Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Individual Differences in Top-down and Bottom-up Prominence Perception

Jason Bishop

City University of New York (College of Staten Island & The Graduate Center)
jbishop@gc.cuny.edu

Abstract

The perception of prosody, like other aspects of speech perception, relies on a combination of bottom-up and top-down information. In the context of prominence perception, the present study explored the interaction of these two types of cues, and individual variation in their effects on listeners. In a "naïve prosody transcription" task, 120 listeners gave prominence ratings to verbs and objects in simple English SVO sentences. First, a known top-down cue to perceived prominence was manipulated: the information structure (focus) status of the verb. Second, a known bottom-up cue to perceived prominence was manipulated: the phonetic duration of the verb. Results showed that both the top-down and bottom-up cues influenced perceived prominence in the expected way, but did not interact. However, both types of cues were found to be modulated by systematic cross-listener variation in cognitive processing style, as estimated by two measures believed to be related to "pragmatic skill".
Index Terms: prominence perception, top-down perception, individual differences

1. Introduction

1.1. Prominence perception

Speech perception relies on a complex combination of bottom-up (signal-based) and top-down (context and experienced-based) information. The present study investigated sources of variability in the process of perceiving prosodic prominence that relate to both types of cues. In particular, two questions were explored:

- *Do top-down and bottom up cues interact in prominence perception?*
- *Do listeners vary systematically in their use of top-down cues to prominence perception?*

These questions were asked in the context of previous work, which shows that one source of top-down information in prominence perception is a sentence's information structural interpretation. More specifically, the study explored whether the top-down effects focus has on perceived prominence interact with properties of the signal, and with properties of the listener. The goal is to understand important sources of variation in the perception of prosody.

Perceived prominence, the kind of prominence this study primarily investigates, is the subjective impression of prosodic strength that a listener experiences in some perceptually measurable way. This contrasts with acoustic prominence and structural (i.e., phonological) prominence, although perceived prominence is sensitive to both. That is, syllables or words may be experienced as prominent because they are assigned structurally prominent positions [1][2], but listeners may also respond to acoustic cues directly [3][4][5]. This distinction will become particularly important further below, where we discuss how additional elements contribute to perceived prominence, namely top-down expectations based on information structure, and systematic variation in listeners' sensitivity to that structure.

1.2. A test case: focus in English

In a previous study, [6] presented evidence that a sentence's information structural meaning can induce perceived prominence in a top-down manner. In particular, he presented simple SVO sentences like *I bought a motorcycle* in different discourse contexts, intended to either broadly focus the entire VP (1a), or narrowly focus only the object (1b). Crucially, the focus interpretation of the SVO sentences varied by context, but the acoustic information of the sentences themselves did not.

(1) a. *What did you do in Nevada? Did you gamble?*
No, I **[** climbed a mountain **]**$_{\mathbf{Foc}}$
b. *What did you do in Nevada? Did climb a hill?*
No, climbed **[** a mountain **]**$_{\mathbf{Foc}}$

The outcome of this experiment was that, when sentences were presented in broad VP focus contexts, listeners reported hearing the verbs and objects as roughly equally prominent on average. However, when the same sentences were instead presented in narrow focus contexts, the verb was perceived as less prominent, and the object as more prominent. This top-down effect can be explained as the result of listeners projecting their expectations; these expectations are based on experience with speakers, whose productions show a similar asymmetry for broad and narrow focus sentences (e.g., [7][8] [9] for English; see also [10][11][12][13] for similar patterns in other West Germanic languages). Thus listeners tend to perceive the signal as conforming to expectations more than it actually does.

1.3. Cue interaction

The finding that focus can serve as the basis for top-down illusions about prominence patterns has been reported elsewhere as well ([14] for Finnish). However, information structure has additional interesting properties whose implications for perception have not yet been explored. In particular, there is much evidence suggesting that focused information undergoes special, "deeper" processing relative to the "shallow" processing afforded to unfocused, or "given" information [15][16][17][18]. This is consistent with recent evidence that links focus interpretations to neural correlates associated with the allocation of attentional resources [19]. The additional processing functions that information structure serves are therefore important and interesting because they serve as another route through which prominence perception could be influenced. More specifically, if focus leads to more detailed processing, possibly through the modulation of

 doi: 10.21437/SpeechProsody.2016-137

attention, it may also affect listeners' sensitivity to bottom-up cues from the signal. That is, focus, a top-down cue, may influence perception via "restorative" mechanisms, as well as through what may be called "processing-based" mechanisms. Notably, multifaceted top-down effects on perception have been reported elsewhere with their origins in rhythmic expectations [20][21]; one of the main questions of this study was whether focus has a similarly dual influence on prominence perception.

1.4. Individual differences

Another important property of information structural interpretation is that it inherently depends on two things: attending to a larger discourse context, and relating a target utterance to that context. Importantly, there is evidence that the extent to which individuals engage in this sort of processing may be subject to considerable variation, some of which seems to be predicted by measures of "pragmatic skill", an aspect of "cognitive processing style" [22]. For example, one measure arguably related to pragmatic skill, the communication subscale of the Autism Spectrum Quotient [23], has been shown to predict individual differences in the use of pragmatic information in sentence [24][25] and lexical processing [26]. The implications for the present study are clear: if listeners do not attend to and integrate discourse context into their comprehension of sentences equally, top-down effects of information structure like the ones of interest here should also not be observed equally. To understand how prominence perception takes place, then—specifically the top-down aspect of it—we also need to understand how this information is dependent on properties of the listener.

1.5. Present Study

The goal of the present experiment was to further explore the focus effect on prominence perception, using it to test two things. First, does information structure (i.e., focus) predict attention to bottom-up acoustic cues to prominence? What we would expect, based on previous work, is that the sensitivity of perceived prominence to small acoustic differences should be higher for focused information than for unfocused information. This was tested by manipulating the acoustic duration of verbs in SVO sentences, like the one in (1), above, and presenting it in contexts where it was either focused (broad VP focus question contexts) or unfocused (narrow object focus question contexts). Importantly, acoustic and structural prominence were held constant, since (a) the recordings of SVO sentences did not vary across information structural contexts, and (b) both the verb and the object in the sentences were always both structurally prominent (i.e., pitch accented). Listeners' impressions of acoustic prominence were expected to be more sensitive to this difference when the verb was part of the focus constituent.

The second goal was to test whether listeners with poorer pragmatic skills showed weaker top-down effects from information structure. If so, there are two possible ways this could happen, based on the discussion above. First, listeners may show a weaker "boost" in perceived prominence as a result of focus. That is, the effect of focus reported in [6] would apply less uniformly across listeners, being systematically weaker in individuals with poorer pragmatic skills. Second, if focus also increases attention/sensitivity to changes in acoustic information, here duration, listeners with poorer pragmatic skills would benefit less from this. The experiment below utilized two separate measures of pragmatic skill used previously in the literature to explore these predictions: scores on the communication subscale of the AQ and scores on the Reading the Eyes in the Mind Test [27], which are both described further below.

2. Experiment

2.1. Methods

2.1.1. Materials

Materials were designed for a prominence rating experiment similar to that in [6], in which verbs and objects served as target words, appearing in sentences that themselves occurred as answers to contrastive-focus WH-questions, as in example (1), above. These were 20 SVO sentences and 40 contexts (20 broad focus-inducing, 20 narrow focus-inducing). The SVO sentences were produced by a male native speaker of American English, trained in intonational phonology, with a slightly falling "hat pattern"; the sentences were all produced as a single prosodic phrase (marked by a L-L%) with a pitch accent on the verb and on the object. The (prenuclear) verb was pronounced with a H* and the (nuclear) object was intentionally pronounced with a pitch accent that was phonetically ambiguous between a !H* and a H* in the ToBI annotation system [28]. Additionally, because the SVO sentences would occur as corrective answers to the WH-questions, the speaker started the utterance with the word "no", pronounced in a separate prosodic phrase from the SVO sentence. An example is shown in Fig 1.

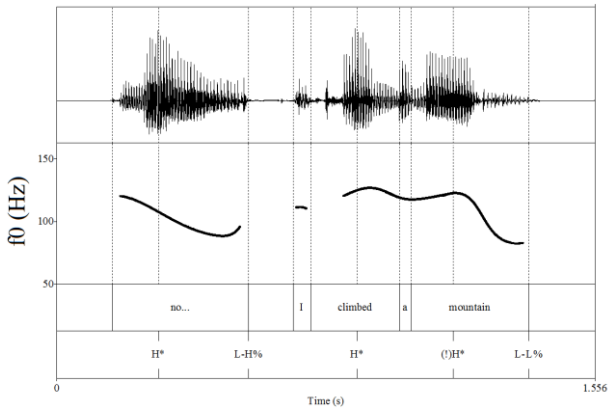

Fig 1. *Example pitch track and waveform for the SVO test sentence "I climbed a mountain."*

Two additional types of conditions were then created, intended to manipulate the acoustic prominence and the information structural status for verbs in the test sentences. First, verbs were manipulated using PSOLA resynthesis in Praat [29] to create a 5-step continuum; these steps were made by either increasing or decreasing the duration from the original recordings, resulting in step sizes ranging from 70% to 130% the of the verb's original duration. Second, lead-in sentences like the ones in example (1), produced by a female English speaker, were used to manipulate whether the verb in the test sentences would be part of the focus constituent.

Finally, 30 additional sentences (by the same two speakers) were also recorded to be used as fillers, although no durational manipulation was carried out for fillers. To increase variability in prominence patterns throughout the course of the experiment (and thus the unpredictability/potential informativeness of both the information structures and the

prominence patterns), the fillers varied the location of the nuclear pitch accent and focus structure. Half of the filler trials contained focus questions that required a nuclear accented subject noun in the answer (e.g., *Who noticed the mice? Did Natasha notice them? No, John noticed the mice.)* or a nuclear accented verb (e.g., *What did you do with your property? Did you rent it out? No, I sold my property.*).

2.1.2. Participants

120 native English-speakers served as listeners in the experiment. For technical reasons, sixty participated at a university on the West Coast of the United States, and 60 at a university on the East Coast of the United States. All participants received course credit or monetary compensation.

2.1.3. Procedure

Listeners participated in a prominence rating task like the one used in [6]; listeners were randomly assigned to one of 6 pseudorandomized orderings of 10 lists that balanced the test sentences across the information structural and durational conditions. Participants listened to each of the question-answer exchanges as many times as they wished, but were discouraged from listening more than three of four times to an item. As they played each question-answer exchange, presented as a numbered "play" button on a computer screen, they were to listen for how "stressed" words in the male speaker's SVO answers sounded. Instructions emphasized to participants that their task was to listen to how his answer sentences were pronounced, in particular how the speaker used his voice to make words stand out from other words. Participants followed along on a printed transcript of the dialogues and provided ratings of from 1 ("not at all stressed") to 5 ("very stressed") for all verbs and objects in the sentences (which were underlined on the transcripts). This produced 4,800 prominence ratings from test sentences (120 listeners × 20 test sentences × 2 words (verbs & objects), i.e., 480 ratings for each durational step in each of the two focus conditions.

In addition to the main experimental task, listeners also completed two measures of cognitive processing style considered to estimate pragmatic skill. The first was the AQ [23]. The AQ consists of 50 items measuring autistic-like personality traits along five dimensions: *Social Skills, Attention to Detail, Attention Switching, Imagination,* and *Communication.* Participants provide "*definitely agree*", "*slightly agree*", "*slightly disagree*" or "*definitely disagree*" responses to statements such as ''*I frequently find that I don't know how to keep a conversation going",* and receive a point for each "autistic-like" response (e.g., a "*definitely agree*" or "*slightly agree*" response would earn a point in the example just given), and thus higher scores indicate poorer, more autistic-like communicative/pragmatic skill. Participants here completed the full questionnaire, although only the communication subscale (henceforth AQ-Comm) scores were of interest, since (as discussed above), this has been most closely linked to pragmatic skill in language processing tasks.

The second measure used was The Reading the Mind in the Eyes test [27] (henceforth "EYES"), a 36-item questionnaire in which the participant is presented with photographs of various human faces. However, only the region around the person's eyes is visible, and the participant must choose from a multiple choice list of four adjectives that best describes the emotional state of the person in the picture. This measure is related to pragmatic skill in that it was designed by the authors to estimate neurotypical adults' "Theory of Mind" abilities—i.e., the ability reconstruct the mental state of another person [30], and higher scores indicate better Theory of Mind/pragmatic skill. The EYES test was included as it potentially provides a measure of pragmatic skill that relies less on the self-report introspection inherent to personality inventory-type instruments like the AQ. The entire experimental session took approximately 50 minutes.

2.2. Analysis

It was necessary to drop four participants from the study, as three were later discovered to not meet the native-speaker requirement, and one was discovered to have a history of auditory processing disorder. Numerical prominence ratings for the remaining participants were to serve as the outcome variable in mixed-effects linear models. The present analysis was limited to verbs, which were the object of both the durational and information structural manipulation. An initial round of modeling included as fixed effects the stimulus-level and listener-level predictors of interest, as well as trial and dialect, although there were no predictions regarding the two regional varieties. Maximal random-effects structure was used [31]. From the initial models, random-effects were retained if doing so did not decrease the fit of the model; fixed effects terms with a large p-value ($p > .1$) were removed if doing so did not result in a decrease to model fit. In the final model, below, the following random-effects parameters were retained: random intercepts for listener and item, by-participant random slopes for trial, and by-item random slopes for focus.

3. Results

First, and unsurprisingly, it was found that verbs in longer durational conditions were rated as more prominent by listeners overall, indicated by a simple effect for verb duration. Second, there was a simple effect for focus, such that verbs that were focused (i.e., presented in broad VP focus contexts) were rated as more prominent, replicating a previous finding [6]. Interestingly, however, both of these simple effects were modulated by individual differences, although based on different measures. First, sensitivity to the durational manipulation of verbs was positively related to EYES scores (indicating that better pragmatic skill on this measure predicted more consistent sensitivity to changes in verb duration), indicated by a significant interaction between verb duration and EYES. There was also a simple effect for EYES, such that higher scores were associated with lower ratings overall, an unpredicted and mysterious overall effect. Both EYES-related effects are apparent in Fig 2. Second, listeners with higher AQ-Comm scores (indicating poorer pragmatic ability) showed a weaker effect for focus, as indicated by a significant interaction between AQ-Comm and focus (Fig.3). The model output is shown in Table 1; absent from the model, because it did not contribute to its fit, was a focus × verb duration interaction, indicating that focus did not modulate listeners' sensitivity to the durational manipulation.

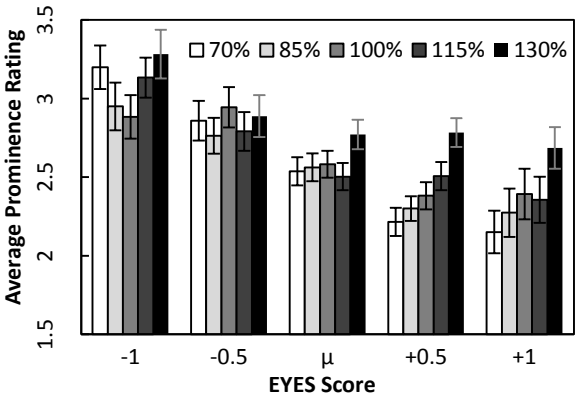

Fig 2. *Effect of the durational manipulation of verbs as a function of listeners' Reading the Eyes in the Mind scores (higher scores indicate better pragmatic skill). Listeners are grouped in terms of standard deviation around the mean score, although this variable was continuous in the statistical model). Error bars show standard error.*

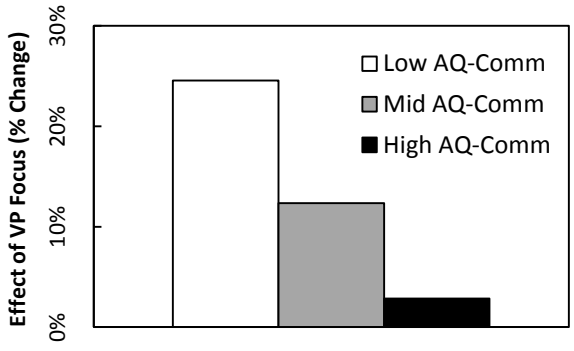

Fig 3. *The size of the top-down focus effect as a function of AQ-Comm scores. (Higher scores indicate poorer pragmatic skill). Bars show the percentage change in prominence ratings for verbs going from narrow focus to broad focus. ("Mid AQ-Comm" represents listeners with scores with 1 s.d. around the mean; "Low" and "High" above or below that, although AQ-Comm was a continuous variable in the statistical model).*

Table 1. *Results for fixed-effects factors in the regression model.*

	β	*SE*	*t*	*p*
(Intercept)	-.014	.103	-0.14	> .1
Trial	.003	.002	1.86	< .1
Dialect (West Coast)	-.139	.123	-1.13	> .1
Vowel Duration	.072	.012	6.05	< **.001**
Focus (VP)	.150	.042	3.57	< **.001**
AQ-Comm	.008	.027	0.32	> .1
EYES	-.056	.016	-3.45	< **.001**
Focus (VP) * AQ-Comm	-.037	.015	-2.48	< **.05**
Vowel Duration * EYES	.009	.003	2.89	< **.01**

4. Discussion and Conclusion

This study was intended to investigate two basic questions, both having to do with the role of information structure in guiding prominence perception. First, it was asked whether focus had the effect of modulating attention to a subtle bottom-up cue, namely duration. The second question was whether a (previously reported) effect of focus was itself modulated by individual differences in pragmatic skill. The answer to the first question appears to be negative, since no interaction between the presence of focus and the durational manipulation of verbs was found. The answer to the second question, however, was positive, since the expected top-down effect of focus on verbs (which increased their perceived prominence) depended strongly on listeners' individual scores on one of the measures of pragmatic skill, namely the communication subscale of the AQ. Finally, and somewhat expectedly, the other measure of pragmatic skill tested, the EYES test, was found to predict some cross-listener variation in sensitivity to the durational manipulation. Thus different measures of pragmatic skill were found to be related to sensitivity to different cues to prominence perception, suggesting the two measures capture something different about cognitive processing style, which has been suggested previously elsewhere [32].

One possibility is that perceived prominence is subject to cross-listener variation in cue weighting [33], and that some of this variation correlates with different aspects of cognitive processing style. Some evidence for this comes from the fact that AQ-Comm, which was not correlated with sensitivity to duration in the present study, may be related more to attention to fundamental frequency/pitch. In a recent study of prominence perception for intonational pitch accents in English, [34] found that perceived prominence for pitch accents of different types was somewhat predictable based on AQ-Comm; in particular, people with very low AQ-Comm scores (indicating above-average pragmatic skill) showed a prominence ranking of L* < !H* < H* < L+H*, but this perceptual separation of pitch accent types became weaker as AQ-Comm scores increased. This finding is broadly consistent with other studies showing a relation between pitch prominence and AQ-Comm [35]. Possibly, then, AQ-Comm correlates with some aspect of listeners that relates to pitch processing or perception, while EYES reflects some underlying sensitivity to metrical/durational cues. If this is true, individual differences in pragmatic skill may help to identify the source of variation in a recent study by [36], who also report cross-listener differences in memory for acoustic cues to pitch accents in English. Clearly predicting attention to acoustic cues to perceived prominence at the level of the individual listener requires further study; the present results, however, suggest that at least some of this variation is systematically tied to cognitive processing style, and so an individual differences approach may be a useful methodology to incorporate into future work.

5. Acknowledgements

The author wishes to thank members of the UCLA Phonetics Laboratory, as well Hannah Kim, Katie Brown, and Amanda Marshall for help with data collection. I am also grateful for helpful comments from two Speech Prosody 2016 reviewers.

6. References

[1] Turnbull, R. submitted. The role of predictability in intonational variability. Ms. The Ohio State University.

[2] Jagdfeld, N., Baumann, S. 2011. Order effects on the perception of relative prominence. *Proceedings 17th ICPhS*, 958-961.

[3] Gussenhoven, C., Rietveld, A. Rump, H., Terken, J. 1997. The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America* 102 (5). 3009- 3022.

[4] Kochanski, G., Grabe, J., Coleman, J., Rosner, B. 2005. Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118 (2), 1038–1054.

[5] Cole, J., Mo, Y., Hasegawa-Johnson, M. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1, 425–452.
[6] Bishop, J. 2012. Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (eds.) *Prosody and Meaning (Interface Explorations)*, 239 – 270. Berlin: Mouton de Gruyter.
[7] Breen, M., Fedorenko, E., Wagner, M., Gibson, E. 2010. Acoustic correlates of information structure. *Language & Cognitive Processes*, 25 (7), 1044-1098.
[8] Jun, S.-A. 2008. Focus: domains, types, and realizations. Talk given at the Yale Linguistics Department Colloquium Series, Yale University, New Haven, CT.
[9] Sityaev, D., House, J. 2003. Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proceedings of the 15th ICPhS*, 1819-1822.
[10] Baumann, S., Grice, M., Steindamm, S. 2006. Prosodic marking of focus domains: categorical or gradient? *Proceedings of Speech Prosody 2006*, 301-304.
[11] Hanssen, J., Peters, J., Gussenhoven, C. 2008. Prosodic effects of focus in Dutch declaratives. *Proceedings of Speech Prosody 2008*, 609-612.
[12] Féry, C., Kügler, F. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36, 680 - 703.
[13] Peters, J., Hanssen, J., Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics* 46: 185 – 209.
[14] Vainio, M., Järvikivi, J. 2006. Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics* 34, 319-342.
[15] Cutler, A., Fodor, J. 1987. Semantic focus and sentence comprehension. *Cognition* 7, 49 – 59.
[16] Blutner, R., Sommer, R. 1988. Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language* 27, 359 – 367.
[17] Sanford, A., Sturt, P. 2002. Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences* 6, 382 – 386.
[18] Sanford, A., Sanford, A., Molle, J., Emmott, K. 2006. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes* 42, 109 – 130.
[19] Kristensen, L., Wang, L., Petersson, K., Hagoort, P. 2012. The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex* 23(8),1836 – 1848.
[20] Breen, M., Clifton, C. 2011. Stress Matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, 64 (2), 153-170.
[21] Zheng, X. Pierrehumbert, J. 2010. The effects of prosodic prominence and position on duration perception. *Journal of the Acoustical Society of America*, 128(2), 851 – 859.
[22] Ausburn, L., Ausburn, F. 1978. Cognitive styles: Some information and implications for instructional design. *Educational Communication & Technology* 26, 337–354.
[23] Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. 2001. The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5-17.
[24] Nieuwland, M., Ditman, T., Kuperberg, G. 2010. On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language* 63, 324 – 346.
[25] Xiang, M., Grove, J. Giannakidou, J. 2013. Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*. 4:708. doi: 10.3389/fpsyg.2013.00708.
[26] Bishop, J. 2012. Focus, prosody, and individual differences in "autistic" traits: Evidence from cross-modal semantic priming. *UCLA Working Papers in Phonetics* 111, 1 – 26. See also Bishop, J. under revision. Focus Projection and prenuclear accents: evidence from lexical processing.
[27] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I. 2001b. The "Reading the Mind in the Eyes" test Revised Version: A study with normal adults, and adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry* 42 (2). 241 – 251.
[28] Beckman, M., Ayers-Elam, G. 1997. Guidelines for ToBI labeling (version 3.0, March 1997). Ms. The Ohio State University Research Foundation.
[29] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Version 6.0.05, www.praat.org.
[30] Premack, David & Guy Woodruff. 1978. Does the chimpanzee have a "theory of mind"? *Behaviour and Brain Sciences*, 4, 515-526.
[31] Barr, D., Levy, R., Scheepers, C., Tily, H. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3), 255 – 278.
[32] Turnbull, R. 2015. Assessing the listener-oriented account of predictability-based phonetic reduction. Ph.D. dissertation, The Ohio State University.
[33] Watson, D. 2010. The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.) *The Psychology of Learning and Motivation*, Vol. 52, 163 – 183. Elsevier.
[34] Colon, J. & Bishop, J. 2015. Autistic traits predict prosody perception in neurotypical adults. Presented at the 27th Meeting of the Association for Psychological Science, New York City.
[35] Jun, S.-A., Bishop, J. 2015. Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier, E. Gibson (Eds), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor*. Springer.
[36] Kimball, A., Cole, J., Dell, G., Shattuck-Hufnagel, S. 2015. Categorical vs. episodic memory or pitch accents in English. *Proceedings of the 18th ICPhS*, 1 – 4.