Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Positional variants of selected Swedish sonorants - part 1: analysis and synthesis of /r/

Carlson, R. and Nord, L.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 30
number: 1
year: 1989
pages: 087-090

http://www.speech.kth.se/qpsr

POSITIONAL VARIANTS OF SELECTED SWEDISH SONORANTS
Part I. ANALYSIS AND SYNTHESIS OF /r/

Rolf Carlson and Lennart Nord *
Dept. of Speech Communication and Music Acoustics, KTH

Abstract

Analysis and synthesis of positional variants of /r/ are reported, and some strategies for synthesis work on sonorants are discussed. The analysis of /r/ allophones shows a complex picture of different acoustic cues. Still, it is possible to realize acceptable /r/ sounds by means of a limited set of synthesis rules.

INTRODUCTION

This is a condensed presentation of our efforts to improve the segmental quality of the text-to-speech system developed at KTH (Carlson & Granström, 1986). Our main aim is to improve the dynamics of segmental interaction and to be able to produce synthetic speech with different degrees of reduction as found in fluent speech. Specifically, we report on the analysis and synthesis of positional variants of Swedish /r/ allophones. Results from other sonorants will follow.

SWEDISH SONORANTS

A number of principles can be used in the structuring of a phoneme space:

1) distributional - phonotactic
2) distributional - morphological
3) descriptive - phonetic
4) descriptive - acoustic

A distribution analysis of Swedish sonorants shows that they occur in a variety of positions, (Sigurd, 1965 and Elert, 1970). As members of clusters, in unstressed endings (suffixes), etc., their acoustic shape varies a great deal and they are sometimes heavily reduced.

A rule search in part of the data bank material at KTH supplied us with data on /r/ occurrences in different contexts, (see Fig. 1). The total number of consonants were 7200 in this material. Of the 765 /r/ phonemes, 361 occurred in unstressed syllables, 245 in stressed syllables ('V:C, 'CV), 58 in stressed, tense positions ('VC:) and 101 in prepausal positions. As many as 500 phonemes occurred in a consonantal environment, which accentuates the need for coarticulation rules in the synthesis. If the material is divided in morphological categories, we find 100 /r/ phonemes as part of the unstressed suffix /er/. The large number of prepausal /r/ phonemes also indicates the importance of rules for those variants.

* Names in alphabetic order.

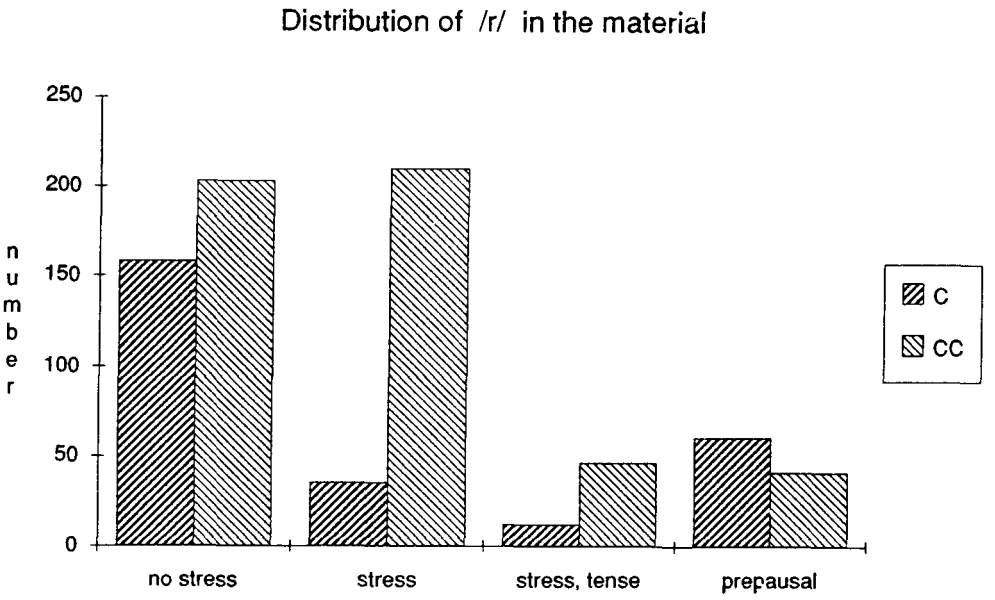

Fig. 1. Distribution of /r/ in the material.

ACOUSTIC REALIZATIONS OF /r/

The Swedish /r/ has earlier been treated in the literature by a number of researchers. The acoustic realization of the different /r/ sounds has been studied, and partly reported earlier (Nord, 1987). Although there is a fair amount of general knowledge about the acoustics of /r/, there is still a lack of data concerning the reduced forms and the acoustics of retroflex coarticulation. At the acoustic level, several patterns will be seen which do not correlate in a very systematic fashion with the phonetic inventory. Sometimes reduced versions appear as an amplitude dip, an F1 dip, an F3 movement, etc.

In order to clarify the picture, at least for one person's pronunciation habits, we made spectrograms of the /r/ sounds in the data bank and classified them according to acoustic criteria. These spectrograms were used in the design of synthesis patterns (see below).

In our material we recognized five acoustically different allophones that quite systematically occurred in a number of different /r/-contexts: (i) a vocalic, sometimes very much reduced /r/, mostly in VCV context, (ii) a fricative /r/ in initial prestressed syllable positions and in clusters containing a fricative or an unvoiced stop, (iii) a flapped /r/ after voiced stops (cluster), (iv) an /r/ mainly signalled by a delayed formant transition rate in sonorant sequences, such as /rja/ /lra/ and preceding unvoiced stops or fricatives, (v) /r/ in sentence final position, realized as a constriction followed by a release gesture.

SYNTHESIS OF /r/-VARIANTS

In this section we will give some explicit examples of synthesis rules. A vocalic /r/ was chosen as a base form for the synthesis. The target for each formant is 150 Hz lower than the neutral vowel. The F4 has a still lower value corresponding to a retroflex tongue position. The time constants for all formants were specified in such a way that a minimum was created for each formant in the middle of the segment. The parameter tracks can be seen in Fig. 3a. In this "default version" of /r/, a voiced source is used with a minimum value in the middle, just as for the formants.

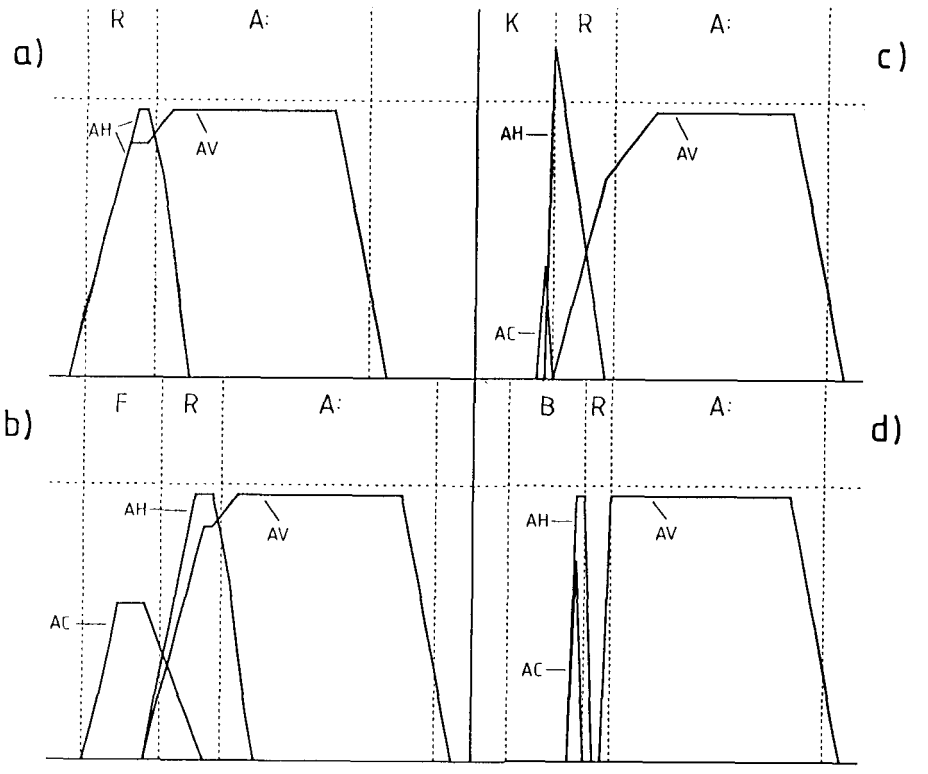

Fig. 2. Source amplitudes for /ra:/, /fra:/, /kra:/, /bra:/. Frication sources are marked by AC and AH, voiced source is marked by AV.

Our next step was to formulate rules for a fricative allophone. The /r/ was marked as fricative when occurring after unvoiced consonants, in word-initial position and in a prestressed position. The realization rule is very simple: give the source amplitude parameters a time constant corresponding to the segment length and add a noise source. The voicing is also reduced at the same time. Fig. 2b shows the excitation level slowly rising as the frication after the /f/ is decreasing. A similar pattern can be seen in Fig. 2a, where the /r/ occurs in initia. After an unvoiced stop, (Fig. 2c), the noise source is falling off during the whole segment, just as it did after the fricative. The frication is filtered by the formants and has a close similarity to aspiration noise.

Next, a flapped /r/ that often occurs after voiced stops, is shown. We have chosen to synthesize it by making the duration short, and to have fast movements of the source amplitudes. No frication is added after the burst.

So far, we have shown examples of rule-based adjustments that create specific variants of the /r/ phoneme. The adjustments have been accomplished by source variations only. The next form is an example of a change in the formant pattern. Fig. 3c shows control parameters for an intervocalic /k/. A combination with the intervocalic /r/ allophone, Fig. 3a, will not create a good /rk/ cluster. The first formant transition will be unnatural and result in poor synthesis quality. By adding a rule that adjusts the time constant for the first formant, we get a realistic slow fall of the formant into the occlusion of the consonant (Fig. 3b). If the dip in the middle of the /r/ had been kept, we would have had the beginning of a schwa vowel insertion. More specific rules for vowel insertions are under development.

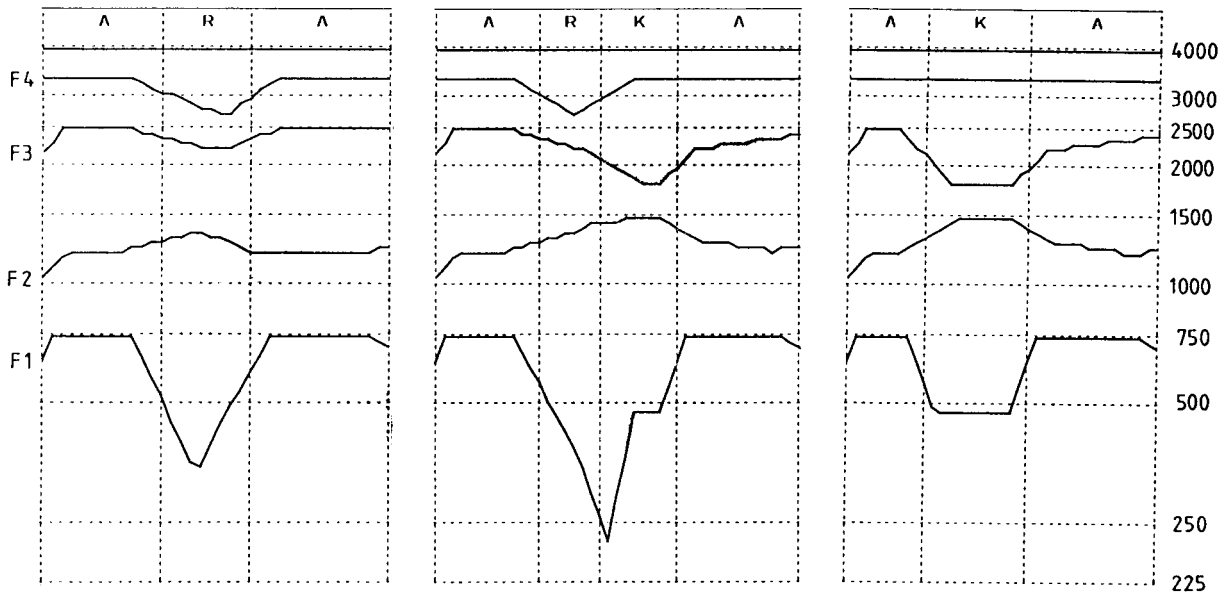

Fig. 3. Formant parameter tracks for /ara/, /arka/, and /aka/.

We have tried to illustrate how a sequence of a few coarticulation rules can model some important variants of the Swedish /r/ phoneme. Additional work will incorporate more allophones used by our reference speaker as well as some dialectal variants.

Acknowledgments

This work has been supported by a grant from the Swedish Board for Technical Development and Swedish Telecom.

References

Carlson, R. & Granström, B. (1986): "Linguistic processing in the KTH multilingual text-to-speech system," *Conf.Rec. IEEE-ICASSP, Tokyo.*

Elert, C.-C. (1970): *Ljud och ord i svenskan*, Almqvist & Wiksell, Stockholm.

Nord, L. (1987): "Some /r/-Realizations in fluent Swedish", paper given at a Symposium at the Dept of Linguistics, Stockholm University, 'The Phonetics and Phonology of Speaking Styles', 4-5 May 1987 (unpublished).

Sigurd, B. (1965): *Phonotactic structures*, Gleerup, Lund.