Pitch Declination and Reset as a Function of Utterance Duration in Conversational Speech Data

Céline De Looze, Irena Yanushevskaya, Andy Murphy, Eoghan O'Connor and Christer Gobl

Phonetics and Speech Laboratory, Trinity College Dublin, Ireland

deloozec@tcd.ie, yanushei@tcd.ie, murpha61@tcd.ie, oconnoe7@tcd.ie, cegobl@tcd.ie

10.21437/Interspeech.2015-624

Abstract

This paper describes the declination trends of f_0 in conversational speech data. A 10-minute dialogue interaction from a corpus of spontaneous speech was annotated to identify inter-silence units (ISU) and turns. Detailed annotation of the ISUs was conducted in terms of communicative types and pitch patterns. f_0 declination was measured by (1) fitting a regression line to f_0 trajectories and (2) by fitting additional regression lines to the data points below and above the original (central) regression line. The slope of declination as well as the height of ISU/turn-initial f_0 peak were examined as a function of the duration of the ISU or turn. The results suggest that declination is indeed present in conversational speech data, at the level of both the ISU and the turn (73% of the analysed ISUs exhibited negative f_0 declination slope). There is a tendency for the steepness of the slope to decrease and the height of IS-turn-initial f_0 peak to increase as the duration of the ISU or turn increases. The results are discussed in the context of Projection and Reaction theories and of Hard vs. Soft pre-planning of speech production. The findings are of potential interest for the development of human-machine dialogue systems.

Index Terms: f_0 declination, reset, turn-taking, prosodic planning, human-machine interaction

1. Introduction

This paper examines how f_0 declination contributes to turn-taking organisation in conversational interactions. This study is part of ongoing research exploring voice source and temporal features, their correlation with melodic characteristics, and their combination for linguistic and paralinguistic signalling.

Turn-taking organisation is the process by which speakers agree on who speaks, when to talk, listen, hold and take turns. To manage turn-taking, speakers would produce, perceive and react to a set of signals (prosodic, pragmatic, syntactic, semantic, visual) (*Reaction Theory* e.g., [1-5]) or would anticipate or project the end of the turn from contextual and structural information (*Projection Theory*, e.g., [6-8]).

Within the frame of Reaction Theory, it has been reported that, in many languages, a level pitch accent or a flat contour at the end of an utterance is indicative of turn-holding while any other terminal contour (such as rises and falls) are indicative of turn-taking [3, 9-14]. In these studies, the prosodic characteristics of the end of utterances (e.g. the last 500 ms, or a second) have been examined. In the present work, in exploring f_0 declination trends at the level of the inter-silence unit (ISU) and the turn, we focus particularly on the initial portions of utterances, as well as on the whole utterances. It is our hypothesis that f_0 declination and reset (or reset at a higher pitch level) play an important role in turn-taking organisation by chunking speech units into larger units and by signalling whether a speaker intends to yield or hold a turn.

Works in the study of prosody have often reported that, in read speech, fundamental frequency declines over the course of an utterance [15-17]. Declination is thought to be the by-product of some physiological processes (e.g. subglottal pressure [15, 18], activity of the laryngeal muscles [19], tracheal pull [16]), or 'controlled' by the speaker and have some specific linguistic and paralinguistic functions. It may however be suspended in certain instances of terminal rises associated with questions or hesitations [20]. Evidence of f_0 declination at supra-utterance levels (e.g. above the level of the Intonation Phrase) has also been reported in many languages, with paragraph initial utterances of spoken texts having higher and wider pitch than paragraph final utterances [21-25]. Pitch declination over the paragraph (or f_0 supra-declination) may participate in signaling the organisational and hierarchical structure of the discourse (e.g. signalling topic changes). It has been further argued that the height of the f_0 peak at the beginning of an utterance and the slope of f_0 declination depend on the utterance length, with longer utterances being marked by higher f_0 peaks and less steep declination slopes [26-28].

While the presence of f_0 declination is well established in the studies of laboratory speech, it is not clear if the same trends are present in spontaneous conversational interactions. We hypothesise that, in interactional speech, f_0 declination can be observed too, both at the ISU and at the turn (supra-ISU) level, and that it plays an important role in turn-taking organisation. ISUs in conversational speech may be embedded within turn units, as speech units in read speech are embedded within paragraph units. A downward shift or decrease in f_0 across successive ISUs by the same speaker may indicate that they belong to the same turn, while an upward shift or increase would signal turn changes.

In a preliminary analysis [29], we have measured pitch range declination at the level of the turn in terms of ISU pitch level (f_0 median) and range (f_0 standard deviation) and have shown that there is a pitch range declination trend between the initial and median ISUs of a turn but a violation of this declination for the final ISU of a turn. We suggested that this was due to a confounding effect of rises in final ISUs. Our results also demonstrated that the higher the number of speech units in a turn, the higher the turn-initial f_0 peak height.

In this paper, we examine f_0 declination at the level of the ISU and at the level of the turn. Measures of f_0 declination and

reset (initial f_0 peak) are different from the previous study. In this latter, f_0 peak height corresponded to the automatically extracted f_0 maximum value at the beginning of an utterance and f_0 declination was explored in terms of pitch range declination (f_0 median and standard deviation). In the present study, f_0 peak height is based on the first accented syllable of the unit (manually annotated). f_0 declination is measured in terms of three regression lines computed from smoothed f_0 contour (described in section 2.4).

We examine (i) whether, as in read speech, f_0 declination operates in conversational interaction, and (ii) whether turn/ISU f_0 reset (measured as the height of ISU/turn-initial f_0 peak) and f_0 declination slope depend on the ISU/turn length. The following hypotheses are tested:

Hypothesis 1: f_0 declination operates at the level of the ISU and at the level of the turn;

Hypothesis 2: the steepness of the f_0 declination slope is correlated with the length of the ISU/turn (the steeper the ISU/turn f_0 declination slope, the shorter the ISU/turn);

Hypothesis 3: the length of the ISU/turn is correlated with the magnitude of f_0 reset (the higher the ISU/turn f_0 reset, the longer the ISU/turn).

2. Material and method

2.1. Speech data

The speech data for this study consisted of a 10-minute dialogue interaction between two female speakers of Irish English, in their early 20s, taken from the Dublin Institute of Technology Emotional Speech Corpus [30]. The interaction was elicited in a shipwreck scenario game where the participants were given 10 minutes to jointly rank-order 15 items in terms of usefulness for their survival. Recordings were made with participants in separate booths using a professional Neumann microphone connected to an Apple Mac-based Digidesign Pro-Tools Mbox2 recording system. The audio signal was recorded using Pro-Tools software as two separate audio streams and digitised at 96 kHz/24 Bit. Audio was then downsampled to 16 kHz/8 Bit.

2.2. Extraction of inter-silence units (ISU) and turns

As the first step, automatic annotation of the audio data into speech and silence intervals was conducted. Binary voice activity detection (VAD) using the VAD algorithm proposed in [31] was carried out on both speaker channels. The threshold for silence interval duration was set to 100 ms to avoid false detection of pauses for speech events like plosives [32]. Silent intervals below the threshold were bridged. Figure 1 illustrates schematically the output of the VAD process.

Automatic annotation of the audio data into speech and silent intervals was subsequently manually corrected. The speech-chunks between the silent intervals of or above the 100 ms threshold are referred to as inter-silence units (ISU). For each ISU the type of speaker-transition (pause, gap, overlap) was also identified. Pause is defined as a silent interval within the same speaker turn; gap is a silent interval after which a speaker change takes place, and overlap occurs when the second speaker begins to talk while the first one is still talking. Turn is defined here as a stretch of speech produced by the same speaker; backchannels are not considered as turn-taking/speaker change. Overall, 396 ISUs and 248 turns were identified. The majority of the turns consisted of one or two ISUs (65% and 21% respectively). Turns containing 3 ISUs made up 7% of the data, and those with 4 ISUs - only 3%. Turns with 5-7 ISUs made up less than 3% of the data. Note that 75% of the ISUs consist of one Intonation Phrase (IP).

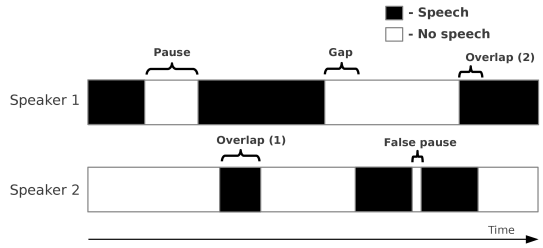

Figure 1. *Schematic representation of the VAD process illustrating the types of speaker transition (pause, gap, overlap).*

2.3. Corpus annotation

The identified ISUs were subsequently annotated using Praat [33]. In addition to the orthographic tier, communicative types and pitch pattern annotation was also conducted following the description in [14]. The communicative types included, for example, Declaratives (DEC), Incomplete Declaratives (IDEC), Yes/No questions (YN), Wh-questions (WHQ), Incomplete questions (IQ), Imperatives (IMP), Exclamations (EXCL), Backchannels (BC), Hesitations (HES) etc. The pitch pattern annotation was done using IViE [34] and included the description of pitch accents, syllable prominence and boundary tones. The midpoints of stressed (S) and unstressed (w) vowels were also annotated.

The general description of the distribution of communicative types and pitch patterns in the conversation analysed is given in Figure 2. Backchannels, Incomplete declaratives, Hesitations, Incomplete questions, Declaratives and Exclamations are the most frequently occurring communicative types in the analysed conversation. The pitch patterns are predominantly falling (42%), followed by level (23%) and rising (11%) pitch patterns.

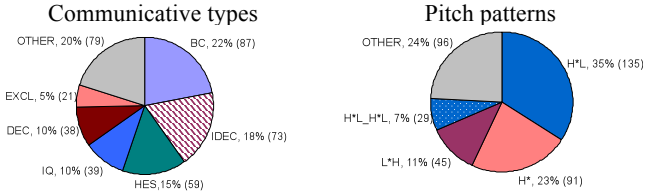

Figure 2: *The distribution of communicative types and pitch patterns in the analysed dialogue. BC = Backchannels, IDEC = Incomplete declaratives, HES = Hesitations, IQ = Incomplete questions, DEC = Declaratives, EXCL = Exclamations. H*L=falling pitch, L*H=rising pitch, H*=level.*

2.4. Data analysis

The analysis of f_0 was done in Praat [33]. To avoid possible pitch tracking errors at the pitch curve extrema, pitch floor and pitch ceiling, when creating a Pitch Object, were automatically adjusted to the speaker's pitch range (cf. [35] for more details). A number of measures of declination were used, similar to [36] and [37]. First, declination was modelled by fitting a least-squares regression line to the f_0 trajectory and the slope

of the regression line was used as a measure of declination (a similar method was used in [36]). Furthermore, following [37], two additional declination lines were fitted to the values above and below the original (central) declination line, and the slopes of the upper and the lower regression lines are also examined. The regression lines are illustrated in Figure 3. The span of the declination is measured as the difference between the upper and the lower regression lines at the beginning and at the end of the ISU or turn. Positive values indicate narrowing of declination span over time (as in Figure 3), negative values are indicative of expanding of the span and suspension of declination trends.

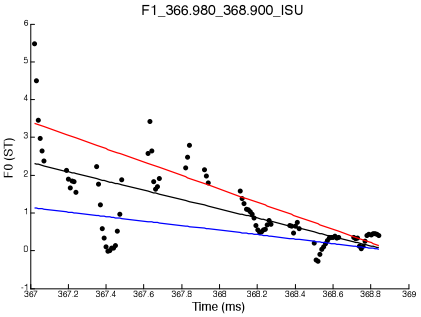

Figure 3 : *Example of regression lines fitted to the f_0 curve of an ISU 'Just in case we'll pick up any signal at all'.*

The data extracted for the analysis included (i) ISU and turn duration in s, (ii) f_0 values (in semitones; ST) over the course of each ISU and turn, (iii) the timing and height (in ST) of the ISU or turn-initial f_0 peak measured at the midpoint of the first stressed vowel (S) in the ISU or turn, (iv) the slopes of the regression lines (in ST/s) and the declination span (in ST).

3. Results

3.1. Declination trends and communicative types and pitch patterns

Declination trends were analysed in a subset of data that excluded HES and BC as well as speech units shorter than 500 ms. Furthermore, we only consider regression line slopes within the range of +/- 20 ST/s (the slopes outside that range are likely to be aberrant values). In total, 159 turns and 189 ISUs were analysed. Overall, the results point at the presence of f_0 declination in the conversational data analysed: 73% of the ISUs and 64% of the turns have negative f_0 slope. Note that as the slopes for the central, upper and lower regression lines were highly correlated ($r=0.9$; $p < 0.001$), the results reported here are based on the central regression lines data. With regard to declination span, results show that f_0 declination goes with a narrowing of the span in 55% of cases at the turn level and in 52% of cases at the ISU level.

Figure 4 shows the distribution of communicative types and pitch patterns in ISUs with negative and positive f_0 slopes. Communicative types showing negative f_0 slopes are mainly Incomplete Declaratives, Declaratives and Incomplete Questions and they are mainly realised with a falling or level pitch. Communicative types in which f_0 slopes showed positive trend include mainly Incomplete Questions and Incomplete Declaratives and the predominant pitch patterns are rising and level pitch.

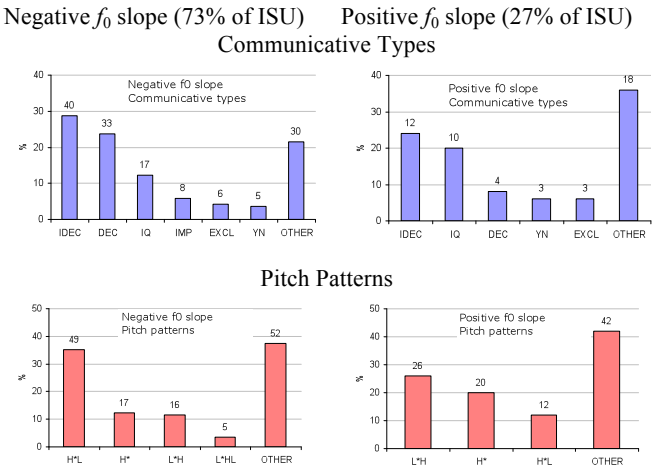

Figure 4 : *The distribution (Y-axis %) of communicative types and pitch patterns in the subsets of data showing negative (left panels) and positive (right panels) f_0 declination slopes. DEC = Declaratives, IDEC = Incomplete declaratives, IQ = Incomplete questions, IMP = Imperatives, YN = Yes/No questions, EXCL = Exclamations, H*L=falling pitch, L*H=rising pitch, H*=level. The numbers on top of the bars represent frequencies of each category.*

In order to investigate the relation between ISU/turn f_0 declination/reset and duration, we excluded from our analyses positive slopes. The final subset of data consisted of 136 ISUs and 95 turns.

3.2. f_0 declination and duration of ISU or turn

Figure 5 shows the mean values of the negative slopes plotted as a function of ISU/turn duration. The ISUs/turns were grouped according to a set of quantised durations by rounding each ISU/turn length. The mean value of their slopes was then plotted for each duration step. We can see that the shorter the ISU (panel a) or the turn (panel b), the steeper the slope of f_0 declination. Linear regression analyses (using the R Statistical Software) with f_0 slope as dependant variable and duration (after logarithmic transformation) as independent variable confirm significant correlation between ISU and turn f_0 negative slopes and duration (ISU level: $F(1,136)=32.31$, $p<0.001$; Turn level: $F(1,143)=90.24$, $p<0.001$).

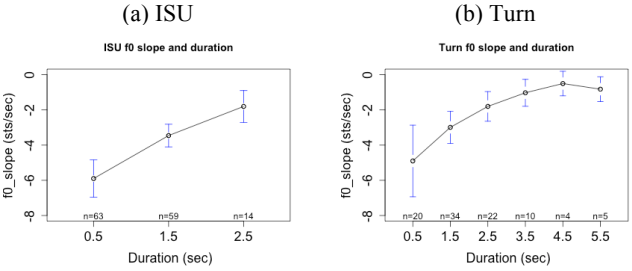

Figure 5: *ISU/turn f_0 declination slope (in semitones per second) vs. ISU/turn duration (in seconds). The ISUs/turns were grouped according to a set of quantised durations by rounding each ISU/turn length. The mean value of their slopes was then plotted for each duration step. Durations were rounded to the nearest number shown on the x-axis.*

3.3. The height of ISU/turn-initial f_0 peak and turn duration

Figure 6 shows the mean values of the ISU/turn initial f_0 peak plotted as a function of ISU/turn duration. The ISUs/turns were grouped according to a set of quantised durations by rounding each ISU/turn length. The mean value of their f_0 peaks was then plotted for each duration step. We can see a tendency towards the increase in the height of initial f_0 peak as the duration of the turn increases (panel b). At the ISU level (panel a), the tendency seems weaker. Linear regression analyses with f_0 peak height (in semitones) as dependant variable and duration (after logarithmic transformation) as independent variable confirm significant (though weak) correlation between turn-initial f_0 peak and duration (Turn level: $F(1,137)=9.447$, $p=0.003$). No significant correlation was found at the ISU level ($p=0.057$).

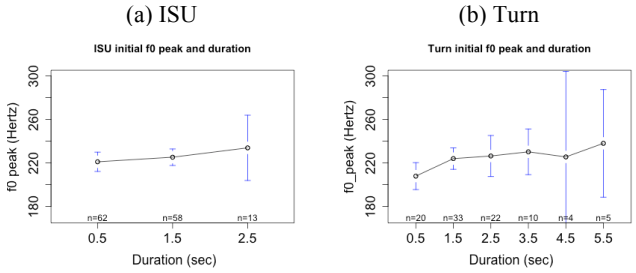

Figure 6: *ISU/turn-initial f_0 peak (in Hz) as a function of turn duration (in seconds). The ISUs/turns were grouped according to a set of quantised durations by rounding each ISU/turn length. The mean value of their f_0 peaks was then plotted for each duration step. Durations were rounded to the nearest number shown on the x-axis.*

4. Discussion and conclusions

This paper investigated f_0 declination and reset in conversational speech data. Note that this preliminary study is based on a small amount of data and will be extended in the future to larger dataset, which will specify the generalisability of the results across different speakers, tasks and languages. This study has served as a pilot testing and developing the methodological approaches to the exploration of f_0 declination in conversational speech.

Our findings suggest that declination operates both at the level of the ISU and at the level of the turn in conversational interaction: 73% of the ISUs and 64% of the turns in our data have negative f_0 slopes. As in read speech, negative slopes are mainly associated with Incomplete Declaratives, Declaratives and Incomplete Questions and are mainly realised with a falling or level pitch. Positive slopes mainly include Incomplete Questions and Incomplete Declaratives and the predominant pitch patterns are rising and level and pitch. In addition, our results suggest that f_0 declination is related to the length of the speech unit: in our data, the steepness of the f_0 slope increased as the duration of the ISU or turn decreased. Finally, our results reveal that the height of the turn-initial f_0 peak is correlated with the turn duration: the longer the turn, the higher its initial f_0 peak. This relation, however, does not hold at the level of the ISU.

Our study corroborates earlier findings on the relation between the initial f_0 peak height as well as slope and the duration of an utterance [26-28]. These findings generally raise the debate of Hard vs. Soft pre-planning of speech production. On the one hand, it is proposed that speakers would be able to plan f_0 contours at a phrase level by adjusting the f_0 height at the beginning of the utterance to the utterance length. A higher initial f_0 may suggest a 'look-ahead' or preplanning mechanism, by which utterance-initial f_0 values are raised proportionate to utterance length [38]. On the other hand, it is suggested that speakers may proceed at a more local level, accent by accent. A lower f_0 at the end of the utterance may mean that adjustment is made 'on-the-fly'. Our preliminary results suggest that speakers may plan their turn, adjusting f_0 initial peak and slope according to the turn length.

Overall, our findings suggest that pitch at the beginning of a turn and f_0 declination of a turn may contribute to turn-taking organisation. This would mean that not only syntactic and pragmatic information but prosody as well, may be used in projecting a speaker change. We believe that both Reaction and Projection theories can account for the underlying functioning of turn-taking organization: speakers could anticipate the end of a turn based on the f_0 peak height at the beginning of the turn (as well as other signals) and on f_0 declination slope, and react to the lately uttered signals, adapting on-the-fly, by readjusting predictions if needed.

The findings reported in this paper could be directly applied to the modeling of human-machine interactions. A lot of work has been lately dedicated in improving the flow of conversation between a human and a computer or virtual agent. Standard methods have used a fixed duration threshold for the computer to begin speaking after the human interlocutor stops [39]. This strategy however does not really mirror what is usually done by humans. Rather than waiting for a silence to come, they rely on syntactic, prosodic, pragmatic as well as visual cues to take the turn. Some studies have therefore investigated the use of these cues (prosodic and syntactic mainly) just before a silence to predict a speaker's hold or change [40]. In this study, our results suggest that turn initial f_0 peak height and slope may also be relevant cues to manage the conversation flow. The height of the initial f_0 peak of a turn could be used by a system to predict the end of the turn, and the signals at the end of the turn (such as a final rise or fall vs. a flat tone) could be used to readjust prediction.

The question that remains to be answered is whether the steepness of f_0 declination slope and reset are used by the participants in conversational interactions to predict turn-taking. In other words, is there a difference between declination slope and reset in the speech chunks preceding pauses and gaps? Future work will explore perceptual salience of the slope, f_0 reset and ISU/turn duration, e.g. by systematic manipulation of the parameters and removing/masking the propositional content. Another direction is to explore how f_0 declination affects other voice source parameters.

5. Acknowledgements

This research is supported by the Science Foundation Ireland Grant 09/IN.1/I2631 (FASTNET).

6. References

[1] A. Kendon, "Some functions of gsze direction in social interaction," *Acta Psychologica,* vol. 26, pp. 22-63, 1967.

[2] V. Yngve, "On getting a word in edgewise," *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society,* pp. 567-577, 1970.

[3] S. Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of Personality and Social Psychology,* vol. 23, pp. 283-292, 1972.

[4] A. Cutler and M. Pearson, "On the analysis of prosodic turn-taking cues," in *Intonation in Discourse*, C. Johns-Lewis, Ed., ed London: Croom Helm, 1985, pp. 139-155.

[5] C. E. Ford, B. A. Fox, and S. A. Thompson, "Practices in the construction of turns: the 'TCU' revisited," *Pragmatics,* vol. 6, pp. 427-454, 1996.

[6] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language,* vol. 50, 1974.

[7] E. A. Schegloff, "Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences," in *Analyzing Discourse: Text and Talk*, D. Tannen, Ed., ed: Georgetown University Press, 1982, pp. 71-93.

[8] J. P. De Ruiter, H. Mitterer, and N. J. Enfield, "Projecting the end of a speaker's turn: a cognitive cornerstone of conversation," *Language,* vol. 82, pp. 515-535, 2006.

[9] J. Local and J. Kelly, "Projection and 'silences': notes on phonetic and conversational structure," *Human Studies,* vol. 9, pp. 185-204, 1986.

[10] C. E. Ford and S. A. Thompson, "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns," in *Interaction and Grammar*, E. Ochs, E. A. Schegloff, and S. A. Thompson, Eds., ed Cambridge: Cambridge University Press, 1996, pp. 134-184.

[11] M. Selting, "On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation," *Pragmatics,* vol. 6, pp. 357-388, 1996.

[12] H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa, and Y. Den, "An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs," *Language and Speech,* vol. 41, pp. 295-321, 1998.

[13] J. Caspers, "Local speech melody as a limiting factor in the turn-taking system in Dutch," *Journal of Phonetics,* vol. 31, pp. 251-276, 2003.

[14] I. Yanushevskaya, J. Kane, C. De Looze, and A. Ní Chasaide, "The distribution of pitch patterns and communicative types in speech chunks preceding pauses and gaps," presented at the Speech Prosody, Dublin, Ireland, 2014.

[15] P. Lieberman, *Intonation, Perception, and Language. MIT Research Monograph, Vol. 38*. MIT, 1967.

[16] S. Maeda, "A characterization of American English intonation. Doctoral dissertation.," MIT, 1976.

[17] J. 't Hart, R. Collier, and A. Cohen, *A Perceptual Study of Intonation: an Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press, 1990.

[18] R. Collier, "Physiological correlates of intonation patterns," *Journal of the Acoustical Society of America,* vol. 58, pp. 249-225, 1975.

[19] J. J. Ohala, "Respiratory activity in speech," in *Speech Production and Speech Modelling*. vol. 55, W. J. Hardcastle and A. Marchal, Eds., ed: Springer Netherlands, 1990, pp. 23-53.

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

[21] I. Lehiste, "Perception of sentence and paragraph boundaries," in *Frontiers of Speech Communication Research*, B. Lindblom and S. Ohman, Eds., ed New York: Academic Press, 1979, pp. 91-101.

[22] A. M. Sluijter and J. M. Terken, "Beyond sentence prosody: paragraph intonation in Dutch," *Phonetica,* vol. 50, pp. 180-188, 1993.

[23] G. Bruce, "Textual aspects of prosody in Swedish," *Phonetica,* vol. 39, pp. 274-287, 1982.

[24] P. Nicolas and D. Hirst, "Symbolic coding of higher-level characteristics of fundamental frequency curves," presented at the 4th European Conference on Speech Communication and Technology, Madrid, Spain, 1995.

[25] H. den Ouden, L. Noordman, and J. Terken, "Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports," *Speech Communication,* vol. 51, pp. 116-129, 2// 2009.

[26] K. Snider, "Tone and utterance length in Chumburung: an instrumental study," presented at the 28th Colloquium on African Languages and Linguistics, Leiden, the Netherlands, 1998.

[27] E. Couper-Kuhlen, "Interactional prosody: high onsets in reason-for-the-call turns," *Language in Society,* vol. 30, pp. 29-53, 2001.

[28] A. Rialland, "Anticipatory raising in downstep realization: evidence for preplanning in tone production," in *Cross-Linguistic Studies of Tonal Phenomena*, S. Kaji, Ed., ed Tokyo, Japan: Tokyo University of Foreign Studies, 2001, pp. 301-321.

[29] C. De Looze, I. Yanushevskaya, J. Kane, and A. Ní Chasaide, "Pitch range declination and reset in turn-taking organisation," presented at the Speech Prosody 2014, Dublin, Ireland, 2014.

[30] B. Vaughan, "Naturalistic emotional speech corpora with large scale emotional dimension ratings. PhD thesis," Dublin Institute of Technology, 2011.

[31] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters,* vol. 6, pp. 1-3, 1999.

[32] J. Trouvain and M. Grice, "The effect of tempo on prosodic structure," presented at the IVth International Congress of Phonetic Sciences, San Francisco, California, 1999.

[33] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 5.3.85 ed, 2014.

[34] E. Grabe, B. Post, and F. Nolan, "Modelling intonational variation in English: the IViE system," presented at the Prosody 2000, Kraków, Poland, 2001.

[35] C. De Looze, "Analyse et Interprétation de l'Empan Temporel des Variations Prosodiques en Français et en Anglais. PhD thesis," Aix-Marseille University, 2010.

[36] J. Yuan and M. Liberman, "F0 declination in English and Mandarin broadcast news speech," presented at the Interspeech 2010, Makuhari, Japan, 2010.

[37] J. Haan, "Speaking of Questions. An Exploration of Dutch Question Intonation. LOT Dissertation Series, No 52," Netherland's Graduate School of Linguistics, Utrecht, 2002.

[38] B. Connell, "Tone, utterance length and f0 scaling," presented at the International Symposium on Tonal Aspects of Languages with Emphasis on Tone Languages, Beijing, China, 2004.

[39] A. Raux and M. Eskenazi, "Optimizing endpointing thresholds using dialogue features in a spoken dialogue system," presented at the SIGdial 2008, Columbus, OH, USA, 2008.

[40] A. Gravano and J. Hirshberg, "Turn-taking cues in task-oriented dialogue," *Computer Speech and Language,* vol. 25, pp. 601-634, 2011.