Perception of words with vowel reduction

Johan Frid

Department of linguistics and phonetics, Lund University, Sweden

ABSTRACT

This study deals with listeners' ability to identify linguistic units from linguistically incomplete stimuli and relates this to the potentiality of vowel reduction in a word. Synthetic speech was used to produce stimuli that were similar to real words, but where the vowel in the pre-stress syllable was excluded. Listeners then performed a lexical decision test, where they had to decide whether a stimulus sounded like a word or not. The effects of the identity of the removed vowel and of features of the consonants adjacent to the removed vowel were then examined, as well as syllabic features. For type of vowel, lower word rates were found for words with the vowels /a/ and /o/, whereas words with nasals after the reduced vowel tended to result in higher word rates. Furthermore, words that still conformed to the phonotactic structure of Swedish after reduction got lower word rates than words that violated this, possibly because the conforming words are more eligible to resyllabification, which renders them as phonotactically legal *nonsense* words rather than real words.

1. INTRODUCTION

The starting point for this study was the observation that similar vowels in weak, unstressed syllables in Swedish appeared to behave differently in different words. These differences were observed when attempting to produce words with a reduced vowel in the pre-stress syllable, as they might be pronounced in spontaneous speech. Some words became difficult or uncomfortable to pronounce, since without the vowel there, the remaining consonants formed a complex articulatory cluster. Other words were easy to pronounce, but sounded like non-words or other words, or ambiguous. With some words, it was possible to reduce the vowel and still experience them as the same words. The following table shows some examples (in Swedish).

Vowel	Difficult	Other/non-word	Ambiguous	Same word
/a/	invadera, asfaltera, postdatera	relatera	debattera	permanenta
/e/	annektera		profetera	rumestera, orkestrera
/i/	inhibera	recitera, residera, presidera	debitera profitera	trafikera
/u/	intonera	resonera		motionera
/ʉ/	abducera	reducera	debutera	fokusera, strukturera

It appears as if the *reduction potential*, i.e., the possibility of a vowel to become reduced varies with different contexts. In some of the above words, the vowel in the pre-stress syllable seems necessary in order to obtain an articulation sequence of reasonable comfort. In some other words, the vowel can be left out with small or no increase in articulatory complexity. Finally, in some words, the vowel is necessary to distinguish the word from other words. In a perception perspective, this leads to the questions: how do reduced vowels affect the perception of words, and how is the perception of words with vowel reduction influenced by the segmental composition of words? The aim of this study is to examine and try to answer these questions.

2. WORD PERCEPTION

It is well known from studies of spontaneous speech (see e.g., Kohler 1990, Engstrand 1992) that words commonly seem to 'lack' certain speech sounds, that are somehow 'expected' to be found in the acoustic realisation of the words. The quotes around 'lack' and 'expected' are there since they emphasise the fact that this view is based on the idea that words in spontaneous speech are being compared to some model, archetypal pronunciation of words that corresponds to how words are pronounced carefully in isolation. This view may be correct or not; it permeates linguistic analysis of speech so that it is often taken for granted. In this article too, it will be used as a basis of discussion.

An interesting fact that is often taken as support for the above-mentioned view is that listeners very commonly ignore the fact that something is missing, in fact, it is very rare that a listener in a normal conversation even discovers that a reduction, deletion or elision occurred. This phenomenon was studied already nearly 100 years ago, by W.C. Bagley who used Edison phonograph cylinders to record words with missing consonant sounds. Listeners were instructed to repeat the words as reproduced by the phonograph as accurately as possible. Despite these explicit instructions, the subjects often restored the mutilated words to their original form. (More on Bagley's experiments can be found in Cole & Rudnicky 1983.) Another effect was demonstrated by Warren (1970), who exchanged the first [s] in the word *legislatures* for a coughing sound, but subjects were unable to hear that the cough was replacing any speech sound and reported that they had heard the whole word. The cough was heard as background noise. This 'phoneme restoration' effect indicates that listeners are very able to reconstruct a sound that never was in the speech signal from the context; they might not even detect that the sound was missing.

3. HYPOTHESES

A similar process occurs when listening to spontaneous speech. Speakers reduce and delete sounds, but listeners supplement the incomplete speech signal with information from other sources, like context and lexical knowledge and recover the intended message. This leads us to the following expectation: when presented with stimuli that according to the 'norm' is incomplete and the information left out corresponds to a deletion that normally occurs in natural speech, listeners would still be able to identify the stimuli as something linguistic, in fact, the reduction may even be expected and thus facilitate the identification. This includes words like *trafikera* and *fokusera.* However, if the deletion does not occur naturally, as in *postdatera* and *intonera* it will be harder to evoke a lexical recognition.

4. METHODS AND MATERIAL

A method of testing listeners' ability to understand 'reduced' speech is to present them with speech-like stimuli consisting of synthesised words in which one or more segments are deleted. A suitable manipulation is to remove the vowel in the pre-stress syllable, since in Swedish, a vowel in this position is often deleted or reduced in spontaneous speech. The listeners' abilities to recognise such stimuli as words will then indicate how devastating the removal is for that word. Stimuli were constructed by feeding the LUKAS (Filipsson & Bruce 1997) concatenative speech synthesiser with a transcription of a real word, but where the vowel in the pre-stress syllable was removed. This results in a direct transition between the two consonants surrounding the vowel. The words used as bases for stimuli had many different sound structures. These were chosen to cover as many combinations of manner and position of articulation as possible. Stimuli thus exhibited several consonantal patterns that corresponded to articulatory movements of several different degrees of complexity, effort and plausibility. The identity of the deleted vowel also varied. In order to avoid effects of word class, number of syllables in the word and the position of stress in the word, all the words chosen as stimuli bases were four-syllable verbs with primary stress on the third syllable. In total, 630 words were included in the test. The experiment was a 'lexical decision' test, and was run in its entirety on a SUN workstation (UNIX). Each stimulus was played once, and the test persons indicated whether a stimulus was judged as a word by pressing a key on the keyboard. Twelve persons participated in the experiment. The word order was random and different for each subject, and subjects first listened to 15 test stimuli to get used to the synthesised speech. These were not included in the analysis.

5. RESULTS

For each stimulus, the number of judgements as word was counted. The percentage of word judgements was 55,1% for all stimuli and all subjects, indicating that judgement as word was slightly better than chance. The responses were then grouped according to the following features in the stimuli:

a) The type of the deleted vowel.
b) The type of consonant sequence adjacent to the deleted vowel, e.g., *reducera* → [d] + [s].
c) The voicing sequence, e.g., *reducera* → Voiced (VD) + Voiceless (VL).
d) The articulation place sequence, e.g., *reducera* → Alveolar + Alveolar.
e) The articulation manner sequence, e.g., *reducera* → Plosive + Fricative.
f) The openness/closeness of the target syllable, e.g., *re.du.ce.ra* → open.
g) The phonotactic lawfulness of the word after deletion, e.g., *redcera* → legal.

The results for groups (a), (d) and (e) are shown in Figure 1. For the vowels, the average word judgement rate is lower for stimuli where the deleted vowels are /a/ (46%) and /o/ (45,1%) than for the other vowels (58,5% - 60,6%). This difference is significant for both vowels (/a/: $z>2.5$, $p<.006$, /o/:

z>2.54, *p*<.005). Only vowels where n>30 (occurring in more than 30 words) were included in the analysis. For type of consonant sequence, t-l (as in *gratulera*) and b-t (as in *sabotera*) have the highest rates (77% and 78% respectively), whereas k-t (as in *förskottera*) has the lowest rate, 32.2%. However, since there are many different types, the number of tokens is lower. Therefore, all combinations where n>5 were included. The rates for the voicing sequences were: VD-VL=51% (n=153), VL-VL=55,6% (n=76), VD-VD=55,8% (n=190), and VL-VD=57,4% (n=202). For a directional difference between VD-VL and VL-VD, *z*=1.889, *p*<.059, which is just under the 95% level. Regarding the place sequences, the combinations Lab-Alv and Alv-Lab get 58% and 62%. Together (n=159), their rate is significantly higher than the rate of the other place combinations grouped together (n=471), *z*=2.07, *p*<.038. Among the manner sequences (combinations shown in Fig.1 has n>10), combinations with a nasal succeeding the reducable vowel generally get high rates, whereas combinations ending with a plosive tend to get lower rates. The high rate for nasals is largely accounted for by /m/ alone. This sound, regardless of the sound that precedes the vowel, gets an average word judgement rate of 72,3%.

The distinction Open/Closed syllable was determined by checking the number of consonants after the vowel in the target syllable. One or more consonants following the vowel means that the syllable is closed. There were 164 words with a closed target syllable (Clo) and 464 open syllables (Opn). The mean word rate for Clo was 58,3% and for Opn 54%. This difference was not statistically significant, *z*=1,51. Phonotactic lawfulness was determined by checking if the consonant sequence formed after vowel deletion conformed with the phonotactic structure of Swedish (Sigurd 1965). There were 359 words with an illegal phonotactic structure when the vowel was removed and 269 word with a legal. Their average word rates were 59,3% and 49,6%, respectively. This difference is statistically significant, *z*=3.897, *p*<4.86E-05. Words that violate the phonotactic structure of Swedish thus has a higher word judgement rate.

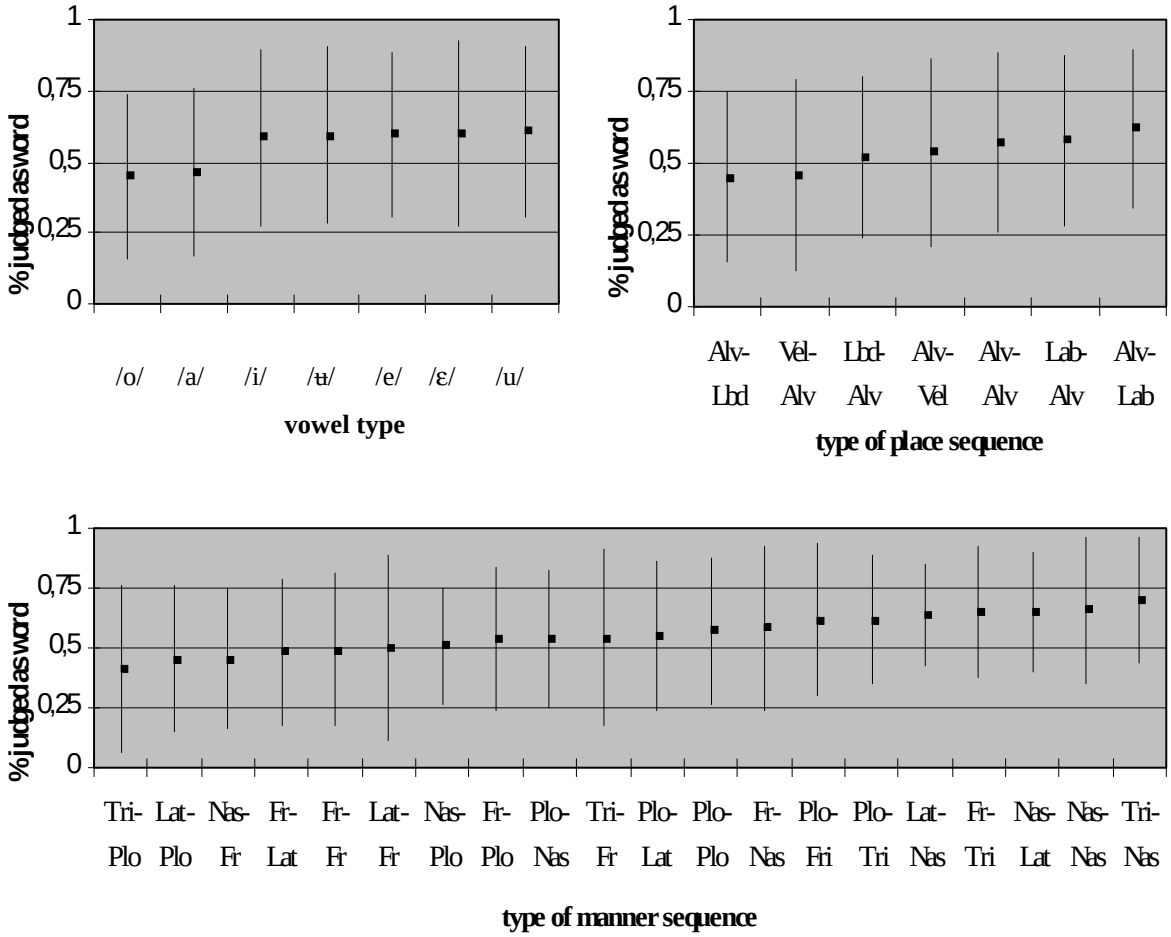

Figure 1. Average number and standard deviations of judgements as word grouped by type of deleted vowel, type of consonant place sequence, and type of consonant manner sequence.

6. DISCUSSION

When a stimulus is recognised as a word, this is interpreted as if the removal of the vowel (=simulated vowel deletion) does not change the impression of the stimulus so much that it becomes incomprehensible and uninterpretable as a word. With this interpretation, the higher the word judgement rate, the less the removal of the vowel affects the perception of a word. High judgement rates mean that this word (or group) is easier to identify in speech, which would indicate that is it more liable to be produced with a reduced vowel. High rates are found primarily when nasals succeed the vowel. Low rates are found e.g., with vowels /a/ and /o/.

The degree of openness of the syllable where the vowel is removed does not have a significant effect on the word judgement rates. Here is a possible explanation: when the syllable is closed, as in *se.lek.te.ra*, the new and possibly disturbing coarticulation is limited to one syllable, but when it is open as in *sa.lu.te.ra* the coarticulation runs across a syllable border, thus affecting two syllables. The number of syllables affected is thus less in the Clo category, which could make those words easier to recognize. On the other hand, if the syllable is closed, this means that the word possibly contains a longer consonant cluster, which could make it more difficult to recognize. We thus have factors that pull in both directions. The results indicate that there is no major difference.

Phonotactic lawfulness apparently is disadvantageous compared to violation of the phonotactic structure. Stimuli that violated the phonotactic structures of Swedish were easier to recognise as words than stimuli that still conformed to the rules after the vowel was deleted. This is perhaps contrary to what one would expect, since listeners naturally are more used to phonotactically legal sequences. However, when a stimulus is phonotactically legal the listeners may resyllabify it: *re.du.ce.ra* becomes *red.ce.ra*, thus detecting three syllables instead of four. In this case the listener might search his/her mental lexicon for a three-syllable word similar to *red.se.ra* and misses the four-syllable word interpretation. When violating the phonotactic structures, resyllabification is less likely to take place. This could explain the higher word rates for violating stimuli. It might be the case that listeners are more inclined to accept a phonotactically legal nonword than to reconstruct a real word from incomplete stimuli.

In general, the variation within groups is very high, indicating that other factors, unexamined here, also play a role. Other factors include the commonness of the word and the synthesis. The frequencies of the words have been estimated, and only a vague correlation was found between word rates and word frequencies. However, the corpus used was probably too small (3m words) and a larger corpus might result in a clearer correlation. All words were examined by the author before the experiment and judged comprehensible. However, subjects not used to synthetic speech may find some of the stimuli strange. By only analysing groups with a large number of tokens, this is not regarded as affecting the outcome in a crucial way.

7. CONCLUSIONS

Some clear tendencies regarding the segmental influence on the potentiality of vowel reduction in pre-stress syllables were observed. When missing from this position, /a/ and /o/ result in lower word rates than other vowels. Words with a nasal succeeding the reduced vowel get higher rates. In general, sounds and sound features do show some influence on the perception of words with reduced vowels, but the individual variance in judgement rates for groups of sounds and sound features indicate only a small role for individual speech sounds in word recognition and points towards a greater role for the listeners' lexicon. Phonotactic lawfulness affects the word rate negatively, possibly since this could lead to erroneous syllabification. Correct syllable structure is therefore also important.

8. REFERENCES

1. Cole, R.A., and Rudnicky, A. (1983) What's new in speech perception? The research and ideas of William Chandler Bagley, 1874-1946, *Psychological Review* 90, 94-101.
2. Engstrand, O. (1992) Systematicity of phonetic variation in natural discourse, *Speech Communication* 11, 337-346.
3. Filipsson, M., and Bruce, G. (1997) LUKAS - a preliminary report on a new Swedish speech synthesis, *Working Papers* 46, Dept. of Linguistics, Lund University, 45-56.
4. Kohler, K. (1990) Segmental reduction in connected speech in German: Phonological facts and phonetic explanations, in *Speech Production and Speech Modelling*, ed. by W.J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 69-92.
5. Sigurd, B. (1965) *Phonotactic structures in Swedish*.
6. Warren, R.M. (1970) Perceptual restoration of missing speech sounds, *Science* 167, 393-395.