# DISCONTINUOUS VARIATION IN SPONTANEOUS SPEECH[1]

Olle Engstrand & Diana Krull
Institute of Linguistics, Stockholm

## 1 Introduction

In spontaneous speech, words and phrases are frequently heard to be strongly reduced in relation to their alleged phonological base forms. In a subset of the vocabulary[2], we observe an apparent quantal alternation between word and phrase representations in the sense that the amplitude of a phonetic gesture corresponding to an underlying segment displays no continuous transition between levels of reduction or elaboration. As a rule, moreover, it seems as if speakers tend to adhere to a very restricted number of such discontinuous realizations. Fig.1 provides an example. The figure displays spectrogram tracings of the first and second formants pertaining to all occurrences of the grammatical conjunction så att /so att/ (so that, such that) as pronounced in a spontaneous speech situation by a male Stockholmer (see Sec. 2.1 for details). There are two basic, quite distinct sets of curves pertaining to one reduced (left three columns) and one elaborated (rightmost column) level of pronunciation. The reduced version sounds very much like the word satt /satt/ (sat).

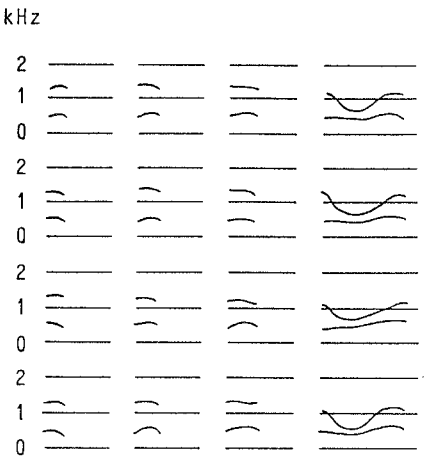


**Fig. 1.** Tracings from broad band specrograms illustrating two basic varieties of the pronunciation of the grammatical conjunction så att. The left three columns represent a reduced level of pronunciation and the rightmost column represents an elaborated level. Subj. JS. (See section 2 for details.)

Do strongly reduced forms such as these instances of så att as a rule contain any kind of acoustic traces of the segments constituting the corresponding phonological base forms? Let us refer to the affirmative answer to this question as the weak hypothesis of segmental invariance[3].

Mere listening and transcribing is probably an insufficient method of investigating this hypothesis. For example, even when a clear segment realization is heard in the natural speech situation, the effect frequently seems illusory when listened for in an excized portion of the speech flow. To test the hypothesis, therefore, careful measurements are necessary. The remainder of this paper discusses a set of such measurements and their subsequent evaluation in terms of listener reactions.

## 2 Experiments

### 2.1 Recording and formant measurements

We used for this experiment approximately half an hour of recorded speech produced by a male native speaker of the Stockholm dialect of Swedish (subj. JS). The greater part of the recording is a lively monologue supported by brief questions and comments from one of the experimenters[4].

Out of the material, we selected and spectrographed (1) all instances of the grammatical conjunction så att (so that), where å stands for the phonologically tense back rounded vowel /o/, and a stands for the phonologically lax open fronted vowel /a/; (2) all sequences saC, where /C/ is any non-nasal consonant; and (3) all preconsonantal instances of the word så (so).

Onset and nuclear (mid vowel) frequencies of the first four formants ($F_1$–$F_4$) were identified and measured for all occurrences of these utterances. $F_3$ and $F_4$ were practically identical and will not be shown here. Measurement data for $F_1$ and $F_2$ are given in Table 1.

**Table 1.** Onset and nuclear frequencies for the vocalic segments in så att, saC and så. Subj. JS.

| NO. | PHRASE | ON/NUC | FORMANT | N | MEAN | S.D. |
|---|---|---|---|---|---|---|
| 1 | så att | on | F1 | 12 | 470 | 49 |
| 2 | så att | on | F2 | 12 | 1332 | 57 |
| 3 | så att | nuc | F1 | 12 | 553 | 42 |
| 4 | så att | nuc | F2 | 12 | 1319 | 62 |
| 5 | saC | on | F1 | 16 | 516 | 78 |
| 6 | saC | on | F2 | 16 | 1348 | 70 |
| 7 | saC | nuc | F1 | 13 | 596 | 85 |
| 8 | saC | nuc | F2 | 13 | 1325 | 83 |
| 9 | så | on | F1 | 10 | 420 | 32 |
| 10 | så | on | F2 | 10 | 1395 | 102 |
| 11 | så | nuc | F1 | 9 | 408 | 40 |
| 12 | så | nuc | F2 | 9 | 1373 | 101 |

The table shows slight differences between the three cases. In particular, a t-test reveals statistical significance at the 1% level for mean $F_1$ onset values both for så vs. så att and for så att vs. saC.

We next turned to the question whether these small differences are at all able to influence listeners' perception in the direction of the respective utterances så att and satt with these portions of the speech flow excized out of their contexts.

## 2.2 A listening test

A series of stimuli containing 12 instances of så att and 12 instances of satt was presented to nine native Swedish speakers of roughly the same dialect as subj. JS. Since the sequence satt occurred only four times in our recorded material, copies of those tokens were introduced randomly to give an equal number of both stimulus types. The resulting set of stimuli was subsequently duplicated and a new random order was created for the duplicate. This procedures gave a total of 432 responses from the nine listeners whose task it was to decide for each stimulus whether it represented så att or satt.

The results are shown in Table 2 in terms of the outcome of a multiple regression analysis using 1-5 predictor variables. The data refer to correctly identified instances of så att. The values of R for the correlation between number of correct answers and each individual variable is given in the middle column, and the cumulative effect is shown in the right column. Calculated across all predictor variables, this effect is R=0.89.

**Table 2.** Result of multiple regression analysis: prediction of number of correctly identified instances of the phrase så att using 1-5 predictor variables. 9 subjects.

| PREDICTOR | MULTIPLE R | |
|---|---|---|
| | Individual variables | Cumulative |
| F1on | -0.60 | -0.60 |
| F2on | -0.60 | 0.77 |
| F1nuc | -0.35 | 0.77 |
| F2nuc | 0.11 | 0.85 |
| Duration | 0.51 | 0.89 |

We thus obtain a high cumulative correlation with contributions from several variables (including nuclear duration which was slightly greater for satt than for så att).

## 3 Comments

The above relationships, of course, are purely statistical, and do not tell us much about the perceptual contribution of each variable. However, $F_1$ onset is the acoustically strongest effect and may therefore play a decisive role. The contribution of a possible spectral difference in the preceding /s/ should also be evaluated. As demonstrated at the Conference, however, the measured effects are likely to be very marginal as a perceptual criterion in natural speech. It is, however, interesting in certain other respects.

Firstly, the example suggests that this speaker in fact tends to preserve slight traces of underlying segment structure even in very strongly reduced forms.

Secondly, since the speaker is very close to completely eliminating a segment, for what reason does he nevertheless preserve it to some extent?

These points may be discussed in terms of various constraints on phonetic speaker-listener interaction. Thus, on the one hand it can be assumed that speakers, in accordance with Lindblom's view of adaptive variability (Lindblom 1987), tend to shape the speech wave so as to achieve a good balance between articulatory effort and perceptual clarity. If this balance were optimal, however, the vowel corresponding to /o/ in så att would be reduced completely since it is highly predictable in context and therefore semantically insignificant to the listener.

Reasonably, however, an optimal adaptation would in itself require a fair amount of data processing. Thus, rather than designing a particular utterance plan to cope with predictable segments, it would seem to be a more convenient speaker strategy to manipulate the basic motor scheme by means of a somewhat more coarse-grained adaptation of gestural amplitude. In particular, the speaker would simplify his task by aiming for either a low (hypo) or a high (hyper) level variety with phonetic traces of even perceptually negligible underlying segments preserved at both levels.

As a third possibility, it may be hypothesized that complete segment reduction is marked stylistically, so that traces of underlying segments will remain and be perceived even though their strictly semantic value is insignificant. The above experiment demonstrates that listeners are, in fact, sensitive to this kind of variation. The difficulty of dealing with it from a semantic-perceptual point of view suggests the need for adding the phonostylistic dimension to the issue of segmental invariance in spontaneous speech.

## Footnotes

1) Supported by the Swedish National Board for Technical Development (STUF) and The Bank of Sweden Tercentenary Foundation.

2) Most typically in grammatically functional morphemes, words and phrases.

3) The strong version would involve the claim that some acoustic aspect of a segment appears unaltered across a variety of contextual transformations.

4) This recording session was conducted by Rolf Lindgren of the Institute of Linguistics.

## Reference

Lindblom, B. 1987. Adaptive variability and absolute constancy in speech signals: two themes in the quest for phonetic invariance. Proc. XIth ICPhS, Tallinn, Estonia, Aug. 1987, Vol. 3, pp. 9-18.