9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Intonation and levels of agreement in interactions between Swedish adolescents

Julia Forsberg[1], *Åsa Abelin*[1]

[1]University of Gothenburg, Sweden
Julia.forsberg@gu.se, asa.abelin@gu.se

Abstract

Intonation features and the interpretation of meaning were studied in recorded interactions between Swedish adolescents performing a map-task game. Dyadic interactions of 6 female speakers from one Gothenburg school were analysed with regard to F0 range and direction of final F0 contour in 193 utterances of "*OK*". These utterances were also analysed in relation to the speaker's role in the interaction as information givers or information receivers. A subset of the isolated utterances of *OK* were included in a web-based listening test with 180 listeners, who were instructed to judge whether 12 tokens of *OK* conveyed different degrees of agreement: *full agreement; partial agreement; partial disagreement;* and *full disagreement*. Results show that listeners were quite accurate in identifying the degree of agreement of the isolated *OK* utterances. Furthermore, indications that F0 contour direction is associated with inferred and interpreted agreement were found.

Index Terms: prosody, interaction, perception, map-task, intonation, discourse particle, agreement, adolescent speech

1. Introduction

The purpose of this paper is to analyse different functions of intonation in the discourse particle *OK* in spoken interaction between Swedish adolescents, recorded when interacting in Swedish and English in a map-task game. The intonation was analysed with regard to F0 characteristics (see [1] for recent work on the relationship between perceived speaker characteristics) and this analysis was related to the results of a listening test. The findings will contribute to an understanding of the acoustics and perception of the discourse particle *OK* in interaction, especially between adolescents.

Research questions on production:

- How is the number of *OK* utterances associated with the speakers' roles as information givers or information receivers, or with which language they speak, or with individual variation?
- In what ways are the intonational features of *OK* associated with speakers' roles as information givers or information receivers?

Research question on interpretations:

- Are listeners able to infer speaker intentions, i.e. degrees of agreement, in isolated utterances of *OK*?

Research question connecting intonation and interpretations:

- How do listener assessments of speaker intentions correspond with intonational features of F0 range and F0 contour direction?

2. Background

The discourse particle *OK* can, in the present context of a map-task game, be seen as expressing degrees of understanding of or agreement with what the previous speaker has uttered. In Svensk Ordbok [2] *OK* is defined as an expression of acceptance. *OK* was borrowed from English in the 1930s and was until the 1960s only used to convey acceptance [6]. It is now one of the most widely used discourse markers expressing acceptance Swedish (alongside others, such as *'ja', 'javisst', 'jaha', jasså')* and can typically start or end a spoken turn, or be uttered in isolation. *OK* is described as a regulating dialogue particle [3] and can be used to initiate a sequence of repair within a turn constructional unit (TCU), for example to redesign a turn [4], within a formulation [5], or to initiate alignment [6]. It can also have a forward-pointing function, indicating a readiness to move on to the next topic [7].

As the *OK* can be uttered using many different prosodic patterns, we hypothesize that these patterns signal attitudes towards what the previous speaker has said. The design of the map-task game elicits different expressions of agreement, since the maps given to the interactants are different. It can be seen as a special case of consulting (namely wayfinding) about which the receiver can express different amounts of confidence. As the two maps are partly different the confidence of the receiver in the giver and in the truth of what she is saying will in some instances be challenged. The attitudes examined in this paper include *full agreement* (from here on referred to as 'agree'), *partial agreement requesting more information* (referred to as 'go on'), *partial disagreement* ('hold on'), and *full disagreement* ('doubt'). These attitudes can be placed on a scale from full agreement to no agreement (but see [8] for a discussion on labelling emotional speech).

Of interest to our study is the methodology used by Couper-Kuhlen [9] who systematically examined the prosody of the specific response particle "*oh*" in a specific sequential location. She studied displays of disappointment through the expression *oh* following a rejection of a request or proposal and analysed the acoustic-perceptual characteristics of these *oh:s* and showed how co-participants in the interactions interpret the *oh:s* as expressing disappointment. *Oh*:s expressing disappointment were more difficult to interpret out of context, and were seen as e.g. sympathy instead of disappointment. They can thus be described as ambiguous out of context, cf. Abelin [10]. Ruusuvuori [11] builds on Couper-Kuhlen's [9] argument that participants in a conversation interpret emotional displays according to a restricted number of emotions that are possible or accepted in a given context. For example, a rejection of a request could be followed either by a display of disappointment, or by a display of surprise, relief, anger etc. These displays could then be distinguished from each other with the help of prosodic cues.

10.21437/SpeechProsody.2018-11

3. Method

3.1.1. Material

In the project Språkbruk i Stockholm och Göteborg (Language use in Stockholm and Gothenburg) [12], a total of 111 16-19 year-old students, males and females, were recorded in 4 schools, 2 in each city. Self-selected peer pairs of informants were asked to complete map-tasks, designed to elicit sociophonetic data in interaction, in Swedish and in English. The informants acted as either information givers (SG: Swedish Giver) or as information receivers (SR: Swedish Receiver, ER: English Receiver). In the present study, utterances of *OK* spoken by 6 female speakers from an inner-city school in Gothenburg are used. The reason for selecting only females from one of the Gothenburg schools is that intonation in Swedish shows a large variability depending on speaker sex, sociolect and dialect. As we are mainly interested in the receiver conditions for the listener test, the 6 speakers (D01, D02, D03, D04, D11 and D12) were included as they had all completed the English map-task as the ER; the other 7 females in the school had only completed the English map-task as the English Giver (EG) and were thus excluded from the present study. None of the included speakers had completed the task with each other; 4 of them completed the task with another female and 2 with a male.

110 occurrences of the word *OK* uttered in a Swedish map-context and 83 in an ER context were included in the data collection (N=193). The analysis below will mainly concern either SR and SG, or SR and ER. Some further occurrences of *OK* were excluded from the analysis due to their being unanalysable, as they contained for example creaky voice, or laughter, or were too quiet.

3.1.2. Analysis of productions

The 193 occurrences of the word *OK* spoken by the 6 speakers were annotated and analysed with regard to F0 range and final F0 contour using Praat. Both the full 193 and a subset of the 12 SR tokens used in the survey (see 3.1.3) were analysed.

A background variable identified and coded for using Praat [13] was whether the utterance came from the interactant giving or receiving information about the route. For intonation, we used Praat to measure the F0 range of the full token. We also, using primarily auditory analysis, annotated whether the F0 contour was falling, rising or flat on the second (usually stressed) vowel or diphthong. Speech containing creak was not measured. Furthermore, the measures of F0 range and F0 contour were correlated with the assessments by the listeners, and with differences in intonation patterns between information givers and information receivers.

3.1.3. Listener tests

Out of the 110 Swedish utterances, 12 instances of *OK* uttered by SR were played to 180 listeners using an online survey format (Google forms). A further 12 instances of *OK* uttered by ER were included in the survey but not used in this study. The survey was distributed using social media and e-mail lists. Listeners were asked to assess which level of agreement the speaker using the discourse particle expressed. The question was asked in a forced-choice format, and the options were "I agree with you" (*agree*, category 1), "I am listening, keep talking" (*go on*, category 2), "Hold on, let me think" (*hold on*, category 3) and "I am surprised by what you are saying" (*doubt*, category 4). These categories reflect the (inferred) meanings of the uttered *OKs*, as they were interpreted by the authors listening to the utterances within their conversational contexts. These classifications were done by the authors independently, with a high level of agreement. The tokens where the authors made different classifications (N=12) were thereafter jointly agreed upon by both authors. This classification was done using conversational/dialogic as well as auditory cues, prior to the survey and the analysis. 5 instances used in the survey were categorised as *agree*, 5 as *go on* and 2 as *doubt* by the authors. None were classified beforehand as *hold on* (compare fig.5).

4. Results and analysis

Table 1: *Numbers of OK tokens uttered in total and OK/minute by ER, SR and SG. Each recording takes 25-64 minutes, average 14 minutes per map*

	Numbers of OK			**OK/minute**		
	ER	**SR**	**SG**	**ER**	**SR**	**SG**
D01	16	3	3	1.08	0.33	0.19
D02	24	11	15	1.74	1.18	1.01
D03	20	4	5	1.43	0.32	0.33
D04	6	7	3	0.23	0.37	0.16
D11	10	5	5	1.11	0.72	0.56
D12	7	22	27	0.47	1.03	1.23
Mean	13.83	8.67	9.67	1.01	0.66	0.58

Table 1 shows that there is a positive correlation (Pearson's R, r=0.9691 n=6 p=0.001417) between the number of *OK* utterances in SR and SG, indicating that there are individual differences which have a stronger impact than the role which the speaker has (SR or SG). When calculating the correlation for numbers of OK/minute we find similar results (Pearson's R, r=0.9393 n=6 p=0.005415). Some speakers simply produce *OK* more often than other speakers do. 4 speakers use OK more times per minute in the ER condition than in either SR or SG. Comparing the Swedish conditions, 5 speakers use OK more when they are receivers than givers. Correlations were not found between ER and SR, or ER and SG, and overall the speakers produce *OK* more often when speaking English.

4.1.1. Acoustic analysis

In the following analyses, no formal statistical testing was run due to the small number of speakers. The analysis of F0 range and F0 contour of 110 Swedish utterances of *OK* in relation to speaker role (SG or SR) shows that there is no great difference in the use of falling intonation between SR and SG, while rising intonation was more common in the speakers' role as information giver (SG). There are too few instances of flat intonation to make a thorough comparison, but these are as (un-) common in SR and SG. However, as there is only a small number of speakers available, individual differences may play a role.

For mean F0 range the analysis shows a tendency that rising intonation has a large F0 range within the category SG. For SR tokens, rising and falling intonations have a similar F0 range; whereas for both ER and SG, falling intonation has a smaller F0 range than rising intonation.

The relationship between the contour direction and the F0 range magnitude holds true for both the subset of 12 and the full set of 110 tokens: rise shows the largest F0 range, while flat shows the smallest.

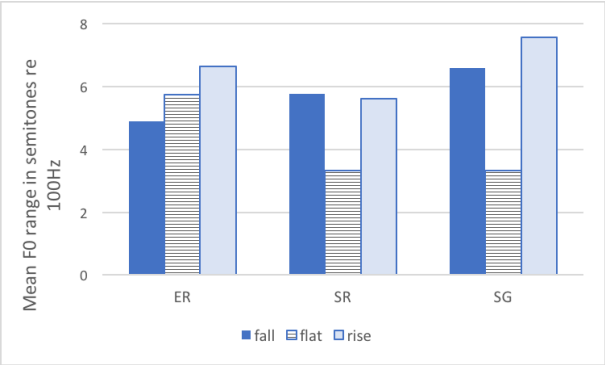

Figure 1. *Mean F0 range (ST re100) measured over the whole word, 193* OK *items. F0 contour direction measured for the final vowel*

Comparing the classifications of the SR and ER utterances to their prosodic characteristics, we can discern three things by observing fig. 2, 3 and 4. Firstly, there appears to be a correlation (in opposite directions) between category (level of agreement) and use of falling vs. rising F0 contour, for both Swedish and English produced by L1 Swedish speakers (fig. 2 and 3). While there are few occurrences of especially the third and fourth categories, these speakers appear to use rising intonation to express disagreement.

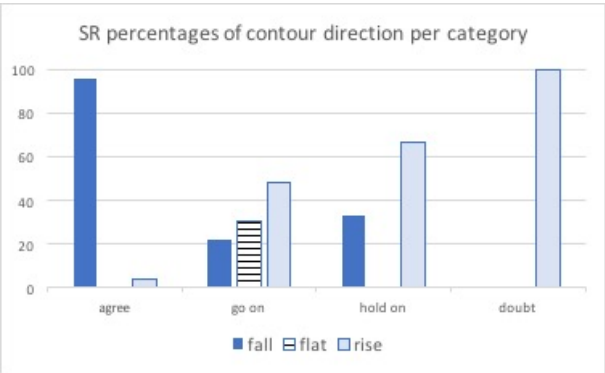

Figure 2. *Percentages for each contour direction by category (1-4; agree-doubt) for SR*

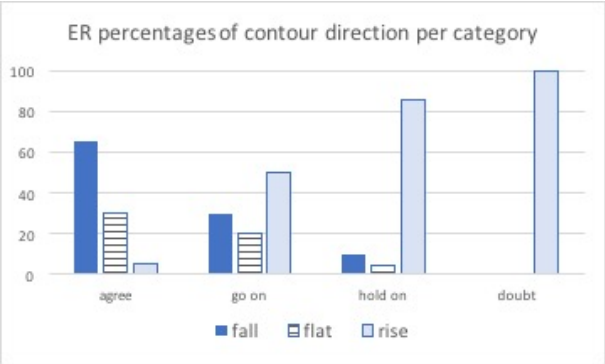

Figure 3. *Percentages for each contour direction by category (1-4; agree-doubt) for ER*

The second observation is that there seems to be no strong relationship between range and category: there is some indication that range increases with level of agreement; however, for full disagreement we have only 2 instances for each language, and for partial disagreement only 3 for Swedish.

The third observation is that ER shows a slightly larger mean F0 range than SR (fig.1 and 4). Each box in fig. 4 indicates the values for the speakers with the highest and lowest F0 range, the median and the first and third quartile. Comparing the ranges in fig. 4 for categories 1 and 2 (full and partial agreement) and 3 ER (partial disagreement for English) using a two-tailed t-test ($p<.05$) shows no significant differences between either of the pairings 1ER-1SR, 2ER-2SR, 1ER-2ER, 2ER-3ER or 1SR-2SR. The remaining categories (to the right of the line in fig.4) were not compared due to the low numbers of tokens.

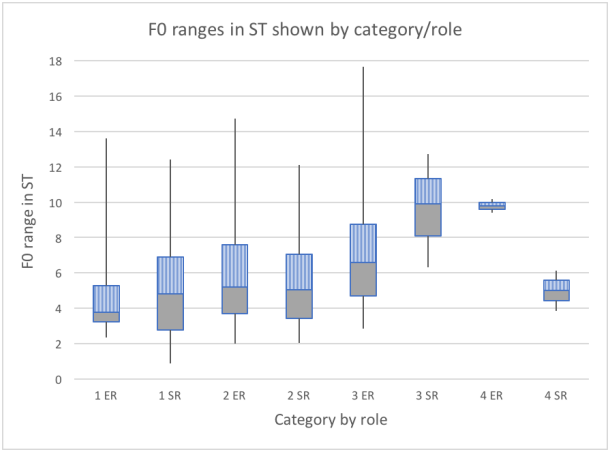

Figure 4. *F0 ranges for each speaker (ST re100) by level of agreement (categories 1-4; agree-doubt) and language (ER or SR)*

4.1.2. *Listener tests*

In the analysis, the categories of agreement were ranked on a scale from *full agreement (agree),* via *partial agreement requesting more information (go on), partial disagreement (hold on),* to *full disagreement (doubt).*

The survey shows that listeners were quite good at assessing the inferred meanings of the *OK* in isolation. Fig. 5 shows how listeners interpret the function behind the uttered *OK* without hearing any conversational context. Listeners tend to be more able to identify the category *agree* than the other categories. No utterances which were pre-classified as *hold on* were included in the listener test; however, some utterances were interpreted by listeners as *hold on*.

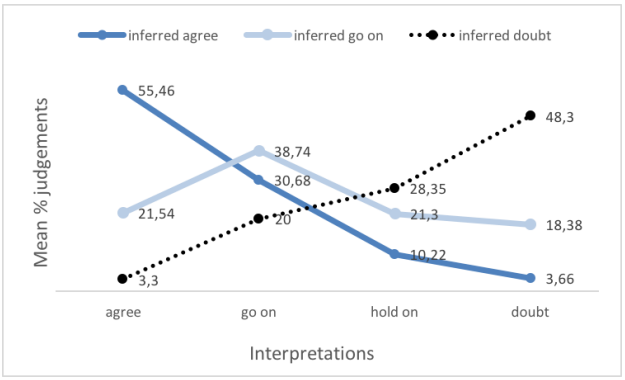

Figure 5. *Illustration of inferred judgements in relation to the listener tests. Mean % based on 180 judgments of 12 OK speech samples.*

Looking at these categories as a scale, it becomes evident that when listeners make an erroneous selection, they usually select one of the adjacent options. Indeed, the interpretations of each inferred category follow the scale perfectly: *Agree* is most different from *doubt*, *go on* lies between *agree* and *hold on and doubt* is most different from *agree*; see fig.5.

4.1.3. Connecting the F0 range and intonation contours with the listener interpretations

The F0 range of *OK* did not vary with the degree of agreement expressed by the speaker for the 12 tokens used in the listener test. On the contrary, the category showing the largest mean F0 range is *go on,* both for the 12 and for the total 52 SR tokens.

In a comparison between the categories of agreement and directions of F0 contour in the 12 tokens used for the listener tests, the results are the following: 'agree' has a falling F0 contour in 5 out of 5 instances; 'go on' has a rise in 3 out of 5 and fall in 2 out of 5 instances; 'doubt' has 2 instances of rise and none of flat or fall. Similar distributions are found in figs.2 and 3 for the full set of receiver tokens. This is in line with Ohala's work on the frequency code [14] which deals with ways of expressing security and insecurity, where low F0 and falling intonation express dominance and high F0 and rising intonation express submission. Rising F0 also had the greatest F0 range and flat F0 had the smallest F0 range, as can be expected.

To summarise, the category 'go on' (request for more information) shows the largest F0 range and the most instances of final rising intonation. 'Agree' shows the most instances of final falling intonation. The number of items is small so all results have to be seen as tendencies only.

5. Conclusions

Conclusions concerning production

- For the total number and frequency of *OK* utterances there are individual differences which have a stronger impact than the role of the speaker (SR or SG). When the speakers spoke English as English receivers (ER), they produced *OK* more often than in the SR and SG conditions.
- SR utterances tend to have a smaller F0 range than SG and ER utterances.
- SG utterances have a larger F0 range for final rising intonation in comparison with falling and flat SG intonation.

Conclusions concerning interpretations of SR

- Listeners were quite accurate at identifying the inferred level of agreement of the *OK*, even though they had not heard the context of the dialogue (cf. fig. 5).

Conclusions concerning the relation between intonation and interpretation of SR/ER

- The largest F0 range was associated with the interpreted category *go on*.
- F0 range is often similar for tokens independent of the final contour, which appears to be a better predictor for assessing inferred intention of the token in isolation.
- Falling F0 contour was most closely associated with interpreted agreement, and slightly more so for ER than SR.
- Rising F0 contour was most closely associated with a request for more information (*go on*) as interpreted by listeners.

6. Discussion and future work

In this material, we find *OK*s having all of the functions described in the background literature, such as in redesigning a turn, within formulations and in the initiation of alignments. The fact that speakers produced *OK* more often when speaking English than Swedish could be explained by the speakers having access to more synonyms to use in the relevant contexts in Swedish, as this is their strongest language.

In comparison to previous results that utterances can be ambiguous out of context, cf. [7], we find that the *full agreement* end of the scale appears to be the least difficult to identify in isolation. The comparison becomes difficult as not all pre-classified categories were present in the listener test. In previous studies the utterance has been given a context and listeners have thus been able to choose from their 'pool' of possible emotional responses, whereas in the present paper listeners do not have access to a given context to which they can connect their distinctions/categorisations. However; a finite number of possible interpretations are given, allowing listeners to construct a (possibly very abstract) context themselves.

Listeners are quite good at interpreting inferred meanings of speakers without having the context but do not reach a 100% match with inferred levels of agreement. The interpretations are, however, close to the inferred meanings and we could argue for these speaker intentions being dimensional rather than categorical. Listeners appear to use prosodic information when interpreting meaning and intention. The F0 contour direction is associated with the level of agreement. This may, however, not give a full picture, and further segmental analysis will be necessary. Given that there are larger differences between F0 range for the different F0 contour directions in SG than in SR, further perception studies into the information giver behaviours (both in Swedish and English) would be beneficial. Perhaps there is greater variation in the frequency of occurrence of *OK* as well as in intonation in the G conditions.

Further research on all 111 speakers (taking into account background variables such as sex, dialect and social factors) in the corpus should be undertaken in order to verify the tendencies found in the present paper. In addition to the prosodic cues, we will in the future also consider the segmental and voice quality information of the *OK*s in relation to the responses given by listeners. It appears that further knowledge is needed in order to get to the essence of what enables a listener to perceive and interpret meaning in isolated utterances, or prohibits him/her from doing so.

7. Acknowledgements

The authors would like to thank the informants who took the time to participate in the perception experiment, as well as those who shared the link in their social media channels, and the speakers who let us use their voices. Thanks to Johan Gross for collaboration with data collection, to Jonas Håkansson for kindly testing the survey prior to its release, to Dom Watt for valuable comments and to the Samtalslabb-group at University of Gothenburg for the valuable discussion on the material.

8. References

[1] P. Belin, B. Boehme and P. McAleer, The sound of trustworthiness: Acoustic-based modulation of perceived vice personality. *PLoS ONE* 12(10): e0185651. 2017.

[2] *Svensk Ordbok [Dictionary of Swedish],* Stockholm: Norstedts, 2009.

[3] J. Lindström, *Tur och ordning: introduktion till svensk samtalsgrammatik. [Turns in order: an introduction to Swedish conversational grammar.], Stockholm:* Norstedts, 2008.

[4] C. Kitzinger "Repair" in J. Sidnell and T. Stivers (Eds.) *The handbook of conversation analysis*. West Sussex: Blackwell, pp.229–256, 2013.

[5] C. Ottesjö, *Att fortsätta och att återgå: studier i koherensskapande praktiker i vardagliga flerpersonssamtal. [To continue and to resume. A study of coherence-creating practices in everyday multiparty conversations.]*, Ph.D dissertation, University of Gothenburg, 2006.

[6] A. Lindström, "Concessive markers in Swedish conversation: A contrastive analysis of the turn-inital particles nåja and okej." *International Conference on Conversation Analysis*, 2014.

[7] W. Beach and A. Lindström, "Conversational universals and comparative theory: Turning to Swedish and American acknowledgment tokens in interaction" in *Communication Theory,* vol.2, no.1, pp. 24-49, 1992.

[8] D. Peres, D. Watt and W. Ferreira Netto, "Emotional Thin-Slicing: A Proposal for a Short- and Long-Term Division of Emotional Speech" *Proceedings of Interspeech 2017*, Stockholm, pp.591-595. 2017.

[9] E. Couper-Kuhlen "A sequential approach to affect: the case of 'disappointment'." In M. Haakana, M. Laakso and J. Lindström (Eds.) *Talk in Interaction – Comparative Dimensions*. Studia Fennica, Linguistica 14. Helsinki, pp. 94-123, 2009.

[10] Å. Abelin, "Expression of emotions in spoken Swedish – a corpus study." *Proceedings of Fonetik 2010*, Lund, pp. 1–5. 2010.

[11] J. Ruusuvuori, "Emotion, affect and conversation analysis." In J. Sidnell & T. Stivers (Eds.) *The handbook of conversation analysis*. John Wiley & Sons, Ltd, Chichester, UK., pp. 330-349, 2013.

[12] J. Forsberg and J. Gross. *"You change your speech depending on who you talk to, but I didn't change much": The map task as a speech event, and the informants' perception of it as a field experiment,* (under review)

[13] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.0.26, retrieved 2 March 2017 from http://www.praat.org/

[14] J. J. Ohala, "Cross-language use of pitch: an ethological view." *Phonetica,* vol.40, no.1, pp. 1-18, 1983.