Hesitation disfluencies in Swedish: prosodic and segmental correlates

Merle Horne, Johan Frid, Birgitta Lastow, Gösta Bruce, and Adina Svensson

Dept. of Linguistics, Lund University, Sweden

E-mail: merle.horne@ling.lu.se, johan.frid@ling.lu.se, birgitta.lastow@ling.lu.se, gosta.bruce@ling.lu.se

ABSTRACT

The conjunctions *att* 'that' and *och* 'and' in Swedish are examined with the goal of finding prosodic and segmental cues distinguishing between their occurrence in fluent and hesitation disfluency contexts (when followed by a pause). Analysis of data indicates the following major tendencies: 1. the mean F0 in the function words does not differ significantly in hesitation and fluent contexts whereas segmental characteristics do indeed differ, 2. Vowel duration in *att* and *och* is longer before hesitations, 3. The duration of the final aspiration phase in *att* and *och* is longer before hesitations, 4. A relationship of complementary distribution is observed to exist between filled pauses and final aspiration, perhaps signalling different kinds of on-going activity in the planning of the following speech fragment. Aspiration phases over 80 ms tend not to be followed by filled pauses.

1. INTRODUCTION

A central issue in speech technology research on recognition and understanding of spoken language is the development of methods for identifying relevant processing units in the stream of speech. Boundaries corresponding to punctuation marks (periods, commas, etc) do not always have clearly specifiable correlates in spoken langauge and thus one fundamental problem that has to be solved is: how do different kinds of phonetic, lexical and syntactic form interact in signalling the boundaries of relevant processing units in spoken language?

Another factor making the processing of spontaneous speech a challenge is the fact that speakers do not always produce complete sentences or clauses. The fact that speakers for example sometimes pause to plan up-coming speech or to access a word or phrase in their mental lexicon has made the study of different kinds of speech 'disfluencies' an important topic for linguists and speech technologists working in the area of speech recognition ([5],[7],[10],[11]). Not only speech technologists, but also psycholinguists working on human language processing have been attracted to the study of spontaneous speech in their quest for a better understanding of how humans produce and comprehend speech ([2],[4],[6],[9]). Spontaneous speech constitutes an excellent material for studying spoken language production and the development of models of language production has much to gain from the study of this mode of language. In particular, the issue of how a conceptualized message is linguistically encoded in different communicative contexts is a central topic for psycholinguists.

2. FUNCTION WORDS

In the pilot project our group has been involved in during the past year, we have investigated the interaction of prosodic and segmental characteristics associated with function words occurring at the boundary of one kind of disfluency, hesitations. Function words such as conjunctions, prepositions and pronouns can be thought of as important cues in parsing syntactic structure since they occur at the boundaries of phrases/clauses. In written-language-based grammars of Swedish, they are described as occurring at the left edge of syntactic constituents. In spontaneous speech production, however, they are observed most often to occur at the right edge of speech fragments bounded by pauses. In (1-2), taken from our data from the *SweDia 2000* project, one can see examples of these 'stranded' function words occurring before hesitation pauses:

(1) och EH || även lite gympa på || vinterhalvåret när EH || fotbollen är slut
'and EH || even a little exercise in || the winter when EH || soccer is over'

(2) att || just å jobba det man || trivs med å göra
'that || just to work (with) what one || likes to do'

where || represents a hesitation disfluency boundary.

According to Clark & Wasow's 'Commit and Restore' model of speech production [4], stranded function words signal that the speaker intends to produce a constituent of the kind signalled by the kind of function word produced, e.g. a clause after a stranded conjunction, a prepositional phrase after a preposition, etc. Thus the recognition of stranded function words (conjunctions, prepositions, pronouns) can be expected to be important for automatic parsing algorithms. Further, according to

ISBN 1-876346-48-5 © 2003 UAB

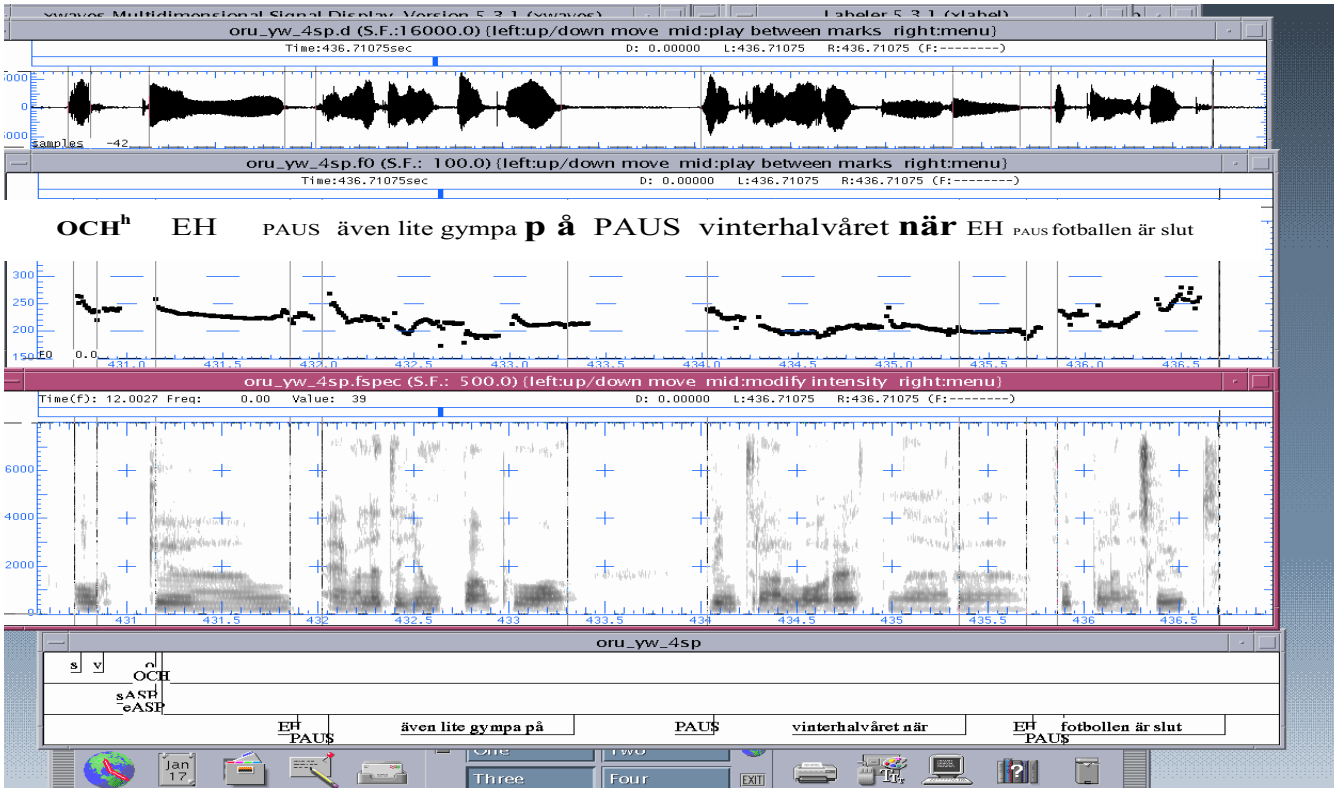

Figure 1. Waves+-screendump for utterance (1) showing from top to bottom: waveform, F0-contour showing associated word groupings, spectrogram, and label-tiers. Notice the 'stranded' function words *och*, *på* and *när*.

Clark & Wasow's 'complexity hypothesis', the probability that a speaker will hesitate in speech production will increase, the more complex the constituent being planned is (where complexity is measurable in terms of a number of lexico-grammatical parameters, e.g. number of (content) words, number of phrasal nodes.

3. HYPOTHESES

An initial hypothesis was that the average F0-level in fluent versions of *att* and *och* would perhaps be different from F0-levels in *att* and *och* in disfluent contexts, in particular if it were the case that the *att*'s followed by a hesitation disfluency were part of a separate planning unit, unlike the fluent *att*'s. Secondly, we expected that *att* and *och* in disfluency environments would be characterized by segmental characteristics that distinguish them from their phonetic realization in fluent contexts. Following the reasoning in Clark & Wasow's 'Commit and Restore' model, one could hypothesize that, since Swedish conjunctions such as *att* 'that' and *och* 'and' occur before major constituents, i.e. clauses, one would expect that their phonetic form before a hesitation would probably differ from that in fluent speech. We decided therefore to investigate prosodic and segmental properties of these function words in hesitation disfluency contexts, i.e. before a pause, and compare them to their properties in fluent speech. Studies like those of Jurafsky et al. [8] and Wesener [12] have investigated reduction of function words in English and German spontaneous speech, respectively. The former study found, e.g. that disfluency contexts cause less reduction, that words are more likely to be reduced the more predictable they are given the two preceding words, and that function words are less reduced in utterance-initial and utterance-final positions. Thus, we also expected that the Swedish function words *att* and *och* in phrase internal position in fluent speech would be more reduced than *att* and *och* at the end of a prosodic constituent before a hesitation pause. In particular, we expected that *att* and *och* in disfluency contexts would be marked by longer duration as well as final aspiration. Finally, we also expected that final strongly aspirated stops before hesitations and filled pauses in hesitations would be characterized by a relationship of complementary distribution. Aspiration and filled pauses can both be expected to reflect on-going activity in speech planning but one can hypothesize that this activity is of a different nature in the two cases. In the case of a filled pause *(EH)* following *att* or *och*, for example, one can hypothesize more on-going cognitive planning

ISBN 1-876346-48-5 © 2003 UAB

activity since the vocal cords are activated, i.e. ready for articulation of the speech being planned.

4. SPEECH MATERIAL

The analyses reported on here have been carried out on spontaneous speech from the *SweDia 2000* project (Bank of Sweden). Speech from 24 speakers has been used from 17 locations in the area of Götaland in southern Sweden. Of these 24 speakers, 14 are women and 10 are men. The speakers are divided into two groups depending on age – younger and older. The material includes 8 younger women, 10 younger men, 2 older women and 4 older men. Decisive for the choice of these speakers has been the frequency of occurrence of the words *att* and *och* in the speech material. In total, 482 cases of *att* and 440 cases of *och* in disfluency environments were labeled and compared with 797 cases of *att* and 204 cases of *och* in fluent environments.

4.1. LABELLING OF SPEECH MATERIAL

The speech material was tagged at hesitation disfluencies involving *att* and *och* as well as in fluent contexts containing these function words. Table 1 shows which tags were used as well as their description:

Label	Description
s	Indicates where the word *att* or *och* starts.
v	Indicates where the vowel in *att* or *och* ends.
o	Indicates end of stop occlusion.
att	Indicates the end of the word *att* in fluent speech.
ATT	Indicates the end of the word *att* in hesitation disfluency contexts.
och	Indicates the end of the word *och* in fluent speech.
OCH	Indicates the end of the word *och* in hesitation disfluencies.
sASP	Indicates the beginning of an aspiration phase.
eASP	Indicates the end of an aspiration phase.
PAUS	Indicates the end of a silent pause.
EH	Indicates the end of a filled pause.
s	Indicates the beginning of an irrelevant stretch of speech
e	Indicates the end of an irrelevant stretch of speech

Table 1: Labels used in tagging the function word data.

5. ANALYSIS OF DATA

5.1. COMPUTATIONAL TOOLS

Analysis of the data has been done using a number of short programs (scripts) developed by B. Lastow and J. Frid. The scripts are in the form of Bourne Shell-scripts or Tcl-scripts and have been executed in a Unix-environment.

5.2. F0 MEASUREMENTS IN *ATT* AND *OCH*

After calculating max- and min-values, averages and standard deviation for F0-values for each vowel, it became clear from examining the results that many of the vowels were potentially glottalized, i.e. they had an unnaturally low F0 (ca 50-60 Hz for a man) together with a high standard deviation. These glottalizaed vowels were consequently deleted from the analysis after an individual examination of the suspected cases. Since the F0-values at the beginning of the vowels is often very instable, we also decided to eliminate the first two measurement points in the calculation of average F0-level for a given vowel. Results for the F0 measurements for the different speakers are given in Table 2.

Person	Fluent att	Disfluent ATT	Fluent och	Disfluent OCH
ars_om_1sp	115	112	112	114
asb_ow_2sp	212	219	210	209
bre_om_1sp	108	108	111	112
bre_ym_3sp	111	110	111	109
flo_yw_1sp	200	205	234	184
fri_om_2sp	113	113	114	106
fri_ym_2sp	123	128	-	126
fri_yw_2sp	201	189	201	203
ham_ym_2sp	95	93	95	97
jam_ym_3sp	106	103	127	112
jam_yw_2sp	217	214	239	216
jar_om_1sp	125	116	113	115
kaa_ym_2sp	100	100	-	115
kor_ym_1sp	102	99	105	105
oru_yw_4sp	216	204	209	216
ost_ym_3sp	95	91	94	107
ost_yw_1sp	208	202	227	215
oxa_ym_3sp	139	130	130	129
oxa_yw_1sp	234	263	272	279
ste_yw_3sp	209	202	217	212
toa_ym_3sp	105	104	119	110
toh_ym_3sp	117	122	141	124
too_ow_1sp	196	194	209	205
too_yw_2sp	196	199	228	219

Table 2: Average F0 levels in *att* and *och* for all speakers.

An analysis of significance with a nonparametric test (Wilcoxon) showed that there was no significant differences in F0-level between *att* and *och* in hesitation contexts and fluent contexts. This finding is interesting since it seems to indicate that an upcoming hesitation after a conjunction has no effect on speakers' fundamental frequency level in the speech fragment being produced up to the point of hesitation. This differs from speakers' production of the segmental form of the function word, which, as will be shown in the following section, differs in the two contexts.

5.3. VOWEL DURATION, WORD DURATION, AND DURATION OF FINAL ASPIRATION

Measurements of vowel and word duration were also

ISBN 1-876346-48-5 © 2003 UAB

made of *att* and *och* in disfluent and fluent contexts. Moreover, measurements of final aspiration phases in *att* and *och*, where these occurred, were also made. Results from comparing fluent and disfluent contexts for potential duration diferences can be summarized in the following way (all differences are significant at the 95%-level in a z-test):

- Duration of *att* and *och* is longer before hesitations (about 130 ms longer).
- Vowel duration in *att* and *och* is longer before hesitations (about 20 ms longer).
- Duration of the final aspiration phase in *att* and *och* is longer before hesitations (about 20 ms longer for *att*, 10 ms longer for *och*).

5.4. PAUSE STRUCTURE IN HESITATION DISFLUENCIES

Variation in the kind of pause structure following *att* and *och* was also observed, i.e. both silent and filled pauses occur. This was not unexpected. However, more interesting was the observed relation between the duration of the final aspiration phases in *att* and *och*, where these occurred, and the kind of pause following the aspiration. A comparison between hesitation and fluent contexts indicated that:

- Silent pauses following the function words *att* and *och* in hesitations are longer than filled pauses (about 240 ms longer after *att*, 100 ms longer after *och*).
- The duration of the final aspiration phase in *att* and *och* is longer before a silent pause than before a filled pause (about 50 ms longer in *att*, 60 ms longer in *och*).

(The above differences were significant at the 95%-level in a z-test)

- The occurrence as well as the length of a filled pause *(EH)* after *att* and *och* appears to be correlated with the degree of aspiration, i.e. after an aspiration phase longer than 80ms, filled pauses do not tend to occur.

These findings indicate that disfluency contexts do affect the function words *att* and *och*'s segmental form in the following way: The function words are *less reduced* when the speaker is about to produce a hesitation in his/her speech production. It seems indeed as if longer and fuller (segmentally nonreduced) forms are associated with planning strategies, perhaps functioning as a signal or giving the speech production mechanism more time to perform planning strategies.

6. CONCLUSIONS

In summary, the study has provided us with a host of information on the segmental and prosodic form of the function words *att* and *och* in hesitation disfluency contexts. Our findings indicate that the function words *att* and *och* are indeed characterized by relatively marked segmental characteristics before hesitations as compared to their form in fluent contexts. Variations in F0-level in fluent and hesitation disfluency contexts on the other hand are not observed indicating perhaps that the planning of the fundamental frequency contour is indeed suprasegmental/autosegmental and controlled by planning strategies independent of the segmental production.

REFERENCES

[1] G. Bruce, O. Engstrand and A. Eriksson, "De svenska dialekternas fonetik och fonologi år 2000 (SweDia 2000: en projektbeskrivning", Proceedings *6:e Nordiska Dialektologkonferensen*, pp. 33-54, 1998.

[2] H. Clark, "Speaking in time", *Speech Communication* **36**, pp. 5-13, 2002.

[3] H. Clark, and J. Fox Tree, "Using uh and um in spontaneous speaking", *Cognition* 84, 73-111, 2002.

[4] H. Clark and T. Wasow, "Repeating words in spontaneous speech", *Cognitive Psychology* **37**, pp. 201-242, 1998.

[5] R. Eklund, "A comparative analysis of disfluencies in four Swedish travel dialogue corpora", *Proceedings of Disfluency in Spontaneous Speech Workshop*, Berkeley, pp. 3-6, 1999.

[6] J. Fox Tree and H. Clark, "Pronouncing ''the'' as ''thee'' to signal problems in speaking", *Cognition* **62**, pp. 151-67, 1997.

[7] P. Heeman, "Speech repairs, intonational boundaries, and discourse markers: modelling speakers' utterances in spoken dialogue". PhD thesis, Univ. of Rochester, N.Y., 1997.

[8] D. Jurafsky, A. Bell, E. Fosler-Lussier, C. Girand and W.D. Raymond, "Reduction of English function words in Switchboard", *Proc. ICSLP 98*, Sydney, Australia, pp. 3111-3114, 1998.

[9] W.J.M. Levelt, *Speaking: From attention to articulation.* Cambridge, Mass.: MIT Press, 1989.

[10] J. Nordling, "Reparationer i spontant tal", B.A. paper, Dept. of Ling., Lund Univ., 1998.

[11] E. Shriberg, *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, Univ. of Berkeley, 1994.

[12] T. Wesener, "Production strategies in German spontaneous speech: definite and indefinite articles", *Proc. ICPhS* 99, San Francisco, USA, pp. 687-690, 1999.

ACKNOWLEDGEMENTS

This research has been supported by grant 2001-06309 from the *VINNOVA* (*Verket för Innovationssystem* 'The Swedish Agency for Innovation Systems') Language Technology Program.