to examine how a sound change in progress is perceived by people at different ages. It is obvious that it may take several generations before the new sound occurs more frequently in the speech

ISBN 1-876346-48-5

community as a whole. There may still be big differences in the production between different generations, but no differences may be noted when it comes to perception.

To improve the test, the formant structure should be changed at more than two points for every stimulus. The ideal is to view the formant contour as a continuum, and then step by step move the whole contour to become more similar to the target vowel's contour.

It is also worth remembering that synthetic stimuli can never be compared completely to the natural speech, and that vowels are complex sounds and therefore cannot be described properly only through F1 and F2.

5. CONCLUSIONS

The results indicated that the informants agreed in their perception, regardless of whether they in their own speech kept the two phonemes apart or not. This suggests that we are slow to relinquish old perceptual categories in an ongoing sound change.

REFERENCES

[1] R. Akahane-Yamada, Y. Tohkura, A. R. Bradlow and D. B. Pisoni, Does training in speech perception modify speech production?, in *Proc. Fourth International Conference on Spoken Language Processing*, Vol. 2, pp. 606-609, Philadelphia, 1996.

[2] T. Janson, Sound Change in Perception: An Experiment, in *Experimental Phonology*, J. Ohala and J. Jaeger, eds., pp. 253–260. Orlando: Academic Press, 1986.

[3] W. Labov, Resolving the Neogrammarian controversy, in *Language 57*, pp. 267–308, 1981.

[4] C.I. Ståhle,: "Mötet uppnas på sundag", in *Studier i dagens svenska*, *Skrifter utgivna av Nämnden för svensk språkvård 44*, pp. 1–15, 1971.

ISBN 1-876346-48-5

ISBN 1-876346-48-5