10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

# The Relationship between Prosodic Ability and Conversational Prosodic Entrainment

*Heike Lehnert-LeHouillier*[1], *Susana Terrazas*[2], *Steven Sandoval*[2], *Rachel Boren*[3]

[1]Department of Communication Disorders, New Mexico State University, USA
[2]Klipsch School of Electrical and Computer Engineering, New Mexico State University, USA
[3]College of Education, New Mexico State University, USA

hlehnert@nmsu.edu, selisath@nmsu.edu, spsandov@nmsu.edu, rboren@nmsu.edu

## Abstract

Conversational entrainment or alignment—the convergence of conversation partners over the course of a conversation in a variety of linguistic features—-is a well-attested conversational phenomenon. The research on prosodic entrainment has shown correlations between prosodic entrainment and several social dimensions of rapport between conversation partners. However, little is known about how skill-level in the entrainment domain affects the ability to converge during a conversation.

The goal of the current study was to investigate whether skill-level of a speaker in receptive and expressive word, sentence, and emotional prosody is correlated with the amount of prosodic entrainment contributed at the conversational level. Twenty native speakers of American English were paired into ten dyads of seven female/female and three female/male conversation pairs. Conversations for each pair were recorded and analyzed. Test scores measuring word, sentence, and emotional prosody were correlated with the amount of fundamental frequency entrainment during conversations.

The results indicate that a negative correlation exists between expressive prosody skill and the amount of f0 entrainment contributed by a speaker. This suggests that speakers with better expressive prosodic skills at the word and sentence level entrain less to their conversation partners. Receptive prosody ability was not correlated with conversational prosodic entrainment.

**Index Terms**: expressive prosody, receptive prosody, human-human interaction, prosodic entrainment

## 1. Introduction

Prosodic entrainment refers to the process by which the prosodic features of two or more conversational partners become more similar over the course of a conversation. The prosodic features that have been found to entrain include fundamental frequency [1, 2], speech rate [3, 4], vocal intensity [3, 5], pause duration [6], and paralinguistic properties related to prosody such as voice quality [2, 7]. The entrainment of these and similar prosodic features has been found to be mediated by a variety of non-linguistic factors. For example, Michalsky and Schoomann [8] found that pitch convergence correlated with perceived attractiveness and likability of conversation partners during dating conversations. Similarly, Ireland et al. [9] found that the amount of entrainment both predicted whether conversation partners were going to initiate a romantic relationship as well as correlate with the stability of an existing relationship. The amount of prosodic entrainment has also been shown to correlate with perceived conversational quality [10] and immediate conversational success in goal-oriented conversations [11]. Along similar lines as [11], a study by Niebuhr and Michalsky [12] found that prosodic entrainment measures can predict features of a more general ability to engage in teamwork, including the quality of the final product created by a team and time-management skills.

Although prosodic entrainment frequently occurs during conversations, there is much individual variability. Variability exists in whether or not conversational partners show entrainment of the different linguistic and paralinguistic features, as well as in the amount of entrainment shown by each individual conversation partner. While some variability is to be expected because entrainment is a phenomenon mediated by a variety of social and contextual factors, the differences between individuals suggests that (in addition to the established social and contextual factors) characteristics of the individual speaker are also important for our understanding of conversational entrainment. However, little is known about which characteristics tied to the individual speaker are responsible for how much a speaker entrains to his or her conversation partner. In particular, whether and how the skill-level of a speaker in the entrainment domain affects prosodic entrainment, has not received much attention. One study by Lewandowski and Jilka [13] investigated the role of phonetic talent of speakers on phonetic entrainment. It was found that those second language learners who showed more phonetic talent also exhibited more phonetic entrainement to their conversation partners compared to those who were less phonetically talented.

To better understand the different contributions of social context and individual speaker characteristics to conversational entrainment, it is of interest to determine whether or not Lewandowski and Jilka's [13] findings generalize to all entrainment domains - the phonetic and the prosodic domain - or whether the particular contributions play out differently in each domain. In other words, is it universally the case that high skill level in a given domain is correlated with more entrainment in that domain? In the current study, we address this question by investigating whether prosodic skill level of a speaker predicts how much she or he entrains to a conversation partner.

## 2. Method

### 2.1. Participants & Procedure

To investigate whether and how prosodic skill level correlates with the amount of prosodic entrainment at the conversation level, twenty native speakers of American English (seventeen female and three male) participated in the current study. All participants were students at New Mexico State University who received course credit in exchange for their participation in the study. All twenty participants reported American English as


10.21437/SpeechProsody.2020-157

their native language and passed a hearing screening to assure that scores on receptive prosody tests were not biased by hearing status of the participants. Prosodic skill level was assessed using the Profiling Elements of Prosody in Speech Communication assessment (PEPS-C) [14]. All participants were administered this standardized prosody assessment. The PEPS-C assesses prosodic ability via fourteen sub-tests, seven expressive sub-tests and seven corresponding receptive sub-tests. The seven prosodic skills assessed are: 1) basic imitation and discrimination of prosodic pattern (Imitation/Discrimination subtests), 2) production and perception of statement versus question intonation (Turn end sub-tests), 3) production and perception of prosody related to liking and disliking (Affect subtests), 4) production and perception of lexical stress (Lexical Stress sub-tests), 5) production and perception of phrasal prominence (Phrasal Stress sub-tests), 6) production and perception of boundary intonation related list intonation (Chunking subtests), and 7) production and perception of sentence focus (Focus sub-tests). For a detailed description of the tasks and rating protocols of the PEPS-C test please refer to [14]. After prosodic skill was assessed, the participants were paired into ten dyads of seven female/female and three female/male conversation pairs. Each pair was then asked to engage in the goal-oriented conversational Diapix task [15]. The conversations were conducted in a sound treated room and digitally recorded onto a PC in waveform audiofile format using a Shure SM48-LC condenser microphone. The recordings were then annotated using Praat [16] TextGrid files such that the utterances for each speaker were marked up.

### 2.2. Fundamental Frequency Entrainment Measures

In order to assess fundamental frequency (f0) entrainment over the course of the conversation, we evaluate two entrainment measures, one based on mean fundamental frequency and the other based on the interquartile range of the fundamental frequency. These methodologies for assessing (f0) entrainment are frequently used in the literature. Our calculations were most similar to those described by [8] who assessed entrainment levels in a speed dating environment. In addition to estimating the level of entrainment, we extend the procedure in [8] to allow an estimate of each speakers's contribution to the global entrainment amount of the conversation.

#### *2.2.1. Mean Fundamental Frequency Entrainment*

To assess whether or not speakers entrained to each other over the course of the conversation, the mean f0 for each of the two speakers engaged in a conversation was determined throughout the first and the last third of the conversation. We denote the mean f0 during the first third of the conversation as $\mathrm{S1_s}$, and $\mathrm{S2_s}$ for speakers one and two, respectively. Similarly, we denote the mean f0 during the last third of the conversation as $\mathrm{S1_e}$, and $\mathrm{S2_e}$ for speakers one and two, respectively. Next, a vector corresponding to the change in the mean f0 per dyad was determined. This vector describes the distance and direction of the f0 change from first to last third of the conversation based on the starting point $S$, where $S = (\mathrm{S1_s}, \mathrm{S2_s})$, and the ending point $E$, where $E = (\mathrm{S1_e}, \mathrm{S2_e})$. A line $\ell$ corresponding to where the two speakers have the same (matching) mean f0 is also defined. The minimum distance of the starting point $S$ to the line $\ell$ was computed and denoted as $d_1$. Similarly, the minimum distance of the ending point $E$ to the line $\ell$ was computed and denoted as $d_2$. For the purpose of this study, the entrainment level was measured as the difference between the minimum distances

$$\Delta\mathrm{Ent} = d_1 - d_2 \tag{1}$$

Thus, a positive value of $\Delta\mathrm{Ent}$ indicates that the two speakers exhibit an increase in f0 entrainment over the course of the conversation while a negative value of $\Delta\mathrm{Ent}$ is indicative of a decrease in f0 entrainment.

After the entrainment measure $\Delta\mathrm{Ent}$ was calculated, the percentage of entrainment contribution by each individual speaker is estimated by defining two contribution measures. Specifically, the first and second speakers' contribution was defined as

$$\mathrm{S1_{resp}} = \frac{d_3}{d_3 + d_4} \quad \text{and} \quad \mathrm{S2_{resp}} = \frac{d_4}{d_3 + d_4} \tag{2}$$

where $d_3 = |\mathrm{S1_e} - \mathrm{S1_s}|$ and $d_4 = |\mathrm{S2_e} - \mathrm{S2_s}|$. The procedure used to determine f0 entrainment is summarized in Figure 1. Finally, the results were then correlated with each speaker's PEPS-C assessment scores as described in the next section.

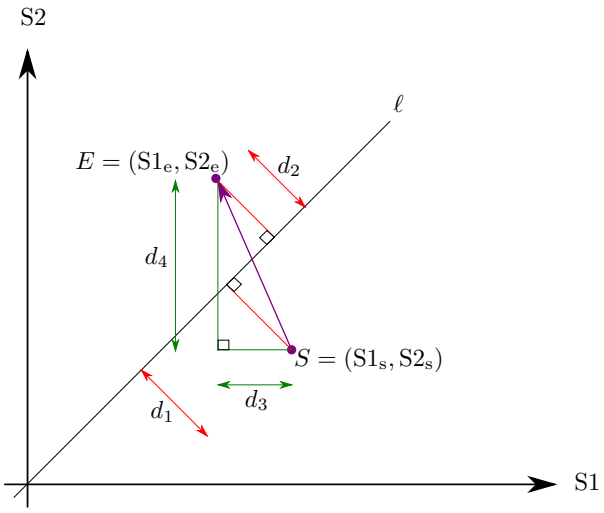


Figure 1: *Illustration of the points, vector, and distances involved in the calculation of the* f0 *entrainment measures.*

#### *2.2.2. Fundamental Frequency Range Entrainment*

In addition to entrainment in terms of mean (f0), we also assess whether or not speakers entrained to each other over the course of the conversation in terms of changes in (f0) range. In order to do so, we compared the interquartile range of the fundamental frequency for each of the two speakers throughout the first and the last third of the conversation. The same general procedure for measuring entrainment level which was used for the the mean fundamental frequency was used for comparison of interquartile range of the fundamental frequency. More specifically, rather than using $\mathrm{S1_s}$, $\mathrm{S2_s}$, $\mathrm{S1_e}$, and $\mathrm{S2_e}$ to represent mean fundamental frequencies, we use $\mathrm{S1_s}$ and $\mathrm{S2_s}$ to represent the interquartile ranges of the fundamental frequency during the first third of the conversation, and $\mathrm{S1_e}$ and $\mathrm{S2_e}$ to represent the interquartile ranges of the fundamental frequency during the last third of the conversation. By reusing this notation, we may proceed to estimate entrainment level using interquartile range by calculating $\Delta\mathrm{Ent}$ using (1) and the individual speaker responsibilities using (2) as previously described for the mean fundamental frequency.

### 2.3. Statistical Analysis

In order to test whether prosodic ability, as assessed by the PEPS-C test, is correlated with prosodic entrainment performance during a goal-oriented conversational task, a series of bivariate correlations were performed between the expressive and receptive sub-test scores on the PEPS-C and percent entrainment contribution for each speaker in terms of mean f0 and f0 range entrainment. Only speakers who were part of a conversational dyad which did show f0 entrainment over the course of the conversation were included. Out of the ten dyads, eight exhibited mean f0 entrainment and nine exhibited f0 range entrainment. Therefore, sixteen speakers (two speakers in each entraining dyad) were included in the statistical analysis assessing correlation in mean f0 entrainment and eighteen speaker were included in the f0 range entrainment analysis.

The bivariate correlations were performed on the overall expressive and receptive scores, which were composed of the sum of sub-scores on the seven expressive and the seven receptive tasks, respectively. In addition to the overall expressive and receptive scores, bivariate correlations were also performed on all fourteen sub-test scores. To account for the fact that some dyads showed more overall f0 entrainment than others, an adjusted entrainment contribution measure was created. This measure normalizes the individual entrainment contribution based on the overall amount of entrainment exhibited by the conversational dyad during the conversation. This adjusted entrainment contribution measure was used for all bivariate correlations in the correlation matrix.

## 3. Results

The results of the statistical analysis reveal a significant negative correlation between the expressive prosody sub-score of the PEPS-C and the adjusted entrainment contribution measure for both mean f0 entrainment ($r = -0.558$, $\rho = 0.025$) and f0 range entrainment ($r = -0.683$, $\rho = 0.002$). This indicates that those speakers who performed lower on measures of expressive prosody in fact tended to entrain more to their conversational partners during this goal-oriented conversational task. Conversely, those speakers who did better on the expressive prosody tasks on the PEPS-C entrained less to their conversation partners. No correlation between any of either of the two f0 entrainment variables and the receptive prosody scores were found. The results of the bivariate correlation analysis for all the sub-scores and individual task scores with the adjusted mean f0 and f0 range entrainment measures are summarized in Table 1. The negative correlation between expressive prosodic ability scores and conversational f0 entrainment is more pronounced in the f0 range entrainment measure compared to the mean f0 entrainment measure. This difference is illustrated in Figure 2.

When considering the individual scores on the seven expressive sub-tests that yield the expressive sub-score, it becomes apparent that the negative correlation between the expressive sub-score on the PEPS-C test and f0 entrainment is primarily due to a small sub-set of the expressive scores. For mean f0 entrainment, the sub-test measuring the ability to produce lexical stress, is predominantly responsible for the negative correlation. The expressive lexical stress sub-test score is the only expressive sub-test score that is also significantly correlated with the adjusted mean f0 entrainment measure ($r = -0.671$, $\rho = 0.004$). This suggests that those speakers who expressively performed better on tasks that required them to produce word stress in words that are solely distinguishable by lexical stress, like *produce* (noun) versus *produce* (verb), entrained less in terms of mean f0 to their conversation partners than those speakers who performed not as well on the lexical stress task.

Table 1: *Pearson's Correlation Coefficient $r$ and $\rho$-value (2-tailed) between PEPS-C sub-test scores and the adjusted entrainment measures for mean* f0 *and* f0 *range. Statistically significant correlations are in boldface.*

| | **Mean f0** | | **f0 Range** | |
|---|---|---|---|---|
| **Description** | $r$ | $\rho$ | $r$ | $\rho$ |
| **Total Expressive** | **-0.558**$^{*}$ | **0.025** | **-0.683**$^{**}$ | **0.006** |
| Imitation | -0.098 | 0.720 | -0.047 | 0.850 |
| Turn end - Expressive | -0.042 | 0.880 | -0.115 | 0.650 |
| Affect - Expressive | -0.094 | 0.730 | -0.251 | 0.310 |
| Lexical Stress - Exp. | **-0.671**$^{**}$ | **0.004** | **-0.539**$^{*}$ | **0.021** |
| Phrasal Stress - Exp. | -0.420 | 0.110 | **-0.477**$^{*}$ | **0.045** |
| Chunking - Expressive | -0.287 | 0.280 | **-0.523**$^{*}$ | **0.026** |
| Focus - Expressive | -0.345 | 0.190 | **-0.674**$^{**}$ | **0.002** |
| **Total Receptive** | -0.191 | 0.500 | -0.187 | 0.460 |
| Discrimination | -0.144 | 0.590 | -0.190 | 0.450 |
| Turn end - Receptive | 0.249 | 0.350 | 0.165 | 0.510 |
| Affect - Receptive | 0.031 | 0.910 | -0.024 | 0.930 |
| Lexical Stress - Rec. | -0.297 | 0.260 | -0.042 | 0.870 |
| Phrasal Stress - Rec. | -0.105 | 0.700 | -0.173 | 0.490 |
| Chunking - Receptive | 0.081 | 0.770 | 0.000 | 0.990 |
| Focus - Receptive | -0.254 | 0.340 | -0.402 | 0.090 |

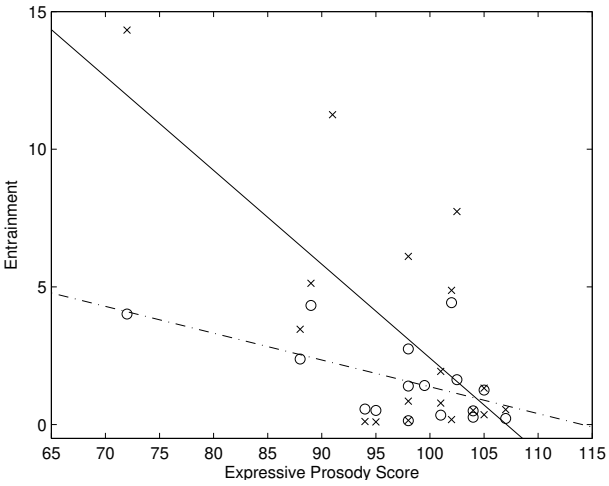


Figure 2: *Negative correlation between expressive prosody ability as measured by the expressive sub-score on the PEPS-C test and mean* f0 *entrainment (dotted line) as well as* f0 *range entrainment (solid line) during a goal-oriented conversation task.*

As illustrated in Figure 2, entrainment in f0 range showed an even stronger negative correlation with the expressive sub-score on the PEPS-C. This can be explained by the fact that several of the seven expressive sub-test scores are negatively correlated with f0 range entrainment. In addition to a negative correlation with the lexical stress sub-test score ($r = -0.539$, $\rho = 0.021$) that was also observed for mean f0 entrainment, the expressive phrasal stress sub-test score ($r = -0.477$, $\rho = 0.045$), the sub-test score for boundary expression ($r = -0.523$, $\rho = 0.026$), and the contrastive stress expression sub-test score ($r = -0.674$, $\rho = 0.002$) were all negatively correlated with f0 range entrainment contribution.

## 4. Discussion

The current study set out to investigate whether prosodic ability of a speaker at the word and sentence level correlates with the amount of prosodic entrainment that a speaker contributes at the conversational level. The results suggest that expressive prosodic performance at the word and sentence level does negatively correlate with the amount of prosodic entrainment contributed at the conversational level. Furthermore, the results suggest that the amount of conversational prosodic entrainment a speaker contributes is independent of how well he or she is able to perceive word and sentence prosody.

The two f0 measures used to assess prosodic entrainment in the current study - mean f0 and f0 range - were both negatively correlated with the expressive sub-score on the Profiling Elements of Prosody in Speech Communication (PEPS-C) assessment. This seems to suggest that those speakers who are more proficient at expressing word and sentence prosody are less likely to use prosodic features, in particular f0, to entrain to their conversation partners during a conversation. However, only a sub-set of word and sentence level expresssive sub-tests on the PEPS-C showed this negative correlation with prosodic entrainment at the conversational level. The expressive sub-test that was negatively correlated with both mean f0 and f0 range entrainment was the lexical stress sub-test. The tasks on this sub-test require the participants to expressively produce lexical stress. Although differences in word stress in English are expressed via multiple acoustic features, including f0, vowel quality, intensity, and syllable duration, pitch differences have long been reported to be the acoustic cue that listeners tend to rely on the most [17]. Similarly, the expression of phrasal stress in pairs like *blackbird* versus *black bird*, the expression of phrasal boundaries in pairs like *chicken-fingers and fruit* versus *chicken, fingers and fruit*, as well as the expression of contrastive stress in pairs like *green COW* as opposed to *GREEN cow* all are perceived by relying on pitch differences.

The negative correlation between word and sentence prosodic skill and conversational f0 entrainment might seem unexpected, given that studies like the one by Lewandowski and Jilka [13] found that better phonetic ability was positively correlated with more phonetic entrainment, while our study suggests that better expressive prosodic ability is correlated with less prosodic entrainment. Several possibilities exist to explain this difference. It is possible that the observed significant negative correlation and/or lack of other significant correlations is due to the limitations of this study. Perhaps, a larger sample size could yield different results. It is also possible that the entrainment behavior observed during one conversational task is not sufficient to deduce overall ability to entrain to a conversational partner. Perhaps multiple observations of entrainment behavior with a variety of conversation partners would provide different results.

However, should the observed pattern hold beyond the above mentioned limitations, it is conceivable that the results reflect individual differences in the expressive use of fundamental frequency. Fundamental frequency serves many prosodic and para-linguistic functions [18]. It is known to correlate with prosodic prominence at the word, phrase, and sentence level. It is also known to be a prosodic feature that entrains during conversations, as well as to serve a para-linguistic purpose when used to express emotional prosody. Given this multitude of uses of f0 for prosodic purposes at different levels of linguistic organization, some speakers may expressively use f0 more so to convey lexical stress and less to express social-pragmatic features as observed in conversational entrainment. On the other hand, others may choose to dedicate more of their expressive f0 capacity to conversational entrainment while using other correlates of word and sentence prosody, such as duration, intensity, or vowel quality (among others) to express word prominence.

## 5. Conclusions

The results of the current study suggest that better skill or performance in the linguistic domain of conversational entrainment - in this case prosody - does not necessarily translate to better or more conversational entrainment in that domain. In fact, in some instances, namely when it comes to lexical, phrasal and contrastive stress production abilities, better skill (as determined by the PEPS-C assessment) correlates with less prosodic entrainment (as determined by changes in mean f0 and f0 range). More research is needed to shed light on the origin of the significant negative correlation between expressive prosody and f0 entrainment at the conversational level. This future research should include larger numbers of participants as well as multiple conversations for each speaker on which overall conversational entrainment contribution of a speaker should be based.

## 6. Acknowledgements

This research was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103451. We also gratefully acknowledge the help of the following lab assistants: Sarah Carillo, Gabriella Puentes, Angelica Ortiz, Samantha Homan, Sarah Krein, Evelyn Madrid, and Alyssa Hernandez.

## 7. References

[1] M. Babel and D. Bulatov, "The role of fundamental frequency in phonetic accommodation." *Language and Speech*, vol. 55, no. Pt 2, pp. 231–48, 2012.

[2] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2011, pp. 3081–3084.

[3] J. Local, "Phonetic detail and the organisation of talk-in-interaction," in *Proceedings of International Congress of Phonetic Sciences (ICPhS XVI). Saarbrücken*, 2007, p. ID 1785.

[4] H. Giles, N. Coupland, and J. Coupland, "Accommodation theory: Communication, context, and consequence," in *Contexts of Accommodation*. Cambridge University Press, 1991, pp. 1–68.

[5] M. Natale, "Convergence of mean vocal intensity in dyadic communication as a function of social desirability." *Journal of Personality and Social Psychology*, vol. 32, no. 5, pp. 790–804, 1975.

[6] J. Edlund, M. Heldner, and J. Hirschberg, "Pause and gap length in face-to-face interaction," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2009, pp. 2779–2782.

[7] S. A. Borrie and C. R. Delfino, "Conversational Entrainment of Vocal Fry in Young Adult Female American English Speakers," *Journal of Voice*, vol. 31, no. 4, pp. 513.e25–513.e32, 2017.

[8] J. Michalsky and H. Schoormann, "Pitch Convergence as an Effect of Perceived Attractiveness and Likability," in *Interspeech 2017*. ISCA, 2017, pp. 2253–2256.

[9] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker, "Language Style Matching Predicts Relationship Initiation and Stability," *Psychological Science*, vol. 22, no. 1, pp. 39–44, 2011.

[10] J. Michalsky, H. Schoormann, and O. Niebuhr, "Conversational quality is affected by and reflected in prosodic entrainment," in *9th International Conference on Speech Prosody 2018*. ISCA, 2018, pp. 389–392.

[11] N. Lubold and H. Pon-Barry, "Acoustic-prosodic entrainment and rapport in collaborative learning dialogues," in *MLA 2014 - Proceedings of the 2014 ACM Multimodal Learning Analytics Workshop and Grand Challenge, Co-located with ICMI 2014*. Association for Computing Machinery, Inc, 2014, pp. 5–12.

[12] O. Niebuhr and J. Michalsky, "PASCAL and DPA: A Pilot Study on Using Prosodic Competence Scores to Predict Communicative Skills for Team Working and Public Speaking," in *Proceedings of Interspeech 2019*. International Speech Communication Association, 2019, pp. 306–310.

[13] N. Lewandowski and M. Jilka, "Phonetic Convergence, Language Talent, Personality and Attention," *Frontiers in Communication*, vol. 4, 2019.

[14] S. Peppé and J. McCann, "Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version," *Clinical Linguistics and Phonetics*, no. 4-5, pp. 345–354, 2003.

[15] R. Baker and V. Hazan, "DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs," *Behavior Research Methods*, vol. 43, no. 3, pp. 761–770, 2011.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Amsterdam, 2017. [Online]. Available: http://www.praat.org

[17] D. B. Fry, "Experiments in the perception of stress," *Language and Speech*, no. 1-2, pp. 125–152, 1958.

[18] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge University Press, 2008.