Simplification of phonotactic structures in unscripted Swedish

Olle Engstrand
Diana Krull
Department of Linguistics, Stockholm University
email: olle@ling.su.se

Informal listening suggests that unscripted Swedish shows a tendency to produce alternating contoid and vocoid articulations which relate to more complex consonant and vowel structures at the phonological level. To test this hypothesis, two unscripted monologues and, for comparison, a careful text reading were analyzed. The speech material was segmented using criteria based on the so-called sonority hierarchy. The results largely corroborated the hypothesis in showing that contoid-vocoid units appeared considerably more frequently in unscripted speech than suggested by conventional phonotactic analysis, and that some reduction effects appeared more frequently in the unscripted than in the read speech. The possibility that this reflects an underlying articulatory organization of spontaneous speech in terms of typologically basic CV units is discussed.

1 Introduction

Swedish grammars usually describe words and morphemes as having a complex phonotaxis permitting heavy consonant clusters in all positions (Sigurd 1965). However, clear instantiations of these structures appear to be relatively rare in most varieties of unscripted speech. Informal listening rather suggests the hypothesis that sound sequences produced in casual speaking styles are usually simpler than suggested by a phonological analysis. In particular, there appears to be a tendency to produce alternating contoid and vocoid articulations which relate to more complex consonant and vowel structures at the phonological level. Informal listening and preliminary measurements further suggest that these reduction effects may result in substantial temporal adjustments. These adjustments, in turn, appear to invite the hypothesis that the duration of syllable-sized contoid-vocoid sequences is less variable than the duration of the more elaborated syllable-sized units found in formal speaking styles.

The present paper focuses on the former of the two above-mentioned hypotheses, i.e., that Swedish unscripted speech tends towards simplification of phonotactic structures. The latter hypothesis, that casual speaking styles display a greater tendency towards 'syllable timing' than do more formal speaking styles, will be examined in a companion paper. It may, however, be a consequence of the former hypothesis since reduction effects such as cluster simplification and syllable opening may tend to

DOI:10.1017/S0025100301001049 Journal of the International Phonetic Association (2001) 31/1

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

equalize the amount of time that must be spent on each syllable-sized unit. A syllable-sized analysis unit would seem to be suitable in both these studies. The reason is that syllable structure provides a crucial basis for specifying rhythmic structures in speech. In particular, while CV-like units seem to appear relatively frequently in casual speaking styles, CV-syllables are also typologically outstanding in the languages of the world (e.g. Ruhlen 1987). It is, thus, worthwhile testing the idea that, from a typological point of view, Swedish may turn out to be rhythmically and phonotactically more basic when considered at the phonetic rather than at the phonological level.

2 Segmenting unscripted speech in terms of contoid-vocoid sequences

Based on the above considerations, segmentation was carried out to identify all contoid-vocoid sequences occurring in an unscripted speech material. This segmentation method, of course, differs in crucial respects from a conventional morpho-phonologically based syllable parsing (e.g. Gårding 1967). In particular, a phonetic segmentation strictly based on syllable-sized articulatory contoid-vocoid sequences is bound to disregard structurally defined syllable divisions. Consider, for example, the word *kappa* 'coat', which is conventionally syllabified as *kap+pa*. The /p/ is usually considered a geminate the elements of which make up coda and onset of the respective syllables. The structural reason is that short vowels appear in stressed syllables only when followed by a consonant cluster or a long consonant. Regarding the long consonant as a geminate simplifies the rule since only one structural condition needs to be taken into account. This analysis, then, follows from language-specific, formal considerations. In terms of contoid-vocoid sequences, however, the word *kappa* may just as well be regarded as consisting of the sequences [ka] and [pa], irrespective of whether traces remain in the first vowel of the following bilabial articulation (see below).

For another example of a structurally motivated syllabification, the word *matsal* 'dining-room' is represented as *mat+sal*. From an articulatory point of view, however, it may also be represented by the contoid-vocoid sequence [ma] followed by the contoid-vocoid-contoid sequence [tsal], where [ts] stands for an affricate. This example makes the important point that, even though affricates are not recognized as PHONOLOGICAL units of Standard Swedish, there is no obvious reason to assume that they could not be PRODUCTION units. Likewise, the phrase *en dag* 'a day' may in casual speech correspond to the sequence [ⁿdɑ] with a prenasalized [d] (the word *dag* is pronounced without the final /g/ even in fairly elaborated speaking styles). Thus, even though prenasalized stops do not form part of the Swedish phoneme inventory, they are quite common in casual speech and may well constitute underlying articulatory control units. Of course, this view of the nature of the motor input to the speech production mechanism contrasts with gesture-based accounts of phonetic reduction such as that of Browman & Goldstein (1992).

As a rule of thumb for dividing consonant sequences, Jespersen's so-called sonority hierachy was used (e.g. Jespersen 1926). Even though the sonority hierarchy was formulated for an entirely different purpose, it turned out to be a suitable device for dividing up clusters in an intuitively natural way. The original purpose of the sonority hierarchy was, of course, to predict the order in which sounds tend to appear at syllabic onsets and codas. Specifically, consonants tend to appear in order of increasing sonority in syllable onsets, and in the reversed order in syllable offsets (the 'mirror effect'). This is illustrated by the Swedish word *brand* [brand] 'fire', in which the

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

relatively sonorous [r] and [n] segments are found adjacent to the syllable nucleus, while the less sonorous voiced stops appear in peripheral positions.

For the present purpose, sequences of contoids were considered as onsets as long as they constituted well-formed initial clusters according to the sonority scale. For example, an unusually clear rendering of the words *avbrott* 'interruption' and *anse* 'consider' (grammatically *av+brott* and *an+se*) would be parsed as [ɑv] plus [brɔt] and [an] plus [se], respectively. Alternative segmentations such as [vbrɔt] or [nse] are not well-formed according to the sonority hierarchy. However, less exaggerated pronunciations that occur more commonly in unscripted speaking styles would typically sound like [ɑ^{v}brɔt] and [ãse], respectively. The [ɑ] of the former word would normally have a slight diphthongization (as a trace of the /v/ segment or as a transition towards the /b/, or both); and in the latter word, an amount of nasalization would be preserved as a residue of the nasal consonant. Thus, the inital syllables in [ɑ^{v}brɔt] and [ãse] will usually display traces of the 'underlying' consonant. Articulatorily, however, [ɑ^{v}] and [ã] are vocoids in the sense that they are produced with an open vocal tract. (The use of spectrograms to verify auditory impressions is described below and illustrated in figure 1.)

3 Speakers, materials and procedures

The unscripted speech material analyzed for the present study comes from lively monologues produced by two male native speakers (JS and RL) of a Stockholm variety of Central Standard Swedish. At the time of the recordings, both subjects were in their early forties and staff of the Department of Linguistics. The recordings were made in 1988 as part of the collection of spontaneous speech materials for a project on speaking styles. The monologues were supported by brief comments and questions by one of the experimenters. To enable a comparison with a different, more formal speaking style, a carefully read text was recently recorded in Central Standard Swedish with a female actress in her forties (GT). All recordings were made in an anechoic chamber at the Phonetics Laboratory using high quality professional equipment.

The recorded speech was digitized at 16 kHz, labeled and analyzed using the Soundswell 3.50 program package (http://www.hitech.se/development/products/soundswell.htm). The analysis was made of parts of the respective recordings (JS: 2.8 minutes, RL: 3.8 minutes, GT: 6.1 minutes). Part of the material was analyzed jointly by both authors. For the remaining part of the material, individual analyses were first carried out by both authors, and the agreement between the analyses was checked later. The segmentation was usually rather straightforward and occasional disagreements were settled quite easily. The analysis was basically auditory, but spectrograms were used to verify the auditory impressions.

In a number of phonological VC_1C_2V structures (such as the word *anse* 'consider'), traces of C_1 could frequently be heard in the adjacent vowel, as noted above. In such cases, spectrograms were examined extra carefully. Presence of consonantal properties (such as weakening of higher formants, turbulence noise or silence) was taken as indicative of a contoid, whereas a lack of such properties was taken to suggest that the sequence in question ended with a vocoid. Occasionally, however, auditory and spectrographic criteria did not agree completely. This problem is typified by the spectrograms of the word *anse* (spoken by author OE) in figure 1. The left spectrogram shows a careful pronunciation of the word with a clearly visible and audible nasal. In the right spectrogram, which represents a more casual pronunciation, a fully developed nasal cannot be heard. However, the spectrographic representation leaves open the possibility that there may have been a brief moment of nasality combined with oral

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

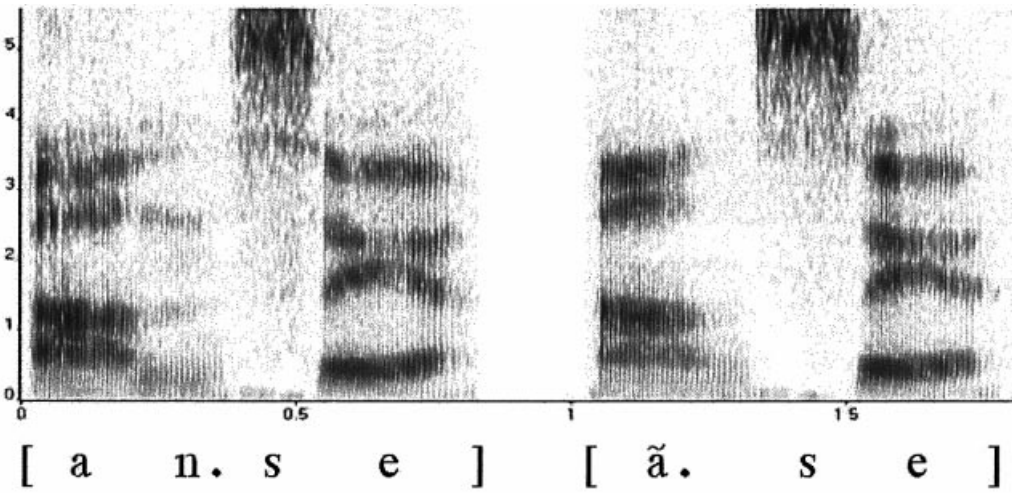


Figure 1 The left spectrogram represents a careful pronunciation of the word *anse* 'consider'. The nasal consonant segment is clearly audible and visible. The right spectrogram illustrates a relatively casual pronunciation of the same word in which the nasal segment cannot be heard but may be weakly represented on the spectrogram. Further discussion in text.

closure giving rise to a nasal contoid. In such a case, the unambiguous auditory impression would nevertheless take precedence as long as it was shared by both experimenters. These cases, then, would be analyzed as the sequences [an] plus [se] and [ã] plus [se], respectively, the latter thus having the phonetic structure vocoid plus contoid-vocoid.

4 Results

In the first part of this section, the incidence of 'open' contoid-vocoid sequences (as well as the relatively few instances of single vocoids and sequences beginning with clusters) as identified according to the above criteria is compared to the proportion of syllabic units obtained using phonological criteria. The latter units are sub-morphemic, mono-morphemic or bi-morphemic. The disyllabic word *kappa* 'coat', for example, constitutes a single morpheme of which *kap* and *pa* are phonologically motivated constituents, as noted above. In contrast, the monosyllabic word *ser* 'sees', is a bi-morphemic combination of the stem *se* 'see' and the conjugational suffix *-r* (present tense). And, finally, a word such as *fram* 'forward' constitutes a monosyllabic morpheme unit.

4.1 Incidence of contoid-vocoid sequences

A comparison of the incidence of contoid-vocoid sequences vs. phonologically determined syllables is shown in table 1. For each speaker, numbers and percentages pertaining to the phonetic sequences are shown in the first row, and data pertaining to the corresponding phonological units are given in the second row. The total number of syllable-sized units, phonetic as well as phonological, is shown in the column headed 'No. of syllable-sized units'. It can be seen in this column that the number of contoid-

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

Table 1 Incidence of phonetic sequences vs. phonological syllables in unscripted speech (male speakers RL and JS) and careful reading (female speaker GT).

Speaker	Analysis	No. of syllable-sized units	V		CV		CCV		'Open' syllable-sized units	
			No.	%	No.	%	No.	%	No.	%
RL	Phonetic	822	20	2	482	59	97	12	599	73
	Phonological	900	16	2	204	23	21	2	241	27
JS	Phonetic	453	18	4	284	63	58	13	360	81
	Phonological	491	25	5	158	32	12	2	195	40
GT	Phonetic	977	46	5	467	48	92	9	605	62
	Phonological	997	21	2	274	27	20	2	315	31

vocoid sequences resulting from the analysis was smaller than the number of phonologically defined syllables, particularly in the unscripted speech (RL and JS). This difference was due to various phonetic contraction effects (see below).

In the V, CV and CCV columns, the number of observations and the corresponding percentages are shown to the left and right, respectively. With regard to the phonetic analyses, V, CV and CCV stand for single vocoids, contoid-vocoid sequences and contoid-contoid-vocoid sequences, respectively. The righmost column ('Open syllables') simply shows the sums of the V, CV and CCV columns. The data show that there are more phonetic sequences with final vocoids than there are phonologically open syllables. For example, the figures for subject RL indicate that 73 percent of all identifiable sequences ended in a vocoid whereas only 27 percent of the phonologically defined syllables were open. It is noteworthy that this effect is equally clear for subject GT's careful reading. The phonetic analysis yielded similar results for the three sequence types (single vocoid, contoid-vocoid, contoid-contoid-vocoid) produced by the male subjects, except for the single vocoid units. Also, the number of contoid-vocoid sequences was much greater than the number of contoid-contoid-vocoid or single vocoid units.

4.2 Some illustrative examples

The previous section has demonstrated that, in both spontaneous speech and text reading, the number of phonetic sequences ending in a vocoid was greater than the number of open phonological syllables. This difference is partly an automatic effect of the parsing criteria, and partly a result of a number of phonetic reduction processes. Some illustrations are provided in the next few paragraphs.

According to the parsing criteria, single contoids followed by vowels were counted as sequence-initial, and clusters of contoids were also regarded as initial if well-formed according to the sonority hierarchy. An example is given in figure 2. The spectrogram represents the phrase *fram emot att* 'forward to'. Here, the phonetic parsing criteria convert the phonologically defined syllable sequence *fram+e+mot+att* into [fr̥ã.me.mu.tat], where periods mark boundaries between contoid-vocoid sequences. Thus, there are three 'open' sequences and one 'closed' sequence. This automatic

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

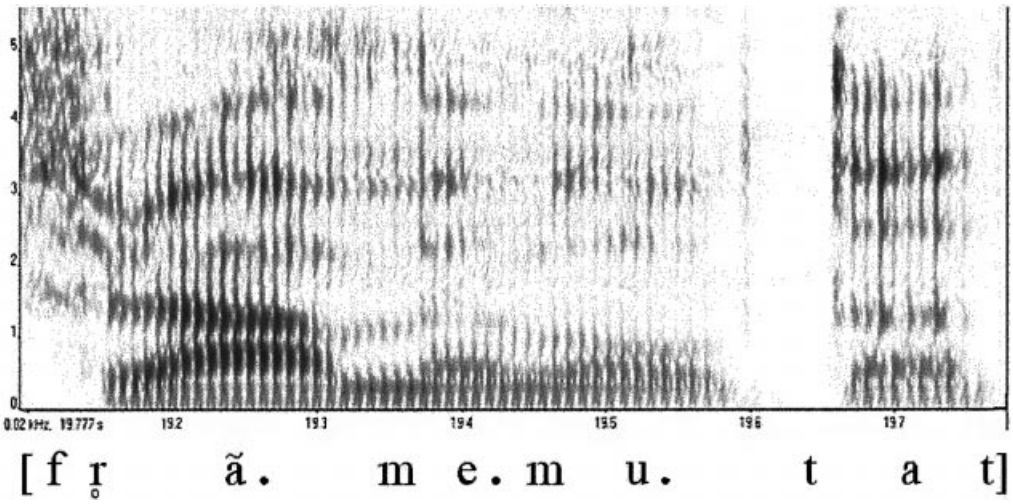


Figure 2 Spectrogram of the phrase *fram emot att* 'forward to' [fr̥ã.me.mu.tat]. Boundaries between contoid-vocoid sequences are marked with periods.

transfer of syllable-final consonants to sequence-initial position accounted for 39 percent of all cases in which an 'open' phonetic sequence corresponded to a closed phonological syllable in speaker JS, and for 44 percent in speaker RL. In the read speech produced by speaker GT, the corresponding proportion amounts to 82 percent.

In the spontaneous speech material, syllable-final consonants frequently did not appear as contoids in the phonetic analysis; they were 'deleted' in this sense even though traces may have been left in the preceding vowel. This effect accounted for 30 percent of all cases in which an 'open' phonetic sequence corresponded to a closed phonological syllable in speaker JS, for 36 percent in speaker RL and for only 12 percent in speaker GT. Since this effect was much less prominent in the read speech, it may be a genuine feature of unscripted speech. Several additional effects resulting in phonetic simplifications in the unscripted speech are exemplified below.

The majority of syllable-final consonants that did not correspond to contoids in the phonetic analysis were nasals and rhotics. However, missing nasal contoids frequently left a clearly audible amount of nasality on the preceding vocoid (cf. discussion above). For example, figure 3 shows a spectrogram of the phrase *och sen så* [ɔsẽsɔ] 'and then'. In this phrase, the word *och* 'and' is pronounced [ɔ] as usual in most speaking styles. This pronunciation of *och* thus exemplifies a 'phonologized' final C deletion. The [sẽsɔ] portion derives from *sen så*. It thus provides an illustration of a missing nasal contoid leaving a nasal residue in the preceding vowel.

Figure 4 illustrates several simultaneous effects resulting in the phonetic phrase [sã.na.dʲẽ] deriving from *så han hade ju en* . . . 'so, of course, he had a . . .'. The relationship between the six syllables and the three contoid-vocoid sequences involves:

a) two cases in which an /h/ phoneme does not leave visible traces in the spectrogram: *han* 'he' and *hade* 'had';
b) the adjacent vowels resulting from absence of /h/ in *så han* (phonologically /so han/) are merged into a slightly diphthongized and nasalized [a];
c) the following sequence becomes [ⁿa] because of absence of /h/ in the word *hade*;
d) the final vowel of *hade* is elided, and the /j/ of the particle *ju* 'of course' appears as palatalization of the (approximantized) /d/; and

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

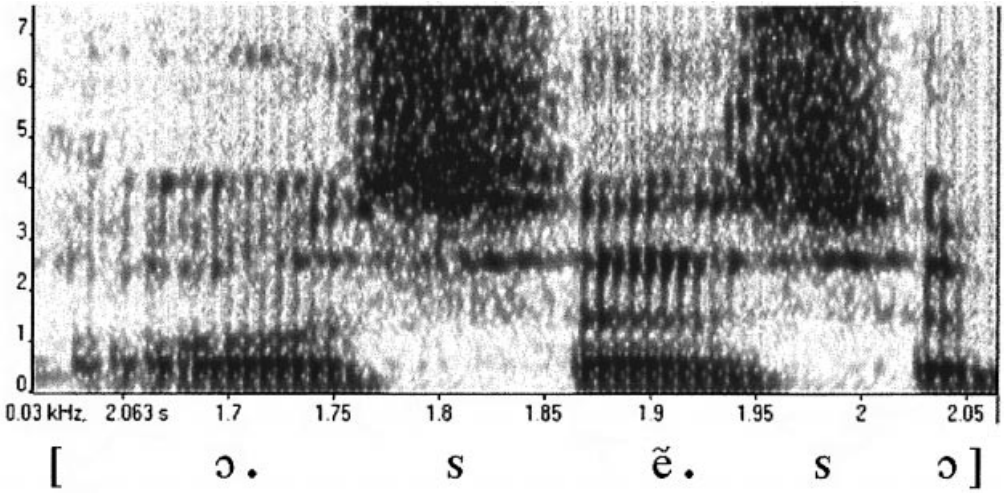


Figure 3 Spectrogram of the phrase *och sen så* [ɔ.sẽ.sɔ] 'and then'. Speaker JS.

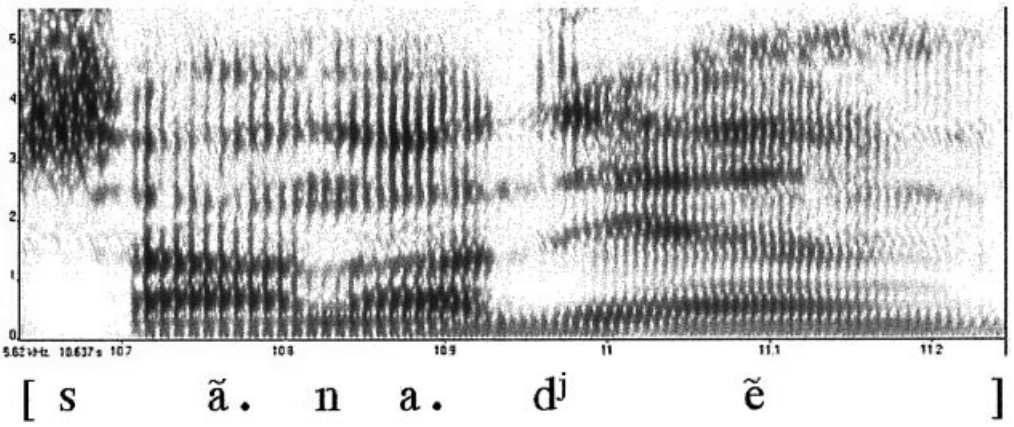


Figure 4 Spectrographic representation of the phrase *så han hade ju en . . .* 'so, of course, he had a . . .' [sã.na.dʲẽ]. Speaker JS. See text for further discussion.

e) the phrase-final nasal (from the indefinite article *en*) is heard as nasalization on the preceding vowel (the possible brief nasal contoid suggested at very the end of the spectrogram is not audible; cf above discussion).

It is noteworthy that these effects are not strictly linear. In many cases, a 'feeding rule order' is implied such that one effect provides the necessary precondition for another effect to take place. In point (b) above, for example, the absence of /h/ caused a hiatus situation which, in the next step, was resolved by elision of one of the vowels. It should be noted, however, that such a rule scheme is purely descriptive and does not necessarily reflect the structure of the utterance plan on a one-to-one basis.

Figure 5 provides further illustrations of phonetic processes resulting in a number of contoid-vocoid and contoid-contoid-vocoid sequences. Again, some of these processes must be described as ordered in relation to one another. The phrase is *därför att*

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

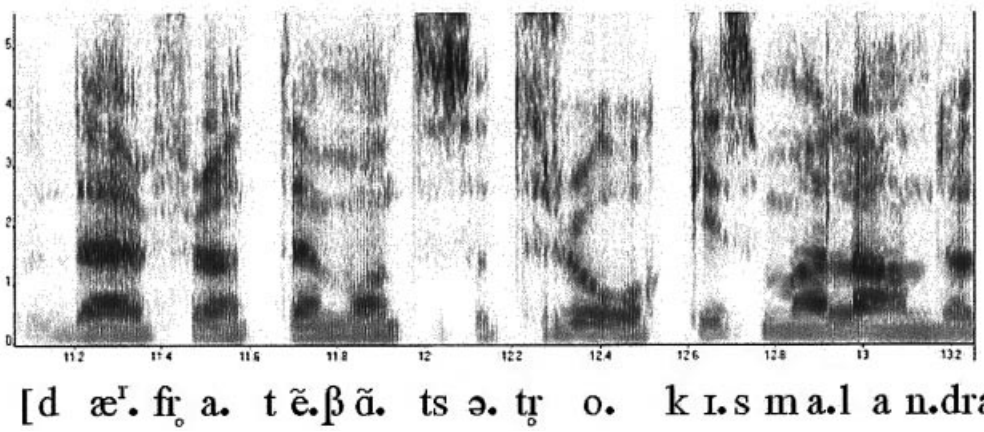


Figure 5 Spectrographic representation of the phrase *därför att den var inte så tråkig som alla andra* 'because it was not so boring as all the others' [dæʳ.fr̥a.tẽ.βɑ̃.tsə.tr̥o.kɪ.sma.lan.dra]. Speaker JS. See text for further discussion.

den var inte så tråkig som alla andra 'because it was not so boring as all the others'. The parsing criteria yield the analysis [dæʳ.fr̥a.tẽ.βɑ̃.tsə.tr̥o.kɪ.sma.lan.dra]. The relationship with the phonological syllable structure can be described as follows:

a) The first /r/ (pertaining to the syllable *där*, which is part of the word *därför* 'because') appears as a slight rhotic quality at the end of [æ] resulting in the contoid-vocoid sequence [dæʳ];
b) the vowel in *för* does not correspond to a vocoid in the spectrogram, and the voicelessness of the /f/ spreads progressively to the now adjacent [r]; this yields the contoid-contoid-vocoid sequence [fr̥a];
c) stop devoicing and the absence of a contoid reflecting the final nasal consonant in *den* 'it' gives the contoid-vocoid sequence [tẽ];
d) *var inte så* 'was not so' materializes as [βɑ̃.tsə] – the /r/ of *var* 'was' is dropped (this is a phonologized deletion), resulting in [ɑ] and [i] coming together; the [i] has hardly left any trace in the spectrogram (this, too, is probably a phonologized deletion). The [n] of *inte* 'not' appears as nasality on the /ɑ/, and the /t/ of the same word forms a phonetic affricate with the /s/ of the following word *så* 'so'; this affricate initiates the following contoid-vocoid sequence [tsə] (with [ts] regarded as an affricate);
e) the final /g/ of *tråkig* 'boring' is deleted to give still another CV (this is a phonologized deletion of the derivational suffix *-ig* which is also met with in less casual speaking styles);
f) the words *som alla andra* 'as all the others' correspond to the syllables [sma.lan.dra]. Here, the /ɔ/ vowel of *som* 'as' does not correspond to a vocoid. This results in the sequence [sma], where the [a] comes from *alla* 'all'; and finally,
g) the remainder of *alla* enters into the contoid-vocoid-contoid sequence [lan] which involves contraction of the two /a/ vowels: This is the only phonetically 'closed' sequence in this phrase according to the employed criteria. In summary, then, this sentence has 15 phonological syllables and 10 phonetic syllable-sized sequences, of which 9 are 'open' and one is 'closed'.

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

5 Summary and conclusions

The purpose of this study was to test the hypothesis that sound sequences produced in unscripted Swedish speech may be greatly simplified in relation to phonotactic structures. Specifically, it was hypothesized that Swedish has a tendency to produce alternating contoid and vocoid articulations which relate to more complex consonant and vowel structures at the phonological level. An additional hypothesis, to be tested later, was that these reduction effects may result in temporal adjustments such that the duration of syllable-sized units in spontaneous speech is less variable than the duration of the corresponding, segmentally more elaborated sequences found in more formal speaking styles. Segmentation in terms of syllable-sized contoid-vocoid sequences was carried out using criteria based on Jespersen's sonority hierarchy. The results largely corroborated the hypothesis in showing that, as indicated by the high incidence of simple contoid-vocoid sequences, alternating contoid and vocoid articulations were more frequent in the unscripted speech than suggested by conventional phonotactic analysis, and that this was to a large extent due to identifiable reduction effects.

It should be noted, however, that the present results are based on a very small number of speakers. Individual speaking styles are known to vary in many ways and over a wide range. More speakers would need to be analyzed before the results can be considered conclusive. It is also obvious that the 'Ohrenphonetik' employed here is bound to be quite subjective. Even though spectrographic evidence was used to verify the auditory analyses, the nature of the task is mainly perceptual. Consequently, various biases may have played a role. On the one hand, the authors' own expectations could have introduced a 'liberal' bias; i.e., they may have led us in the direction we thought we were heading. On the other hand, a more 'conservative' bias could have come from our own phonological knowledge of Swedish; failing to disregard the formal structure of our material, we may have overlooked some phonetic effects that would have spoken in favor of our hypothesis. These problems could perhaps be avoided to some extent by using non-Swedish listeners who are also blind to the purpose of the experiment.

Of course, the discrepancy between phonological and phonetic representations of speech is nothing new. For example, several studies have demonstrated that quite drastic reduction phenomena may occur in many languages. It is particularly noteworthy that these studies have shown that reductions appear to occur in a principled way rather than being just sloppy versions of more elaborated speaking styles (see, e.g. Engstrand & Krull 1988; Engstrand 1992; Kohler 1990, 2000; Lindblom 1987, 1990). In addition, the present study has demonstrated that phonetic reduction also appears to involve considerable simplifications such that alternating contoid and vocoid articulations relate to more complex consonant and vowel structures at the phonological level. Finally, the segmentation criteria used in the present study were partly based on the assumption that articulatory activity underlying casual speech is organized in terms of typologically basic CV units. This, however, is the linguist's intuition. We do believe that is is also a psychological reality to language users, but this is still an open question.

References

Browman, C. & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica* **49**, 155–180.

Engstrand, O. (1992). Systematicity of phonetic variation in natural discourse. *Speech Communication* **11**, 337–346.

Engstrand, O. & Krull, D. (1988). On the systematicity of phonetic variation in spontaneous

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press

speech. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* (*PERILUS*) **8**, 34–47.

Gårding, E. (1967). *Internal Juncture in Swedish*. Lund: Gleerup.

Hardcastle, W. & Marchal, A. (eds.) (1990). *Speech Production and Speech Modelling*. Dordrecht: Kluwer.

Jespersen, O. (1926). *Lehrbuch der Phonetik*, 4. Aufl., 190–191. Leipzig & Berlin: Teubner.

Kohler, K. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In Hardcastle & Marchal (eds.), 69–92.

Kohler, K. (2000). Investigating unscripted speech: implications for phonetics and phonology. In Diehl, R., Engstrand, O., Kingston, J. & Kohler, K. (eds.), *Emergence and Adaptation. Studies in Speech Communication and Language Development*, 85–94. Basel: Karger.

Lindblom, B. (1987). Adaptive variability and absolute constancy in speech signals: two themes in the quest for phonetic invariance. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallinn 1987, vol. 3, 9–18.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H and H theory. In Hardcastle & Marchal (eds.), 403–439.

Ruhlen, M. (1987). *A Guide to the World's Languages*. London: Edward Arnold.

Sigurd, B. (1965). *Phonotactic Structures in Swedish*. Lund: Uniskol.

https://doi.org/10.1017/S0025100301001049 Published online by Cambridge University Press