Individual variation in F0 marking of turn-taking in natural conversation in German and Swedish

Martina Rossi, Kathrin Feindt, Margaret Zellers
ISFAS, Kiel University, Germany
mrossi@isfas.uni-kiel.de, kfeindt@isfas.uni-kiel.de, mzellers@isfas.uni-kiel.de

Abstract
The linguistic mechanisms organizing turn-taking in conversation are still not fully understood. Especially disputed is the relevance of various linguistic features to signal the disposition to yield the floor. The present study adds to this discussion by examining the role of prosody for turn-taking in two different languages, German and Swedish. F0 movement is measured at three points—offset (P1), 200ms (P2) and 500ms (P3)—before the turn end, and normalized. Sentence type (declarative, question), type of speaker change (change, keep, backchannel) and transition (gap, no-gap-no-overlap, overlap) were also annotated, among other features. Preliminary results show that German uses a much wider span of F0 values compared to Swedish. Since F0 has a lexical-phonological function (i.e. pitch accent) in Swedish, the potential prosodic structure is restricted. On the other hand, the flexibility of German manifests itself in extreme F0 movements and an accommodation of F0 between interlocutors. Although there is evidence for accommodation of F0 in German, this is not as strongly demonstrated in the Swedish data. As our F0 normalization should exclude a physiological explanation, we argue for an explanation based on entrainment.

1. Introduction
In conversation, speakers produce acoustic-prosodic cues, along with morphosyntactic and gestural ones, to signal to listeners their intention to either cede the floor or continue speaking. In order for the listener to launch their turn in time or to remain silent, and thus to comply with the conversational expectations of keeping silent gaps and overlapped speech portions reduced to a minimum [1, 2], the variation of phonetic cues related to turn-taking is hypothesized to be initiated early within the current turn, from half a second [3] up to one full second before the end of the turn [4]. The precise location of the transition space, however, is yet to be defined, as well as the relative contribution to turn transition of several proposed phonetic cues across languages. The current study adds to the ongoing discussion about the turn-taking mechanisms by observing the variation of a phonetic parameter which has been assigned a central role by researchers in signaling speakers’ conversational intentions, i.e. F0, towards potential turn boundaries (PTBs) [4].

To test the extent of F0 variation as a turn-taking cue, we adopted a cross-linguistic perspective and analyzed fundamental frequency contours in German and Swedish. The comparison of these two languages is motivated by the fact that, even though related, they differ significantly in their prosodic phonology. German is an intonation language in which pitch accents mark prominent syllables and boundary tones mark the end of an intonational phrase, whereas Swedish is a pitch-accent language [5] in which a two-way lexical pitch accent contrast distinguishes word pairs and focus is signaled with an additional high tone [6, 5]. Moreover, boundary tones in Swedish rarely show a rising pitch contour; even questions end mostly with a falling contour [7, 8].

The objectives of this study are thus to investigate the differences in F0 marking of PTBs before both speaker change and floor holds and, furthermore, to observe how speakers of German and Swedish use F0 variation to signal their intentions for the following turn in the ongoing conversation.

2. Methodology
2.1. Dataset
The production data analyzed for this study consists of two-party spontaneous conversations in German and Swedish. The German data comes from the Lindenstraße task in the Kiel Corpus of Spoken German [9], while the Swedish conversations come from the Spontal Corpus [10]. Subjects from the two corpora are native speakers of German and Swedish respectively. In both corpora, speakers were recorded in separate channels, allowing for a phonetic analysis of speech even when it was produced in overlap. Thus far we have annotated and analyzed a total of 4 conversations involving 8 different speakers. Each subject has been given an alphanumerical identification code (e.g. SG01A or SS02B) indicating their L1 being either German (SG) or Swedish (SS) and the conversation in which they took part (01 or 02, with two conversations per language). The letter “A” or “B” at the end of the code identify the two participants in the same conversation.

The current sample was not balanced for the speakers’ gender: all German speakers are female while 3 Swedish speakers are male and 1 is female. This issue and its possible consequences on our results and their interpretation will be addressed in Section 4; however, to exclude the variation related to physiological factors, F0 values have been normalized with the procedure reported in 2.3. The dataset analyzed for the present study is comprised of 521 potential turn boundaries in declarative form. All annotations were carried out using Praat [11].

2.2. PTBs annotation
PTBs are defined as locations in conversation where it is possible, although not obligatory, for one speaker’s turn to end; they are roughly comparable to transition relevance places [1] or SYNCOMPS [12]. For the current study, PTBs were identified following the automatic detection of silent pauses in the individual audio tracks. They were later manually annotated for syntactic/pragmatic completeness, sentence type (declarative, question or tag question; however, the current study focuses only on declarative utterances), sequential structure and transition type.

On the basis of what followed the PTB, one of the following
sequential structure labels was assigned:

- “c” for change: the other conversational participant takes
  up the next full turn;
- “k” for keep: the current speaker holds the floor by con-
  tinuing his/her turn;
- “b” for backchannel: the other conversational participant
  produces a minimal response and the first speaker con-
  tinues his/her turn.

The transition type label was applied to describe how the
transition between the above-mentioned sequential structures
occurred:

- “g” for gap: between the turns there is a marked silent
  pause, i.e. longer than 120 ms [13];
- “o” for overlap: between the turn there is a marked por-
  tion of overlapped speech in which both speaker talk si-
  multaneously for more than 120 ms [13];
- “n” for no-gap-no-overlap: the transition between turns
  occurs smoothly, with potential gaps or overlaps shorter
  than 120 ms.

2.3. F0 points

To observe the F0 movements leading up to the turn bound-
ary and to investigate the extension of the transition space, F0
values were automatically extracted at three test locations:
P1, i.e. at the offset of speech; P2, i.e. at 200 ms preceding
the boundary; P3, i.e. at 500 ms preceding the boundary. The
extraction was carried out using Praat’s setting for semitones
above 1 Hz. The values obtained were then normalized for each
speaker by calculating a baseline F0 value for the speaker and
subtracting it from the measured values in the data. The base-
line F0 was calculated using a similar method to [14], using
an automatic extraction of values and assigning the value at the
first percentile as the baseline F0 value for the speaker. Thus all
reported F0 values are measured in semitones (st) relative to
the speaker’s baseline F0.

3. Analysis and Results

Results from an exploratory analysis of the data give evidence
for a different use of F0 between the two groups of speakers,
with a few cross-linguistic similarities. The first main differ-
ence between the two languages is the overall F0 span of data
points at the test locations. Where German speakers’ F0 range
of values at turn ends extends from 3 st to around 20 st, Swedish
F0 span of data points at turn ends remains overall closer to the
speakers’ baseline, going from 2 st to 8 st: the range of F0 val-
ues observed for the Swedish speakers appears to be much nar-
rower then that used by the German speakers. In speaker change
cases, the variation of the P1 (F0 values at the boundary) shows
that German speakers end their utterances at around 10 st above
their baseline, while they end higher for keeps, i.e. at around 14
st. On the other hand, Swedish speakers end higher for speaker
change, at around 5 st, and lower for keep, at around 3 st.

A recurrent F0 contour pattern for changes and keeps, how-
ext, did not clearly emerge from our dataset, due to the very
high degree of inter-speaker variability observed for German
subjects and, to a lesser extent, for Swedish ones (see Fig. 1).
For example, in the German data, where subject SG01A ends
at 15 st above the baseline, subject SG02B ends much lower at
2.5 st. These are, however, speakers from two different conver-
sations. If we observe German speakers interacting with each
other in the same conversation, the two pairs appear to behave in
a similar way, seemingly accommodating their F0 movements
to each other. The F0 values at the three test locations average
at 15 st in conversation 01, while they are closer to the speakers’
baselines in conversation 02, at 5 st, with some variability. Lo-
cal similarity for the German speaker pairs appears in the data
for keeps as well (see Fig. 2). Speaker SG02A displays a high
degree of variability for the two F0 points (P3, P2) preceding
the turn final one (P1), which falls however to the same average
of the other conversational participants, i.e. 5 st above the base-
line. In conversation 01, both speakers behave once again very
similarly and maintain the F0 at 13 st above their baseline.

On the other hand, in the Swedish data there is no evidence
for possible F0 accommodation between speaker pairs. For in-
stance, in speaker change cases (Fig. 1), participants of conver-
sation 01 fall to 2.5 st and 5 st respectively: in conversation
02, an almost divergent behavior can be observed, where one
speaker rises to 8 st while the other falls very close to his base-
line.

In spite of the high degree of variability in the data, we were
able to make some general observations about the two groups’
turn-taking behavior. For both languages, in speaker change
cases followed by a smooth transition, the F0 values showed
a wide span of variability. F0 in speaker changes followed by
a gap, instead, shows a more consistent patterning towards the
boundary, with the German P1 averaging at around 13 st and
the Swedish P1 at around 2.5 st (higher than the other aver-
eges for German, lower than the others for Swedish). The fact
that such speaker change cases are followed by a gap offers a
possible explanation for the situation, i.e. the next speaker in-
terpreted the final rise or the final fall, respectively for German
and for Swedish, as a signal that the current speaker wanted to
continue talking and take up the following turn, too. These val-
ues, in fact, resembles those from the keep condition, with the
German speakers’ F0 at around 14 st above the baseline and
the Swedish at 3 st. Thus, ending higher would be a signal for
German speakers to hold the floor, while for Swedish the same
intention would be signaled by a lower F0.

Sentence completeness gives more insight into the high F0
variability for speaker change cases. In the no-gap-no-overlap
transitions, as well as for overlaps, we observed a high degree of
F0 variability in sentences labeled as complete: it appears that,
when the syntax or pragmatics signal the completeness (or the
forthcoming completeness) of the sentence, then F0 variation is
more free and less restricted.

4. Discussion

We analyzed the F0 marking of turn-taking in two German
speaker pairs and two Swedish speaker pairs (four different
speakers for each language) over the last 500 ms of speech in
conversation. What was immediately apparent from our data
was the high degree of inter-speaker variability in the F0 move-
ments over our three test locations. In the German data, how-
ever, we noticed that the four subjects in the two conversations
analyzed showed a convergent behavior in F0 variation in both
change and keep cases, i.e. the speakers in the same conver-
sations displayed a very similar range of values of F0 at the
three locations analyzed. We hypothesize that this could po-
tentially be evidence for a possible accommodating behavior
between German conversational participants, who may entrain
with each other in F0 and its variation when signalling different
Figure 1: Boxplots of normalized F0 for German speakers (SG01A, SG01B, SG02A, SG02B) and Swedish speakers (SS01A, SS01B, SS02A, SS02B) preceding a speaker change. Numbers "01" and "02" in the German and Swedish participants' identification codes indicate the conversation in which they took part.

Entrainment of acoustic-prosodic parameters between conversational participants was observed by [15], who reported local F0 accommodation at the turn level, as well as [16], with F0 being one of the two most significant features that speakers in collaborative dialogues entrained when there was rapport between participants. At the current stage of this study, our data are still limited to a relatively small set of speakers and the results obtained are descriptive. In fact, our observation is based on an seemingly convergent qualitative distribution of F0 values in the conversations of two speaker pairs in each language, and convergence or proximity measures have not been taken into consideration yet. However, the fact that our data are normalized to the speakers’ baseline allows us to exclude the possibility of a similarity based on physiological factors, so we hypothesize the presence of a local entrainment in F0 between the subjects.

Moreover, as we have already mentioned in 2.1, our corpus is not balanced for subjects’ gender: speakers in the German group are all females, while all except one speaker (i.e. SS02B) in the Swedish group are male. It could be argued that this gender difference could be the explanation for the observed convergent behavior in the German subset, since female speakers have been observed to accommodate more frequently in conversation [17, 18]. However, studies on prosodic entrainment have reported irregular results on the matter and, after their systematic analysis of two large corpora, [19] claimed that gender does not have a significant effect on the accommodation of acoustic-prosodic features, including F0. It is thus still possible to attribute the different behavior observed between the two groups to cross-linguistic differences in the two languages’ intonational phonology, particularly since the differences we find are consistent with what could be predicted on the basis of the phonological systems. As discussed in Section 1, Central Swedish has a two-way lexical pitch accent system in which pitch distinguishes segmentally-identical word pairs, and focused content words and phrase-final words are marked with an additional high tone [6, 5].

Our results are in line with [20, 21, 4]’s hypothesis that Swedish pitch could be already saturated as a signaling tool, thus speakers tend to rely more consistently on other phonetic features, such as segmental duration. On the other hand, German speakers are able to vary pitch more without impact on the lexical content, which leads to them having more flexibility and thus the space to entrain with each others’ F0 in conversation. This does not exclude the possibility of entrainment in Swedish: conversational participants may accommodate to each other in other phonetic features, such as e.g. segmental duration, intensity or voice quality. Further research will investigate this hypothesis, particularly focusing on clusters or constellations of phonetic features.

The observation that Swedish F0 may be already saturated as a signaling cue could also explain the narrower F0 range observed for Swedish in comparison to German: Swedish boundary tones are more constrained by the lexical-phonological function of the pitch, and thus F0 variation range at the boundary will be smaller when compared to that of an intonational language, such as German, although variations in boundary tone...
5. Conclusion

We analyzed F0 variation at three time points towards turn boundaries in spontaneous conversation in German and Swedish. These are two related languages with different intonational phonologies, which appear to lead to cross-linguistic differences in turn-taking signaling. Individual variation was particularly evident in our sample, and it reflects to some extent the way in which German and Swedish deal with F0 as a turn-taking cue. In fact, German speakers display a higher degree of variability, along with entrainment tendencies among speaker pairs, while, on the other hand, we find that Swedish speakers did not show any convergent behavior for F0, and utilize a smaller F0 range when approaching the boundary. The fact that Swedish F0 appears to be less flexible at PTBs may suggest that it is less available to speakers as an intention signalling cue in conversation, while its wider variability range in German, as well as its use as a feature for entrainment, may indicate a greater contribution to the turn-taking mechanism regulating interactions. Further research on these two languages, including a larger sample of speakers, will investigate further how F0, along with other phonetic features, varies to signal turn-taking intentions, both cross-linguistically and between individual speakers.

6. Acknowledgements

This research was supported by the project Sound Patterns and Linguistic Structures at the Transition Space in Conversation (ZE 1178/3-1), funded by the German Research Council (DFG). Many thanks to our student assistants Sophie Caliebe and Oke Matthiesen for their work on the annotations, and to the reviewers for their constructive comments.

7. References

[4] M. Zellers, “Prosodic variation and segmental reduction and their...


