The impact of syntax and pragmatics on the prosody of dialogue acts

Katalin Mády, Uwe D. Reichel, Beáta Gyuris and Hans–Martin Gärtner
Research Institute for Linguistics, HAS, Hungary

Introduction

Task-oriented spoken dialogues have several advantages: (1) Since speakers are involved in a non-linguistic task, they tend to concentrate less on the fact that they are being recorded, (2) various settings allow for the elicitation of repetitions of certain elements (words, names, etc.). (3) Since these tasks create a specific setting, intentions of speakers are more easy to control in terms of information structure, e.g. whether a certain element is given, new, contrastive etc.

In this paper the dialogue structure coding scheme by [1] was used in order to test whether dialogue acts can be classified based on their prosodic features developed by [4]. Dialogue acts (DA) were investigated under two aspects: (1) they belonged to different sentence types such as yes/no questions vs. wh-questions or to DA pairs, e.g. yes/no questions and positive or negative responses to them, alternatively, (2) they differed in their informational weight within the same sentence type, e.g. explaining new information vs. assuring that previous information was understood correctly in declaratives. The goal was to find out whether syntactic and pragmatic categories can be distinguished by different prosodic features.

Materials and methods

Data were taken from the Hungarian version of the object game of the Columbia Games Corpus [2] based on a computer-aided game with two participants and two laptops. Players see objects on their screens that are identical except for one object. The first player describes the position of the blinking object in relation to the other objects. The second player is supposed to place the object in exactly the same position. Participants get a score after each turn (altogether 14 in each game) on a 0 to 100 scale. (See [3] for more detail).

Annotation: The signal was manually segmented into inter-pausal chunks, text-transcribed annotated for dialog acts. F0 was extracted by Praat and preprocessed as described in [4]. Within each chunk prosodic phrases were extracted, and within each dialog act the most prominent syllable. Details on these unsupervised automatic annotation methods are given in [4].

Feature extraction: From this annotation we extracted temporal, energy and f0 features (cf. Table 1) on the entire dialog level (*glob*) and in an analysis window of length 0.3s around the most prominent syllable (*loc*). One part of the syllable-related features refers to its *Gestalt* properties, i.e. to what extent its f0 register is distinct from the underlying intonation phrase. For this purpose a base-, mid- and a topline were fitted both to the syllable as well as to the related prosodic phrase, and for each line pair the RMS was calculated within the syllable analysis window. The second local feature set describes the shape of the f0 contour in terms of the coefficient values of a third order polynomial. All features were extracted within the *CoPaSul* intonation stylization framework by a freely available toolkit [4].

Table 1: Temporal, f0, and energy features for each dialog act. *glob*-features were extracted on the dialog act level, *loc*-features in an analysis window centered on the nucleus of the most prominent syllable within the dialog act

Dialog act level temporal (glob_temp), f0 (glob_f0), and energy (glob_en)	
dur	duration
syl_rate	number of syllables per second
f0_max,med,sd	f0 maximum, median, standard deviation
en_max,med,rms,sd	energy maximum, median, RMS, standard deviation
Syllable level f0 *Gestalt* (loc_gst) and shape (loc_shape)	
loc_bl,ml,tl_rms	RMS between syllable and IP baseline, midline, topline
loc_c0–3	polynomial coefficients 0–3

Results

Mann-Whitney and Kruskal-Wallis tests were used due to the lack of normal distribution in all samples. Significance level was set to $p < 0.05$. 1 First, two pairs of DAs belonging to different sentence types were compared: (1) yes/no questions (QY) and the positive response (RY) given to them, (2) yes/no questions (QY) and *wh*-questions (QW). The above sentence types were distinguished mostly by local features: QY had higher *Gestalt* values than RY, presumably due to the obligatory low accent in yes/no questions, and QY and QW were differentiated by

accent shape, supposedly being linked to a low accent in the first and a falling one in the second question type.

Two types of declaratives, the general category *EXPLAIN* (EX) containing new information and *CLARIFY* (CL) used for reassuring that the speaker has understood previous information properly, i.e. all-given information, were compared. EX was characterised by higher overall energy and longer duration than CL. Two other declaratives, COMMENT (CO) and READY (RE) were also compared to EX. These latter categories do not contain information relevant for the task itself, but either comments such as 'well, this is all I can tell you' or a transition to the next turn 'o.k., so I press the button'. Thus again, their informational weight is lower than that of EX. Again, global features connected to duration, energy, and syllable rate show higher values for EX, while interestingly, *Gestalt* values showed a higher emphasis on the most prominent syllable for comments that often expressed emotions.

Based on [1]'s scheme, DAs were divided into three categories based on their position within a turn during the game. Initiations are mostly questions, responses consist of declaratives, while the category preparation signalises that a speaker is ready for a new turn. Initiations were realised with higher values for most global categories, i.e. duration, f0, energy and syllable rate. Preparation and response were best distinguished by accent shape.

Discussion

In this paper, a first attempt was made to test whether dialogue acts suggested by [1] can be characterised by stylised prosodic parameters. Comparisons were either based on syntactic or on pragmatic categories. While DAs that express grammatical categories such as various question types seem to be connected to different prosodic categories based on local features, DAs that belong to the same sentence type but carry different pragmatic meaning tend to be distinguished along global prosodic parameters. The findings of the present study will be extended by more pragmatic categories. A mid-term goal is to predict DAs simply on the basis of automatic prosodic feature extraction.

References

[1] J. Carletta, A. Isard, S. Isard, J.C. Kowtko, G. Doherty-Sneddon, and A.H. Anderson 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

[2] Agustín Gravano, Štefan Beňuš, Héktor Chávez, Julia Hirschberg, and Lauren Wilcox 2007. On the role of context and prosody in the interpretation of `okay'. In *Proc. 45th Annual Meeting of Association of Computacional Linguistics*, Prague, 800–807.

[3] Katalin Mády and Uwe D. Reichel. 2016. How to distinguish between self- and other-directed wh-questions? In Proc. *Phonetik und Phonologie im deutschsprachigen Raum*, Munich, Germany.

[4] U.D. Reichel 2017. *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*. RIL, MTA, Budapest, Hungary. https://arxiv.org/abs/1612.04765.

Acknowledgements

This work was funded by the Alexander von Humboldt Foundation and OTKA K 115922.