F0 DECLINATION AND SPEECH PLANNING IN FACE TO FACE DIALOGUES

Susanne Fuchs[1], Uwe D. Reichel[2] and Amélie Rochet-Capellan[3]

Zentrum für Allgemeine Sprachwissenschaft, Berlin, Hungarian Academy of Science, Budapest, GIPSA-lab, DCP & CNRS Grenoble
Email: fuchs@zas.gwz-berlin.de

Abstract: In general, f0 decreases over the course of an utterance. This phenomenon is known as f0 declination and has been observed in a variety of languages. F0 declination is yet influenced by different factors such as the type of speech (e.g. read vs. spontaneous speech), the length of the utterance, and it cannot only be explained in terms of physiological mechanisms. This flexibility in f0 declination suggests that it could also depend on cognitive factors and vary in dialogue to signal the organization of turn-taking. In line with previous work, we analyzed f0 behavior in short face-to-face dialogues involving native female speakers of German. Our aim was to analyze the relationship between (1) f0 declination and utterances length as an indicator of anticipation of the upcoming utterance; (2) f0 values and declination according to the main events of dialogue (turn-taking, turn-continuation, turn-ending) to assess whether speaker-listener dyads could use f0 information to negotiate the turns. Our findings suggest that f0 slopes are shallower and less variable for long utterances than shorter ones. Negative f0 slopes are also more often observed than positive f0 slopes, but the direction of change depends on the communicative event. F0 on- and offset values of the regression line defining the f0 slope are related to the direction of f0 changes (positive vs. negative) but not to the communicative event. Altogether, f0 declination is an indicator of the length of the upcoming sentence in dialogues and is to some degree used for the organization of turn taking.

1 Introduction

F0 declination has been described as the gradual decrease of f0 over the course of an utterance [8]. Since it has been reported for a variety of tone and intonation languages (see [5] for an overview), it has sometimes been considered as a universal phenomenon, and therefore several researchers have been looking for physiological origins of the phenomenon [6, 7, 15]. Our previous work [5] based on read and semi-spontaneous speech, evaluated a potential respiratory contribution (decrease of rib cage volume during exhalation) to f0 declination. For this purpose we tested whether f0 declination and respiration (measured as changes in rib cage compression) changed in a similar way according to the number of syllables in an utterance. Additionally, we varied the degree of rib cage compression by means of the number of voiceless obstruents (Steeper rib cage compression occurs due to an open glottis and loss of air in sentences with a higher number of voiceless obstruents) and tested whether a comparable change would be found in f0 declination. The results of these experiments provided evidence that f0 declination and rib cage compression work rather independently. According to Ohala [10], laryngeal mechanisms are also less likely to explain f0 declination, because the laryngeal muscles are not continuously active over the time window of f0 declination (usually an intonational phrase). We searched for an alternative hypothesis and suggested that cognitive and communicative demands may provide further explanations to the fact that f0 declination is frequently found in various languages.

The first aim of this study is to investigate the role of f0 declination in face-to-face dialogues. Specifically, we assessed to what extent f0 declination reflects cognitive processing, i.e. the

anticipation (look ahead mechanism) of the length of the upcoming utterance (measured as the number of syllables). This idea follows up on earlier work [18] showing a consistent relation between the slope of f0 declination and the number of syllables of an utterance. Shallower f0 slopes were found for longer utterances (more syllables) and steeper slopes for shorter utterances (fewer syllables). This may not only apply to negative f0 slopes, but also to positive ones.

This adaptation of f0 declination to the length of the utterance may be relevant for face-to-face interactions and play a communicative role: shallow f0 slopes may signal the listener in a dialogue that the speaker's utterance is long and there is still time for preparing the next turn. In contrast, steep f0 declination could signal shorter speech units and indicate the possibility to take the turn soon. However, an argument against a common use of global f0 declination in dialogue is that it is less frequently found in spontaneous than in read speech (e.g. [17]).

Most of the work so far, has reported that local f0 changes are used for turn taking. In the paper "Why is Mrs. Thatcher interrupted so often?", [1] the main answer to the question was that utterance final lowering in Mrs. Thatcher's speech was often interpreted as a completion of a turn so that her interview partner started to talk. Phrase-final rising pitch may have been a better choice to signal the continuation of a turn [9] and has been reported as a salient trigger for listeners' back channelling [2]. In a recent work, Bögels and Torreira [4] analysed the single versus combined influence of intonational phrase boundaries and linguistic parameters (lexico-syntactic completion) for the interpretation of the speaker's turn end. Their results point to a multivariate approach. The lexico-semantic completion alone cannot predict the ends of speaker's turns, and intonational phrase boundaries are additional cues that listener's use to predict the end of a turn.

The second aim of this study is to investigate f0 declination and local initial as well as final f0 values in inter-pausal utterances. We predict that: A) Negative f0 slopes should be more frequently found in turn-taking or turn-ending utterances, while positive f0 slopes should rather characterize turn continuation. B) Initial f0 should be higher at turn-taking than at turn-ending or turn continuation. C) Final f0 should be lower at turn-ending than at turn-taking or turn continuation.

2 Methods

2.1 Experimental setup

The corpus described in Rochet-Capellan & Fuchs [14] was analysed for the current purposes. It consists of 110 spontaneous face-to-face dialogues of approximately 2.5 min length each. In these dialogues 11 female speakers (S1-S11, mean age= 31 years, std = ± 7) talked in five dialogues to one conversational partner and then in five successive dialogues to another conversational partner (P1=42 years old and P2=28 years old). The two partners were females to avoid potential gender confounding. All interlocutors involved in this study were German native speakers with at least a high school educational background. The interlocutors were sitting and facing each other with a distance of about 1.5 m. The experimenter proposed several topics for a conversation like cooking, holiday, sports and movies. The interlocutors chose among these topics to initiate a dialogue. When one dialogue was finished, they could either continue with the same topic or start a new one.
Acoustics were recorded with 11030 Hz sampling frequency using two directional microphones (Sennheiser HKH50 P48). Respiratory kinematics was recorded synchronously, but will not be considered for the purpose of the current study.

2.2 Acoustic labelling and calculation of f0 slopes

We distinguished between speech and silence phases by determining inter-pausal units (IPUs) with silences longer than 50ms. F0 declination in semitones per second was determined for each IPU.

F0 was extracted by autocorrelation (PRAAT 5.3.16, [3] sample rate 100 Hz). Voiceless utterance parts and f0 outliers were bridged by linear interpolation. Outliers were defined as differing from the average f0 by more than two standard deviations. The contour was then smoothed by Savitzky-Golay filtering [16] using third order polynomials in 5 sample windows and transformed to semitones relative to a base value. This base value was set to the f0 median below the 5th percentile of an utterance and serves to normalize f0 with respect to its overall level.

For f0 level stylization for each IPU we employed the midline extraction method proposed in Reichel & Mady [13], which is illustrated in Figure 1 and can be summarized as follows: within each IPU, a window of length 50ms is shifted along the f0 contour with a step size of 10ms. Within each window the f0 median is calculated. Through the resulting median sequence a line is fitted by linear regression, yielding the midline. The following features were then derived for each IPU from this stylization:

- mRate: the midline (level) declination rate (in ST/s)
- mOnset: the initial midline f0 value (in ST)
- mOffset: the final midline f0 value (in ST)

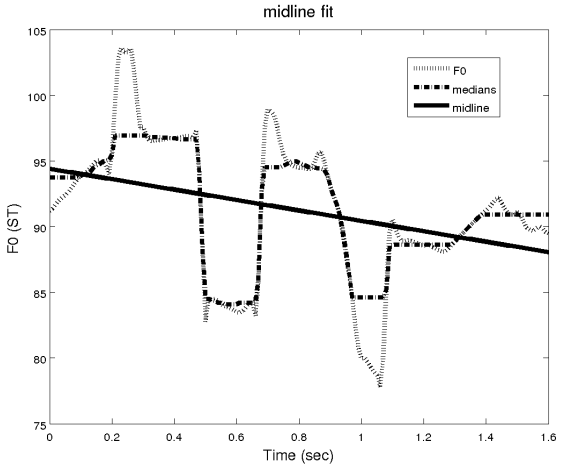

Figure 1: Midline stylization by linear regression through local f0 median values.

The number of syllables in IPUs was inferred from an automatic canonic transcription of the orthographic annotations of the utterance [12].

2.3 Labelling of communicative events

The annotated communicative events were those described in Rochet-Capellan & Fuchs [14]. In brief, IPUs of each interlocutor were separated in initial turns (IPUs at turn taking abbreviated as “InitialOnly”), final turns (IPUs preceding a change in the speaker-listener role, abbreviated as “FinalOnly”), and continuation turns (IPUs between “InitialOnly” and “FinalOnly”, abbreviated as “Cont”. Some turns consisted of single IPUs, these IPUs were both turn initial and turn final, and were abbreviated as “Both”.

3 Results

3.1 F0 declination and cognitive processing

The slope of f0 declination in relation the number of syllables in the upcoming utterance is represented on Figure 2. Negative and positive values of f0 slope are separated to simplify the statistical design and use a linear approach.

Negative as well as positive f0 slopes are shallow in utterances consisting of a large number of syllables, while steeper and more variable in utterances with a smaller number of syllables.

A linear mixed model was run in R (version 3.2.3, [11], library lme4) with f0 slope as the dependent variable and the number of syllables as the fixed factor. For this analysis f0 slopes were log scaled to allow for linearly distributed residuals. The logarithmic scaling was only possible for positive values. Therefore we multiplied negative f0 slopes with -1. The number of syllable was centred, since all values were above 3. The random structure consisted of a random intercept for speaker, respective dialogue by speaker, and partner, and a random slope for number of syllables by speaker. We treated all absolute values of t greater than 2 as significant. The influence of the number of syllables was significant for both, negative f0 slopes (β= -0.057, t= -9.741) and positive f0 slopes (β= -0.073, t= -14.13).

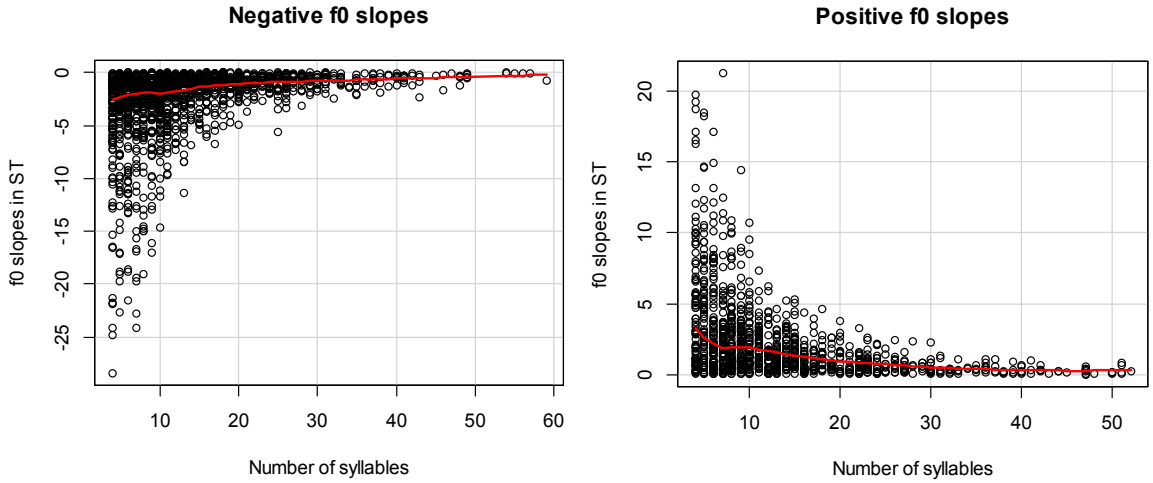

Figure 2: f0 midline slopes as a function of number of syllables. Data are split into f0 slopes <0 (left graph) and >0 (right graph). IPUs with more than 3 syllables were taken into account.

3.2 Communicative role of f0 declination as well as initial and final f0 values

We predicted that negative f0 slopes are more frequent in turn taking or turn ending, while positive slopes are a characteristic of turn continuations. Table 1 summarizes the frequency of occurrence of negative and positive f0 slopes according to the communicative events.

Turn events	All	Negative f0 slopes	Positive f0 slopes
InitialOnly	398	262 (65.83%)	136 (34.17%)
Cont	1473	806 (54.72%)	667 (45.28%)
FinalOnly	415	248 (59.67%)	167 (40.24%)
Both	278	173 (62.23%)	105 (37.77%)

Table 1 – Frequency of occurrence of negative and positive f0 slopes in different communicative events.

Results in Table 1 generally support our predictions when comparing different turn events within a category. The highest occurrence of negative f0 slopes occurs turn initially (~66%), turn finally (~60%) and when turns consist of one inter-pausal unit (62%), while continuation turns have the lowest proportion of negative f0 slopes (~55%). However, negative f0 slopes are generally more frequent than positive ones for all categories, including turn continuations.

The effects of communicative events on f0 slopes as well as initial and final f0 values of the midline are displayed in Figure 3. In general variations according to the communicative events are very small and are mainly observed for f0 slopes (upper track in Figure 3). For the negative f0 slope (top left plot), a marginal effect was found for turn continuation (Cont vs. FinalOnly: β= 0.12822, t=2.238 and Cont vs. Both: β = 2.663, t=2.079, similar statistical model as above, but with communicative event as fixed factor). Shallower negative f0 slopes occur in turn continuations in comparison to turn ends and single turn IPU ("Both"). Positive f0 slopes are also shallower at turn taking and turn continuations than at the end of a turn or single IPU turns (InitialOnly vs. FinalOnly: β=0.433, t=2.725 and InitialOnly vs. Both: β= 0.6741, t=3.67).

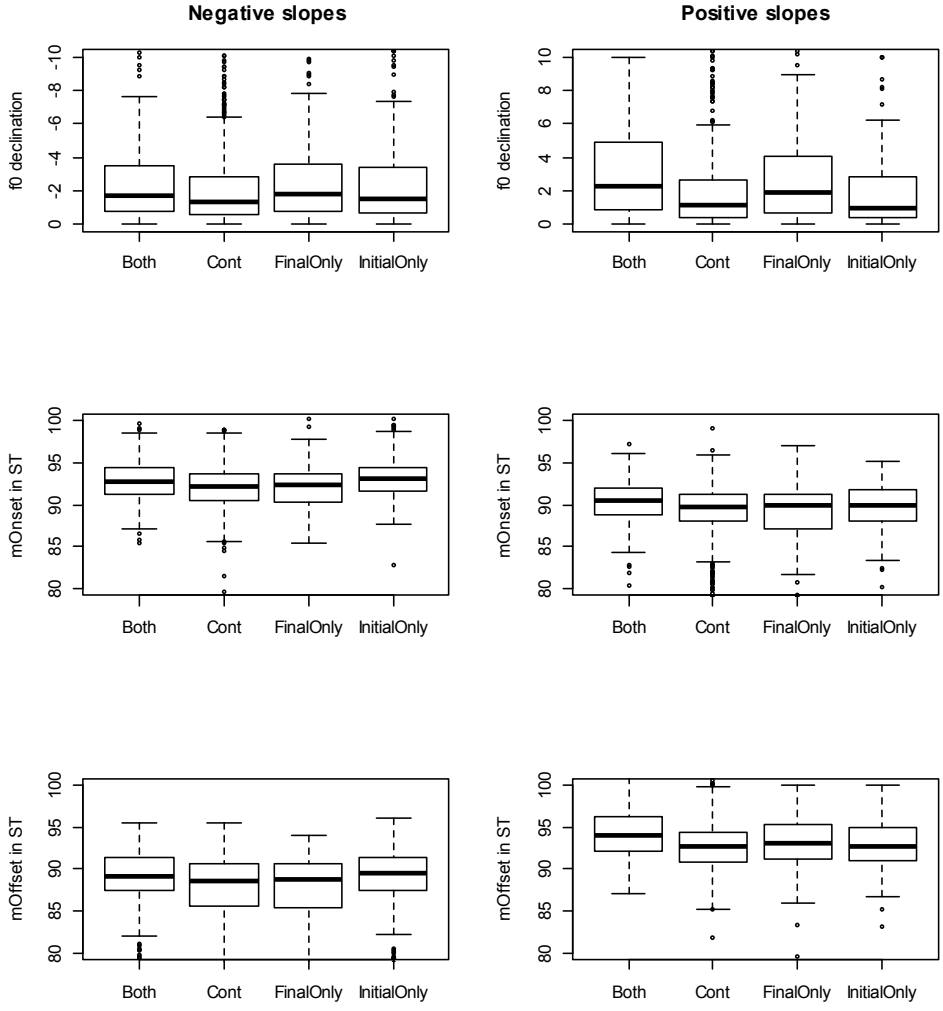

Figure 3: Boxplots for f0 midline slopes (upper tracks), initial f0 value of the midline (middle tracks), final f0 value of the midline (lower tracks) for different communicative events. Data are split into negative f0 slopes <0 (left graphs) and positive ones f0 slope >0 (right graphs).

The initial f0 value of the midline depends on the sign of f0 slope, but is then poorly sensitive to the communicative event. Initial f0 values are greater for negative f0 slopes than positive ones, while the reverse is observed for final f0 values The height of f0 at the onset (or offset)

of an IPU may predict the sign of the slopes (e.g. the evolution of f0 during the IPU), but not the type of communicative events (e.g. is the speaker starting, ending or continuing).

4 Discussion and conclusion

F0 declination of speech utterances is an important feature of spoken languages that may result from a complex interaction between physiological, cognitive, and communicative demands. The first aim of this study was to evaluate if f0 declination is related to the number of syllable in an utterance and could therefore be an indicator for utterance length' anticipation. Our results show, for negative as well as positive f0 midline slopes, that shallow slopes are consistently found in utterances consisting of a larger number of syllables (more than ~ 15 syllables). In contrast, short utterances can have shallow or steep f0 slopes and vary to a large extent. These profiles were very similar to what we observed in spontaneous monologues [5], and extents the findings to face-to-face interactions.

The second aim of this study was to understand the role of f0 declination and local f0 values in turn-taking processes. Many previous studies focused rather on local pitch changes than global f0 declination parameters in the organization of turn taking. For instance, continuation rises may signal the interlocutors that the speaker intends to continue his/her turn. Hence, we expected more positive f0 slopes for turn continuations and more negative slopes for turn taking and turn ending. The former was supposed because Nakajima and Allen [9] reported high declination peaks at topic boundary shifts (could correspond to turn taking) and lower ones at elaboration boundaries (could correspond to turn continuations). The latter was suggested, since turn continuations may go hand in hand with rising f0 patterns at the end of an utterance [1]. When data were separated by the sign of the f0 slopes, our results confirmed the predictions. However, in general we observed more negative slopes than positive ones for all data.

Furthermore, our findings revealed consistently shallower positive and negative midline slopes for turn continuations in comparison to final turns. The effects were however, not very strong. Local f0 values might have played a more important role in this respect. Thus, we also looked at initial and final f0 values of the f0 midline. It was evident, that these single f0 values did not show a relation with respect to turn events, but differed when comparing positive and negative f0 slopes. However, we think that we cannot draw strong conclusions about the role of single f0 values on turn taking mechanisms, because we automatically calculated these single f0 values at the on- and offset of the f0 midline. The algorithm we used is very robust against outliers (see Figure 4), hence the initial and final f0 values of the f0 midline may not correspond to the actual f0 peaks at these boundaries. A manual analysis looking for pitch accents (phonological features) at the boundaries of inter-pausal units maybe more time-consuming, but fruitful approach. Besides these limitations of our work, we think that turn-taking mechanism are manifold and multi-modal, including not only acoustic cues, but also nonverbal gestures, eye gaze, posture, respiration and many more. Hence, interlocutors can rely on a huge variety of mechanisms to signal turn taking, continuation or end. In this sense, we do not expect very strong effects for the use of one particular cue only. Taken together, f0 declination is an indicator of the length of the upcoming sentence in dialogues and is to some degree used for the organization of turn taking.

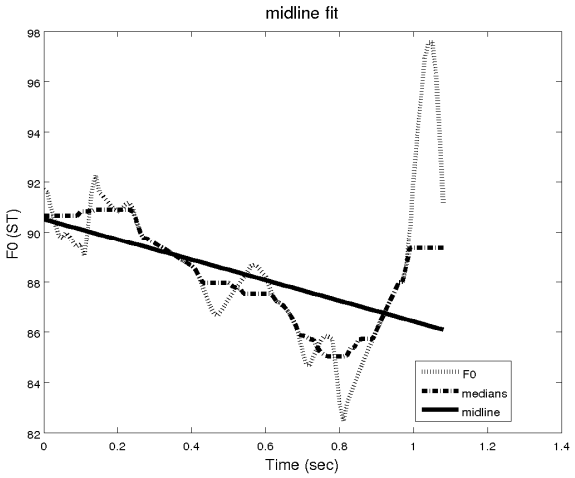

Figure 4: Example for the robustness of the algorithm. Negative f0 midline co-occurs with final continuation rise.

Acknowledgements

This work was supported by a grant from the Ministry for Education and Research (BMBF 01UG0711) to the Laboratory Phonology group at ZAS. We like to thank Anna Sapronova for initial annotation of the texts, Caroline Magister and Jörg Dreyer for support during the experiment.

References

[1] Beattie, G.W., Cutler, A. & Pearson, M. 1982. Why is Mrs. Thatcher interrupted so often? *Nature*, 300, 744–747.

[2] Benus, S., Gravano, A. & Hirschberg, J. 2007. The prosody of backchannels in American English. *Proceedings of the ICPhS*, Saarbrücken, 1065–1068.

[3] Boersma, P. & Weenink, D. 1999. Praat: *A system for doing phonetics by computer.* [Computer program]. Retrieved from http://www.praat.org/

[4] Bögels, S. & Torreira, F. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52, 47–56.

[5] Fuchs, Petrone, Rochet-Capellan, Reichel & Koenig 2015. Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics,* 52, 35– 45.

[6] Hombert, J.-M. 1974. Universals of downdrift: Their phonetic basis and significance for a theory of tone. In W. R. Leben (Ed.), *Papers from the fifth conference on African linguistics* (Supplement 5, pp. 169–183).

[7] Lieberman, P. 1967. Intonation, perception and language. Cambridge, MA: MIT Press.

[8] Maeda, S. 1976. A characterization of American English intonation (Unpublished doctoral dissertation). MIT.

[9] Nakajima, S. & Allen, J.F. 1993. A study on prosody and discourse structure in cooperative dialogues. *Phonetica*, 50, 197–210.

[10] Ohala, J.J. 1978. The production of tone. In V.A. Fromkin (Ed.) *Tone: A linguistic survey* (pp. 5–39). NewYork: Academic Press.

[11] R Core Team. 2015. *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org/

[12] Reichel, U.D. 2012. PermA and Balloon: Tools for string alignment and text processing. *Proc. of Interspeech*. Portland, Oregon, paper 346

[13] Reichel, U.D. and Mády, K. 2014. Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian. *Proceedings of Interspeech*. Singapore, pp 111–115.

[14] Rochet-Capellan, A. & Fuchs, S. 2014. Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 20130399.

[15] Strik, H. & Boves, L. 1995. Downtrend in f0 and Psub. *Journal of Phonetics*, 23, 203–220.

[16] Savitzky, A. & Golay, M.J.E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627–1639.

[17] Svetozarova, N.D. & Kuosmanen, A. 2003. Declination and finality in spontaneous and read speech in Russian. *Proccedings of the ICPhS*, Barcelona, 1297–1300.

[18] Yuan, J.& Liberman, M. 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, 65, 67–74.