DODGING THE QUESTION IN COMPETITIVE SPOKEN DIALOGS: SEMANTIC AND PROSODIC CHARACTERISTICS

Uwe D. Reichel[1]*, Piroska Lendvai*[2]

[1]*Institute of Phonetics and Speech Processing, University of Munich*
[2]*Göttingen Centre for Digital Humanities, University of Göttingen*
reichelu@phonetik.uni-muenchen.de

Abstract: We examined conversational non-cooperation in human–human task-oriented dialogues in English. Our approach was to identify prosodic and semantic patterns that characterize replies to information-seeking dialogue acts of task-solving, in a corpus holding competitive and cooperative scenarios. Appropriate and inappropriate replies were manually defined on the dialog act level; all information-containing replies were characterized in terms of speech signal and semantic content (vocabulary, entropy, sentiment). We found that communicative means for dodging a question include reduced content-providing, increased negative sentiment and a stronger F0 declination trend. Some prosodic markers also indicate that holding back information increases the speaker's cognitive workload.

1 Introduction

Our study focuses on a specific non-cooperative dialog strategy: dodging the interlocutor's question, i.e., avoiding to provide solicited information, in human–human task-solving dialogues. Deceptive discourse is an actively investigated area, including signal data [1], and there are studies on incorporating non-cooperative behavior in human–machine dialog management, in particular as a beneficial strategy of virtual agents [2, 3]. However, non-cooperation in terms of strategically applied speech acts has so far been less addressed in human–human (spoken) dialog research. [4] recently presented an elaborate scheme for annotating and analyzing naturally occurring dialog in terms of linguistic non-cooperation. They measured how much participants of political interviews deviated from correct dialog behavior – as defined by discourse obligations, cf. [5] –, derived from a corpus of cooperative dialogs. Non-cooperation was scored for discourse roles (interviewer, interviewee), and relative to the interlocutor. [4] concluded that incorporating prosodic aspects would be a welcome add-on to the approach.

We propose to analyze prosodic and semantic aspects of dialog-act-level non-cooperation, using statistical methods. The lack of resources to this end is remarkable: (non-)cooperative behavior is annotated in corpora typically on the level of the task but not on the level of dialog acts (DAs). In DA schemes, i.e. taxonomies that explicate communicative functions, dimensions that can be relevant for annotating non-cooperation are only recently suggested; e.g. to establish functional dependence relations between DAs for indicating which question is replied by an answer [6]. We chose to manually define a set of non-cooperative DA pairs that violate the Gricean *cooperative principle* [7], in order to test our hypothesis that in competitive settings respondents operationalize various communicative means to dodge the question, and these exhibit characteristic patterns in terms of prosody as well as content semantics.

The Gricean cooperative principle is described by four maxims: *Quantity:* make your contribution as informative as is required; *Quality:* do not say what you believe to be false or for which you lack adequate evidence; *Relevance:* contribute to the ongoing conversation topic, and

Table 1 – Our tagset, originally developed by [11] with examples from the corpus employed.

code	orig. label	explanation	examples
AC	Acknowledgment	listener feedback	*Ok., Yeah.*
AL	Alignment	checking attention	*Ready?*
CH	Check	request confirmation	*Like this?*
CL	Clarify	give additional information	*It's more like it's inwards ...*
EX	Explain	provide information	*It looks like a batman symbol*
IN	Instruct	Command	*Look at me!*
QW	Question-W	Wh-question	*So what image do you think we have?*
QY	Question-YN	Yes-no question	*Does it have a door?*
RE	Ready	Indicating readiness	*Ok!, Alright!*
RN	Reply-No	No-reply.	*No.*
RW	Reply-W	A reply to QW	*Like a diamond face.*
RY	Reply-Y	Yes-reply	*Yes.*

Manner: avoid obscurity of expression and ambiguity. Related to the maxim of relevance, [8] developed the *Relevance Theory*. Within this framework, the *relevance* of an utterance for the hearer is defined as a function of positive cognitive effect, i.e. the importance of the conveyed information, and of processing effort for the hearer to extract and make use of this information. Within these frameworks we expect non-cooperative behavior in terms of discourse obligations [5] to be reflected in:

- violating the Gricean maxim of relevance, i.e. provide less adequate reactions to questions, which we will examine in terms of distributional characteristics of dialog acts,
- violating the maxim of quantity by holding back information, which will examine by text-based statistical means,
- decreased empathy with the interlocutor, in terms of text-based sentiment analysis, and
- violating the maxim of manner by increasing the listener's effort to process the reply by prosodic means.

2 Data

We used parts of the Illinois Game Corpus [9, 10] that contains Tangram game dialogs (a puzzle of seven pieces to be combined to various shapes) by American English speakers in cooperative and competitive settings. Both dialog partners were presented with Tangram silhouettes reciprocally hidden from the partner. The task was to decide if the silhouettes were the same, by verbally describing them to each other. In the cooperative setting the partners solved this goal in a joint effort. In the competitive setting, partners were required to solve the task competitively: solving it first entailed winning. The data was manually DA-annotated using the tag set of [11], briefly described in Table 1. Two annotators labeled the data, mismatches were resolved. We selected all acts related to questions about task-solving ($n = 755$), for which either new content or the grounding of content is expected by the interlocutor: i.e. yes-no questions (QY), wh-questions (QW), and checks (CH). To each information-seeking DA we linked the subsequent DA of the interlocutor (skipping hesitations) which we then examined on (1) DA distribution level, (2) text level, and (3) prosodic level. Features are listed in Table 2 and introduced next.

3 Analysis of cooperative and competitive behavior

3.1 Dialog act distributional properties

First we counted separately for the cooperative and competitive dialog condition the amount of missing interlocutor reactions (*answQ*, cf. Table 2), i.e. all cases where an information seeking

Table 2 – Extracted features. The direction of significant differences ($\alpha = 0.05$ after type 1 error correction) is indicated in the final column. Further specifications: *w.s.*: weakly significant ($\alpha = 0.1$), *n.s.*: not significant, *n.t.*: not tested.

feature	description	significant difference
DA distribution		
answQ	prop. question answered	coop > comp
appropReply	prop. appropriate replies	coop >comp (w.s.)
H(reply|quest)	conditional entropy	coop < comp (n.t.)
Text features		
sentiment pos	positive sentiment replies	coop >comp
sentiment neg	negative sentiment replies	coop <comp
H	word unigram entropy in reply	coop >comp
newH	new words unigram entropy	coop >comp
newV	proportion of new words	coop >comp
Prosodic features: DA level		
t_transit	turn taking latency (sec)	coop <comp
syl_rate	syllable rate (syl/sec)	n.s.
dur	reply duration (sec)	coop >comp (w.s.)
ml_m	midline mean	coop <comp
ml_c1	midline slope	n.s.
rng_m	range mean	n.s.
rng_c1	range slope	coop >comp
en_med	medium energy	coop <comp
Prosodic features: accent level		
ml_rms	F0 midline deviation	coop <comp
rng_rms	F0 range deviation	coop <comp
c0	0th order poly coef of F0 contour	n.s.
c1	1st order poly coef of F0 contour	n.s.
c1	2nd order poly coef of F0 contour	n.s.
c3	3rd order poly coef of F0 contour	coop >comp (w.s.)

DA of a speaker is not followed by a DA of the interlocutor but by another DA of still the same speaker. Then we subdivided the reactions linked to the information seeking DAs into *appropriate* and *inappropriate* and counted their respective amounts in cooperative and competitive dialogs (*appropReply*). For this purpose for each of the three informations seeking DAs we assigned an exclusive list of appropriate reactions as follows:

- CH: AC, CL, EX, RN, RW, RY
- QW: CL, EX, RW
- QY: AC, RN, RW, RY

Finally, we measured the conditional entropy *H(reply|quest)* of reactions' DAs given information-seeking DAs, in order to compare DA sequence stability across dialog conditions.

3.2 Text-based content analysis

We applied the Vader sentiment analysis tool [12] on the response DAs. Vader returns three scores for positive, negative, and neutral sentiment. Each reply was classified according to the maximum of these three values. To obtain further content features, we preprocessed the texts: removed punctuation, lowercased and stemmed tokens ([13] implemented in [14]). We quantified the amount of information content in terms of entropy (*H*, cf. Table 2) of the contained type set for all response DAs that provide information other than just *yes* or *no*, i.e. for CL, EX,

and RW (cf. Table 1). The underlying word stem unigram probability model (maximum likelihood estimate) was trained on the entire corpus. Additionally, we calculated the entropy for the non-overlapping part of the reply and the question content (*newH*) as well as the proportion of non-overlapping types in the reply (*newV*).

3.3 Prosodic parameterization

The parameterization of the f0 contour was carried out in the contour-based superpositional CoPaSul stylization framework [15]. In this framework f0 is decomposed into a global component, here corresponding to the DA segment, and local components corresponding to pitch accents. From this stylization, feature sets were extracted on DA and accent level.

Preprocessing. F0 was extracted by autocorrelation (Praat 6.0 [16], sample rate 100 Hz). Voiceless utterance parts and f0 outliers were bridged by linear interpolation. The contour was then smoothed by Savitzky Golay filtering [17] 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 a speaker within a dialog and served to normalize f0 with respect to the speaker's overall level. Energy in terms of root mean squared deviation was calculated with a sample rate of 100 Hz in Hamming windows of 50 ms length. Syllables and pitch accents were extracted automatically by the CoPaSul toolkit as introduced and validated in [18].

Dialog act level features. As shown in Figure 1, a base-, a mid- and a topline are fitted through the f0 contour in a DA by means of linear regression. Time was normalized from 0 to 1. Further details are described in [15]. We treat register in terms of *level* and *range* as follows: the midline represents register level aspects, and the pointwise distance between base- and topline represents register range aspects. More precisely, in our approach range is parameterized by means of a linear regression through these pointwise distances. A negative slope of the range regression line thus indicates converging base- and topline, whereas a positive slope indicates line divergence. From the midline and the range regression line the parameters mean and slope were considered for further analyses (features *ml|rng_m|c1*, cf. Table 2). Additionally, the median energy over the DA segment (*en_med*), its duration (*dur*) and the turn taking latency (*t_transit*) to the corresponding question, both in seconds, was calculated.

Accent features. The f0 contour within an analysis window of 0.3 sec length centered on the vowel midpoint of the detected accented syllable is represented by its shape and by its deviation from the underlying IP. For capturing its shape we fitted 3rd order polynomials to the time-normalized contour (cf. [15]), of which we derived the coefficients as shape features for further examination (features *c0–3*, cf. Table 2). The local register in these segments was derived analogously to the DA level as described in the previous paragraph. Deviation of the accent-related from the DA-related register was measured by the root mean squared deviation of the accent-level regression line with the corresponding DA-level regression line stretch (*ml|rng_rms*).

4 Results

Dialog act distribution. The influence of the dialog condition on the amount of answered questions, and on the amount appropriate answers is shown in Figure 2 and was tested by χ^2 tests ($\alpha = 0.05$). In cooperative dialogs significantly more questions triggered a reaction by the interlocutor ($p = 0.0484$, $\chi^2_1 = 3.8965$). The amount of appropriate answers was weakly significantly higher in cooperative dialogs ($p = 0.0935$, $\chi^2_1 = 2.8126$). As seen in Figure 3, the conditional reply probabilities for CH and QY are more equally distributed in the

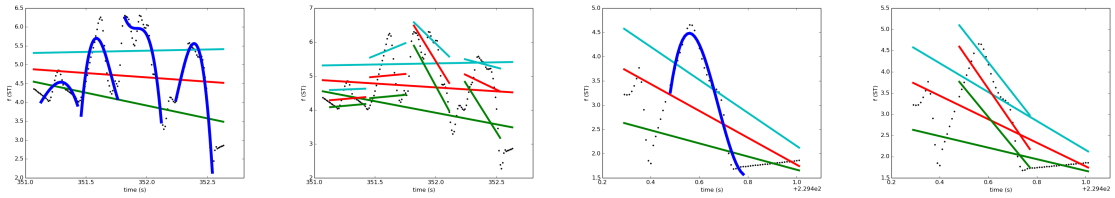

Figure 1 – Stylization examples of dialog act RW in cooperation (**left, left-mid**) and in competition (**right-mid, right**). **left, right-mid:** Superpositional f0 stylization into DA-level register and accent-level f0 shape. **left-mid, right:** Accent register deviation from the dialog act register. The competitive example is characterized by converging dialog act register range and lower duration.

competitive setting, which is also reflected in a higher overall conditional entropy measure H(responselquestion)=2.36, as opposed to 2.05 in cooperative dialogs.

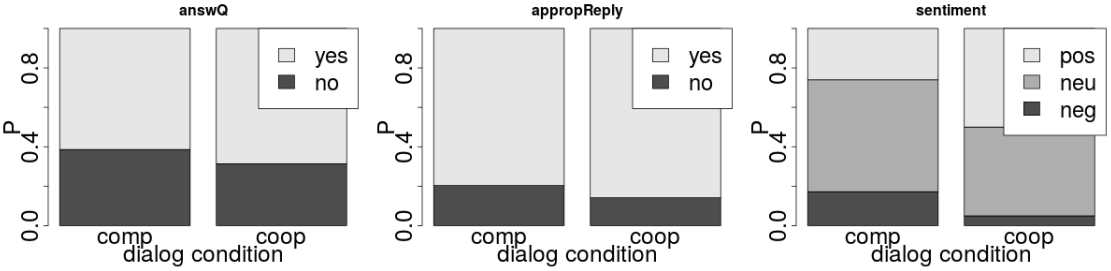

Figure 2 – Proportions P of: **left:** answered questions, **mid:** appropriate replies, and **right:** sentiment polarities of the reply in cooperative and competitive dialog acts.

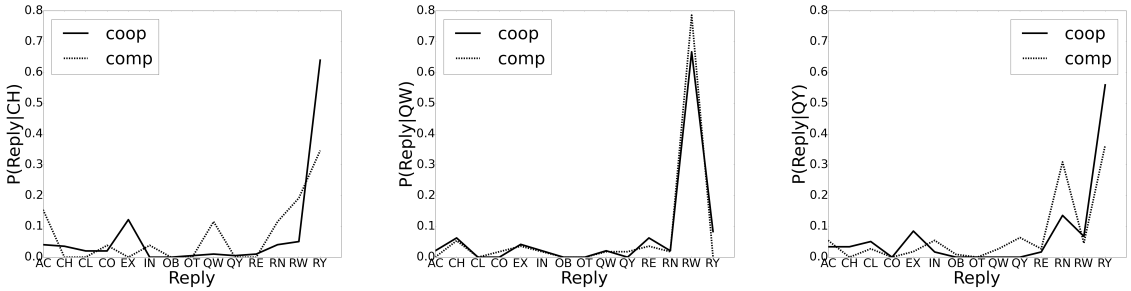

Figure 3 – Conditional probabilities of reply dialog acts to the question types CH, QW, and QY in cooperative and competitive settings.

Sentiment. The influence of the dialog condition on sentiment polarities is shown in Figure 2 and was tested by a χ^2 test ($\alpha = 0.05$). The sentiment proportions differed highly significantly ($p < 0.0001$, $\chi^2_2 = 38.094$) with overall more positive and less negative polarity in the cooperative dialogs.

Text and signal features. For the additional text- and signal-based features we applied linear mixed effect models with dialog condition as the fixed effect and the replying speaker as the random effect for which a random intercept model was calculated. The results are shown in Figures 4, 5, and 6. *P*-values were corrected for false discovery rate [19]. The significance level was set to 0.05. Significant differences are reported in Table 2.

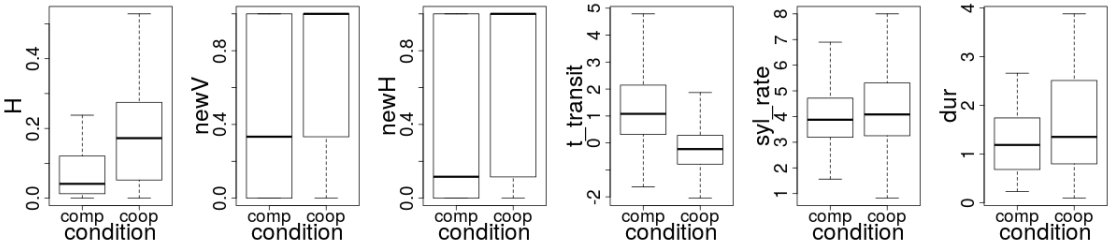

Figure 4 – Text-based and temporal features in reply dialog acts in cooperative and competitive dialogs.

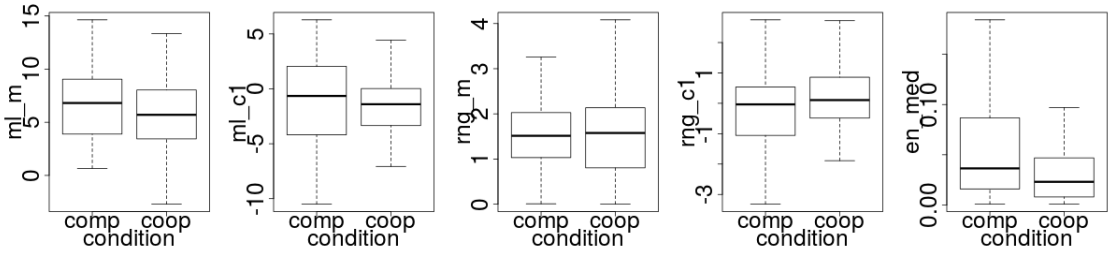

Figure 5 – Prosodic dialog-act level features in reply dialog acts in cooperative and competitive dialogs.

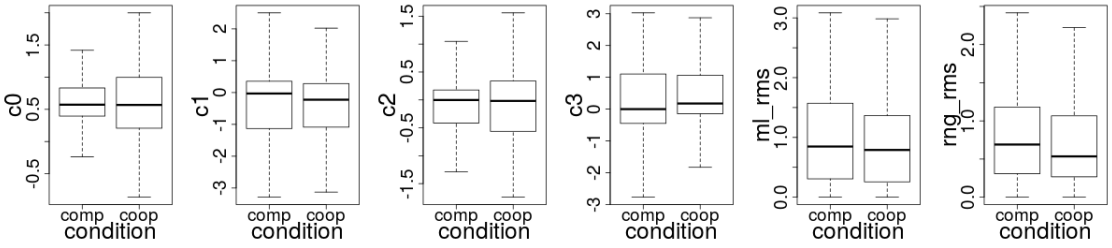

Figure 6 – Prosodic accent level features in reply dialog acts in cooperative and competitive dialogs.

5 Discussion and conclusion

To summarize the results, for competitive dialogs compared with cooperative ones we found:

R1 a higher amount of inappropriate, hard-to-predict and missing replies,
R2 a lower amount of new vocabulary, and a lower entropy,
R3 a lower positive and a higher negative sentiment score,
R4 shorter dialog act duration,
R5 larger turn-taking latencies,
R6 a stronger F0 declination trend, and at the same time
R7 a higher F0 register with respect to both level, and range, as well as
R8 an overall higher energy, and
R9 more prominent pitch accent marking.

Findings R1 and R2 are well in line with the expectations formulated in Section 1, that non-cooperative speakers refuse to fulfill discourse obligations by violating the maxims of relevance and quantity, respectively. As an example, have a look at the following dialog snippet:

A: *Is yours just hopping along?* (QY)
B: *Hopping? How do you mean? What do you mean hopping?* (QW)

Speaker B, who is playing for time, refuses to answer A's question (R1), and does not add any new information (R2) for joint problem solving. Quantity violation is further reflected by R4 and R6 (feature *rng_c1*, cf. the line convergence in the right half of Figure 1) marking the speaker's intend to give short replies only [20]. Results R7–R9, however are harder to interpret. According to the expectations, a cooperative speaker would obey to the maxim of manner to reduce processing costs for the listener, which can be achieved by hyperarticulated speech [21]. Prosodically, this is expected to be encoded in an a reduced syllable rate, increased pitch range, more energy and a more salient marking of important information. However, we did not find a significant difference in syllable rate. R7, R8, and R9 provide further counter-evidence against this hypothesis: all these prominence features have higher values in competitive dialogs.

We thus suggest a possible alternative explanation based on different needs of mental workload [22] in cooperative and competitive settings. In the latter, workload is expected to be higher due to multiple task demands [23], i.e., eliciting information from the interlocutor and at the same time holding back known information. It has been shown by e.g. [24] that an increase in mental workload is accompanied by an increase in f0 level and energy which is in line with our findings R7 and R8. However, f0 range findings are more diverse [24] and pitch accent realization (R9) has to our knowledge not yet been covered. Nevertheless, further support for the workload hypothesis partially suggested by our prosodic findings comes from R5, i.e. larger turn-taking latencies indicating an increase in processing effort in competition.

To summarize, we identified several means by which speakers are dodging the question: by inappropriate dialog acts, by providing low amount of information, and by strong prosodic declination characteristics for early turn termination. As suggested by further prosodic characteristics, non-cooperative behavior requires an increase in mental workload.[1]

References

[1] HIRSCHBERG, J., S. BENUS, J. BRENIER, F. ENOS, S. FRIEDMAN, S. GILMAN, C. GIRAND, M. GRACIARENA, A. KATHOL, L. MICHAELIS, B. PELLOM, E. SHRIBERG, and A. STOLCKE: *Distinguishing deceptive from non-deceptive speech*. In *Proc. Interspeech*, pp. 1833–1836. Lisbon, Portugal, 2005.

[2] TRAUM, D.: *Computational models of non-cooperative dialogue*. In *Proc. LONDIAL Workshop on Semantics and Pragmatics of Dialogue*. 2008. Extended abstract.

[3] EFSTATHIOU, I. and O. LEMON: *Learning non-cooperative dialogue behaviours*. In *Proc. SIGDIAL*, pp. 60–68. Philadelphia, 2014.

[4] PLÜSS, B. and P. PIWEK: *Measuring non-cooperation in dialogue*. In *Proc. COLING*, pp. 1925–1936. Osaka, Japan, 2016.

[5] TRAUM, D. and J. ALLEN: *Discourse obligations in dialogue processing*. In *Proc. 32nd annual meeting of ACL*, pp. 1–8. Morristown, NJ, 1994.

[6] BUNT, H., V. PETUKHOVA, and A. FANG: *Revisiting the ISO standard for dialogue act annotation*. In *Proc. ISO-ACL ISA-13*, pp. 37–50. Montpellier, France, 2017.

[7] GRICE, H.: *Logic and conversation*. In P. COLE and J. MORGAN (eds.), *Speech acts*, vol. 3 of *Syntax and semantics*, pp. 41–58. Academic Press, New York, 1975.

[8] SPERBER, D. and D. WILSON: *Relevance: Communication and Cognition*. Blackwell, Oxford, 1986.

[1] The work of the first author is financed by a grant of the Alexander von Humboldt-foundation.

[9] *PAGE – Prosodic and Gestural Entrainment in Conversational Interaction across Diverse Languages*. http:page.home.amu.edu.pl, 2015.

[10] REICHEL, U., N. PÖRNER, D. NOWACK, and J. COLE: *Analysis and classification of cooperative and competitive dialogs*. In *Proc. Interspeech*, p. paper 3056. Dresden, Germany, 2015.

[11] CARLETTA, J., A. ISARD, S. ISARD, J. KOWTKO, G. DOHERTY-SNEDDON, and A. ANDERSON: *The reliability of a dialogue structure coding scheme*. *Computational Linguistics*, 23(1), pp. 13–31, 1997.

[12] HUTTO, C. and E. GILBERT: *VADER: A parsimonious rule-based model for sentiment analysis of social media text*. In *Proc. 8th International AAAI Conference on Weblogs and Social Media*, pp. 216–225. Ann Arbor, Michigan, 2014.

[13] PORTER, M.: *Snowball*. `http://snowball.tartarus.org/`, last visited: January 15th 2018.

[14] STAHL, P., P. LJUNGLOF, L. BENZAHIA, A. CHELLI, and A. ARIES: *Nltk snowball implementation*. `http://www.nltk.org/_modules/nltk/stem/snowball.html`, last visited: January 15th 2018.

[15] REICHEL, U.: *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*. RIL, MTA, Budapest, Hungary, 2018. Https://arxiv.org/abs/1612.04765.

[16] BOERSMA, P. and D. WEENINK: *PRAAT, a system for doing phonetics by computer*. Tech. Rep., Institute of Phonetic Sciences of the University of Amsterdam, 1999. 132–182.

[17] SAVITZKY, A. and M. GOLAY: *Smoothing and differentiation of data by simplified least squares procedures*. *Analytical Chemistry*, 36(8), pp. 1627–1639, 1964.

[18] REICHEL, U.: *Unsupervised extraction of prosodic structure*. In J. TROUVAIN, I. STEINER, and B. MÖBIUS (eds.), *Elektronische Sprachverarbeitung 2017*, vol. 86 of *Studientexte zur Sprachkommunikation*, pp. 262–269. TUDpress, Dresden, Germany, 2017.

[19] BENJAMINI, Y. and D. YEKUTIELI: *The control of the false discovery rate in multiple testing under dependency*. *Annals of Statistics*, 29, pp. 1165–1188, 2001.

[20] YUAN, J. and M. LIBERMAN: *F0 declination in English and Mandarin broadcast news speech*. *Speech Communication*, 65, pp. 67–74, 2014.

[21] LINDBLOM, B.: *Explaining phonetic variation: A sketch of the H&H theory*. In W. HARDCASTLE and A. MARCHAL (eds.), *Speech Production and Speech Modelling*, pp. 403–439. Kluwer Academic Publishers, Dordrecht, 1990.

[22] GOPHER, D. and R. BRAUNE: *On the psychophysics of workload: Why bother with subjective measures? Human Factors*, 26(5), pp. 519–532, 1984.

[23] WICKENS, C.: *Processing resources and attention*. In D. DAMOS (ed.), *Multiple Task Performance*, pp. 3–34. Taler&Francis, Bristol, 1991.

[24] HUTTUNEN, K., H. KERÄNEN, E. VÄYRYNEN, R. PÄÄKKÖNEN, and T. LEINO: *Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights*. *Applied Ergonomics*, 42(2), pp. 348–357, 2011.