Pitch range declination and reset in turn-taking organisation

Céline De Looze, Irena Yanushevskaya, John Kane and Ailbhe Ní Chasaide

Phonetics and Speech Laboratory
School of Linguistics, Speech and Communication Sciences, Trinity College Dublin, Ireland
[deloozec, yanushei, kanejo, anichsid]@tcd.ie

Abstract

This paper examines how pitch range declination and reset contribute to turn-taking organisation. This is part of a broader study of voice prosody, i.e., how pitch, voice quality and temporal features combine for various prosodic functions, both linguistic and paralinguistic. The present study first investigates the effect of the speech unit position in a turn on its pitch range. We also test the effect of the number of speech units in a turn as well as the turn duration on the turn-initial f_0 peak height at the beginning of the turn. Our results suggest a pitch range declination trend between the Initial and Median speech units of a turn but a violation of this declination for the Final units of the turn. They also demonstrate that the higher the number of speech units in a turn or the longer the turn, the higher the turn-initial f_0 peak height. We discuss our findings along the debate on Projection and Reaction theories and that of Hard vs. Soft pre-planning of speech production. We address how these findings may be useful to formulate a holistic model of prosody and to enhance human-machine interactions.

Index Terms: pitch range, declination, reset, pause, gap, turn-taking, reaction vs. projection theories, hard vs. soft pre-planning.

1. Introduction

Spoken interaction is a joint activity where all participants are involved in the co-construction of meaning and in the establishment and maintenance of social relationships. Turn-taking organisation is an instance of such coordination, for which participants agree on who speaks, when to talk, listen, hold and take turns [1].

Depending on the situational context, turn-taking (employed herein as speaker change) can be realised with a silent interval (or gap), no-gap-no-overlap (when the start of one speaker's turn perfectly coincides with the end of the other speaker's previous turn) or with a transition overlap (when the start of one speaker's turn overlaps with the other speaker's previous turn, excluding backchannels). In this organisation, turn-holding (employed herein as speaker hold) corresponds to several utterances of the same speaker within a turn, that are separated by a silent interval (or pause).

To manage turn-taking, speakers would produce, perceive and react to a set of signals (prosodic, pragmatic, syntactical, semantical, visual) (*Reaction Theory* e.g., [2, 3, 4, 5, 6]) or would anticipate or project the end of the turn from contextual and structural information (*Projection Theory*, e.g., [1, 7, 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 a turn-holding while any other terminal contour (such as rises and falls) is indicative of turn-taking [4, 9, 10, 11, 12, 13].

Works in the study of prosody have often reported that, in read speech, fundamental frequency declines over the course of an utterance [14, 15, 16]. Declination is thought to be the by-product of some physiological processes (e.g. subglottal pressure [14, 17], activity of the laryngeal muscles [18], tracheal pull [15]) or 'controlled' by the speaker and may have some specific linguistic and paralinguistic functions. It may however be violated in certain instances of terminal rises associated with questions or hesitations [19].

Evidence of f_0 declination at supra-utterance levels (e.g. above the level of the Intonation Unit) has also been reported in many languages, with paragraph initial utterances of spoken texts having higher and wider pitch range than paragraph final utterances [20, 21, 22, 23, 24]. This pitch range declination over the paragraph (or f_0 supra-declination) may participate in signaling the organizational and hierarchical structure of the discourse (e.g. signaling topic changes). [24] reported, for Dutch, that the lower a text segment is embedded within the hierarchy of a text, the lower the pitch range. [23] observed that, in French, intonation units at the beginning of a paragraph have a wider pitch range than medial and final position intonation units, the difference being greater between the initial and medial units than between the medial and final units.

We hypothesize that, in interactional speech, f_0 declination can also be observed, both at the utterance and at the turn (supra-utterance) level, and that it plays an important role in turn-taking organisation. Speech units in conversational speech may be embedded within turn units, as speech units in read speech are embedded within paragraph units. As suggested by Couper-Kuhlen [25], describing data of conversational speech, a downward shift or decrease in f_0 across successive utterances by the same speaker may indicate that they belong to the same turn, while an upward shift or increase would signal turn changes. In particular, she suggests that : "*Beginning an intonation phrase relatively high in one's voice range allows room for subsequent intonation phrases to be positioned lower and thus affords the possibility of declination units, which can be used to structure a 'big package'. Because high onsets initiate pitch declination units, they can be thought of as projecting 'more to come' in this case, further intonation phrases within the declination unit. In this sense, they provide prospective prosodic cues to the 'big package' that is under way*" ([25]:43).

It has been further argued that the height of the f_0 peak at the beginning of an utterance is indicative of the utterance length, with longer utterances being marked by higher f_0 peaks [26, 27, 28]. Similarly, we hypothesize that, in interactional speech, the first unit of a speaker's turn may be marked by a pitch reset whose height depends on the number of speech units within the turn or on the turn duration. Note that the relation between initial f_0 peak height and utterance length is still controversial as other studies did not observe any [29, 30].

10.21437/SpeechProsody.2014-209

In this paper, we investigate (i) whether pitch range declination operates in interactional speech at the level of the turn and (ii) whether pitch reset at the beginning of a turn depends on the number of utterances within the turn or on the turn length.

We test two hypotheses:

- **Hypothesis H1**. Initial Inter-Pausal Units (IPUs) within a turn have a higher and wider pitch range than Median and Final IPUs. To test this, the effect of the IPU's position (Initial/ Median/ Final) in a turn on the IPU's pitch range is investigated.
- **Hypothesis H2**. The f_0 peak at the beginning of a turn is higher when the number of IPUs in the turn is larger or when the turn is longer. To test this, we investigate the effect of the number of IPUs in a turn and the turn duration on the turn's initial f_0 peak.

It should be noted that most prosodic research on turn-taking has tended to focus on the duration of silent intervals, and on the chunk of speech preceding these intervals. In examining f0 declination trends, we focus particularly on the initial portions of utterances, as well as the relationships among the utterances of a turn. This work complements parallel work on the final intonational contours in the same data [41] as well as ongoing research exploring voice quality and temporal features and their correlation with melodic characteristics.

2. Experiment

2.1. Data

The speech data was extracted from a corpus of task-oriented dyadic interactions in Irish English [31]. It consists of 6 gender-paired interactions (involving 6 female and 6 male speakers). The interactions are based on a shipwreck scenario game where participants are presented with 15 items and are given 10 minutes to rank them in order of usefulness to their survival. Recordings were carried out with participants in separate isolation booths using a professional Neumann microphone connected to an Apple Mac-based Digidesign Pro-Tools Mbox2 recording system. The audio signal was digitised at 96 kHz/24 Bit and recorded using Pro-Tools software as two separate audio streams. Audio was then downsampled to 16 kHz/8 Bit.

2.2. Annotation and measurement

2.2.1. Annotation

The data was annotated in terms of speech units and silences automatically, similarly to [32]. A binary voice activity detection (VAD) was carried out on both speaker channels for each dyadic interaction, using the VAD algorithm proposed in [33].

Pauses, gaps, no-gap-no-overlaps (NGNO) and transition overlaps (TOV) were determined from the binary VAD on both speaker channels. A schematic output of the annotation is shown in Fig. 1. A minimum duration for pauses was set to 100 ms (threshold set empirically to avoid speech events like plosives to be annotated as pauses), which means that every speech unit separated by a pause of less than 100 ms were chunked together into a single speech unit. Note that decisions on such threshold settings may significantly affect the resulting duration distributions [34]. This automatic procedure resulted in 176 gaps, 121 TOV and 498 pauses.

Herein, TOV were excluded from the analyses. Speech units separated by a pause are referred to as Inter-Pausal Units (IPU). A turn is defined as a speech unit composed of one or several IPUs. The terms 'turn-taking' and 'speaker change' are used interchangeably. No annotation of backchannels (BC) have been included, which means that BC have been automatically annotated as speaker changes.

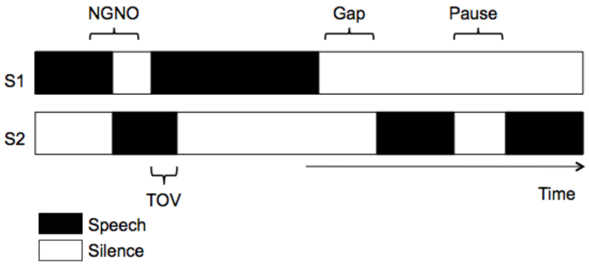

Figure 1: Schematic of a dyadic conversation between speaker 1 (S1) and speaker 2 (S2), illustrating occurrences of pauses, gaps, no-gap-no-overlaps (NGNO) and transition overlaps (TOV).

In the data, turns are composed of one to thirteen IPUs (mean=2.5, sd=2.12). Most common turns contain one IPU (40% of the data). To test Hypothesis H1, we excluded turns composed of one IPU as pitch range declination across several IPUs cannot be investigated in such turns. We also excluded turns composed of more than six IPUs as they are under-represented in the data (less than 3%). From the resulting subset, IPUs were labelled Initial, Median and Final according to their position in a turn. In total, 292 IPU were labelled Initial, 243 Median and 311 Final. To test Hypothesis H2, the subset of selected data consisted of turns composed of one to six IPUs.

2.2.2. Measurement

Pitch range corresponds to the tonal space actually used in speech and is commonly described along two dimensions : its level (height) and span (range) [19]. Pitch range declination corresponds to the f_0 downward shift or decrease in pitch range across successive utterances within a turn and can be compared to the concepts "paragraph declination" [35], "supra declination behavior" [22] or "superordinate f_0 declination" [36] used in the literature for read speech.

Different acoustic measurements have been used to estimate pitch range level and span. Descriptive measurements such as f_0 mean and median based on f_0 values extracted every 0.01s on voiced segments or based on f_0 inflection points (or tonal targets, e.g. peaks and valleys) have been employed to account for utterance f_0 level. Measurements such as the difference between f_0 maximum and minimum values of utterances, f_0 95th and 5th percentiles, f_0 90th and 10th percentiles of utterances as well as f_0 standard deviation have been used to account for utterance f_0 span.

In this work, for each IPU and turn, standard descriptive measurements (f_0 maximum, minimum, median and the difference between f_0 maximum and minimum), representative, respectively, of an IPU's and a turn's initial f_0 peak, final valley, pitch range level and span, were computed (cf. Table 1). All these measurements are given on a logarithmic scale, the octave scale (i.e. log2(Hertz)), which is equivalent to the semitone scale. To account from cross-speaker differences, the data was normalized using z-scores. Pitch measurements were extracted using the phonetic software Praat [37]. To avoid possible pitch tracking errors at the pitch curve extrema, and enable their auto-

Features	Measurement	Abbreviation
initial f_0 peak	maximum f_0	$f_0$max
final f_0 valley	minimum f_0	$f_0$min
pitch level	median f_0	$f_0$med
pitch span	$f_0$max$-f_0$min	$f_0$maxmin

Table 1: Features computed for each turn (T) and Inter-Pausal Unit (IPU). Initial f_0 peaks are extracted at the beginning of each T and IPU, final f_0 valley at the end.

matic extraction, pitch floor and pitch ceiling, when creating a Pitch Object, were automatically adjusted to the speaker's pitch range (cf. [38] for more details).

2.3. Statistical analyses

A series of statistical analyses were carried out to test the effect of the independent variables (i.e. IPU position and number of IPUs in a turn) on the dependent variables (i.e. pitch range measurements). The analyses included a series of linear mixed models [39]. In our models, the position of the IPU (IPU-POS), the number of IPUs in a turn (N-IPU) and the duration of a turn (TDUR) were treated as fixed factors. Speaker was included as a random factor to take into account inter-speaker variability. The p-values were calculated using the method of Monte Carlo sampling by Markov chain (pMCMC = Monte Carlo Markov Chain [40]). In all our models, significance is set at a pMCMC $\alpha< 0.01$.

3. Results

3.1. H1 - Declination at the turn level

The effect of the speech unit position (IPU-POS) - 3 levels, Initial/ Median/ Final - on the IPU's pitch range is investigated. We assume that Initial IPUs have higher and wider pitch range than Median and Final IPUs. For the dependent variables $f_0$min, $f_0$max, $f_0$med and $f_0$maxmin, IPU-POS is tested as fixed factor and SPEAKER as random factor.

Results reveal a significant effect of IPU-POS on $f_0$min, $f_0$max, $f_0$med and $f_0$maxmin, being respectively higher and wider for Initial-IPUs than for Median-IPUs (respectively : t= -2.102, pMCMC< 0.01; t= -1.921, pMCMC< 0.01; t= -1.129, pMCMC< 0.01;t= -1.316, pMCMC< 0.01). This suggests a pitch range declination trend (i.e. lowering and narrowing of the pitch range) between the Initial and Median IPUs of the turn. The declination is however violated at the end of the turn, as it may be confounded with instances of rises associated with certain question forms or hesitations.

Two examples of pitch range declination trend observed in the data are given in Fig. 2 and 3. In Fig. 2, the turn is characterized by a declination trend over the three IPUs, the final IPU having a lower and narrower pitch range than the preceding IPUs. In Fig. 3, the turn is characterized by a declination trend over the first two IPUs which is violated on the final IPU, the latter having a higher pitch range than the preceding IPU.

3.2. H2 - Turn's initial f_0 peak as a function of the number of IPUs in a turn and turn duration

We investigate the effect of the number of IPUs (NIPU) in a turn and the turn duration (TDUR in log) on the turn's initial f_0 peak. We assume that the f_0 peak at the beginning of a turn is higher when the number of IPUs in the turn is larger or when the turn is longer. The models include NIPU and TDUR as fixed factor and SPEAKER as random factor.

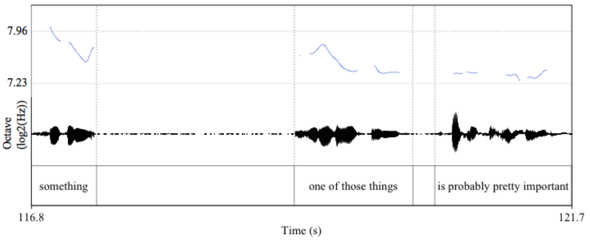

Figure 2: Example of a turn composed of 3 IPUs, uttered by a female speaker.

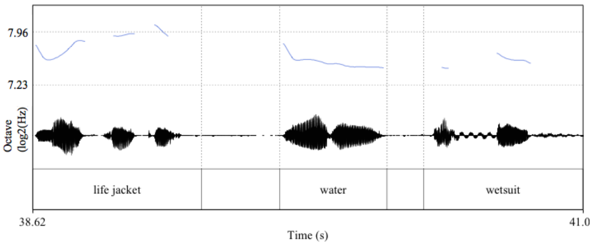

Figure 3: Example of a turn composed of 3 IPUs, uttered by a female speaker.

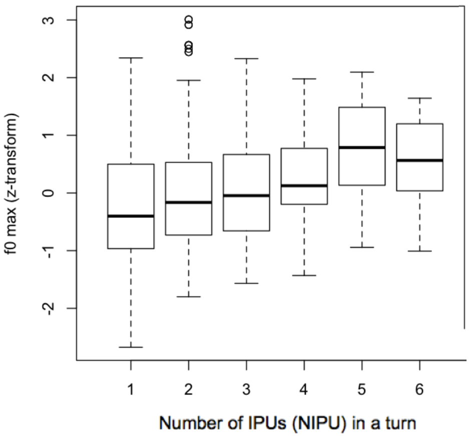

Figure 4: Turns' initial peaks (or $f_0$max) according to the number of IPUs in a turn.

Results reveal a significant effect of NIPU (pMCMC< 0.01) and TDUR on $f_0$max (t=12.208, pMCMC< 0.01). The higher the number of IPUs in a turn and the longer the turn, the higher the initial f_0 peak of the turn. Figure 4 indicates how the height of the f_0 peak increases with the number of IPUs in a turn. These findings suggest that the initial f_0 peak of a turn may be a salient cue in projecting the end of a turn. They corroborate earlier studies that observed a strong relation between an utterance's f_0 peak height and the utterance duration.

4. Discussion and conclusions

In this paper, we have investigated how pitch range declination and reset contribute to turn-taking organisation. We have first tested the hypothesis H1 that Inter-Pausal Units in dialogue speech are embedded into turns as utterances in read speech are embedded into paragraphs, in such a way that Initial IPUs are higher and wider in pitch range than Median and Final IPUs. Our results suggest a declination trend (lowering and narrowing of the pitch range) between the Initial and Median IPUs of the turn, but not over the final IPU.

In a parallel study, using the same data, we have shown that 49% of units preceding a change of turn are declaratives, 35% questions and 10% backchannels and that these communicative types are associated with a falling tune (in 67% of the cases) or a low-rise tune (in 29% of the cases) [41]. The proportion of rises is the lowest in Declaratives (10%), it is higher in Incomplete Declaratives and WH questions (17%), and is the highest in Backchannels (24%). In Incomplete Questions, the pitch is predominantly rising (only 12% of samples have falling pitch). In Yes/No Questions, both falling and rising pitch pattern is used, with a slight preference for rises (about 54% in total).

Taking into account the effect of the utterance communicative type (e.g. declarative, incomplete, Wh-question, Yes/No question, hesitation, backchannel) on pitch modifications will allow us to better understand its singular role in turn-taking organisation. Pitch variations convey multiple functions in speech, e.g. signaling questions vs. statement, focus, topic changes and turn-taking, as well as in the signaling of intentions, attitudes and affect. The many linguistic functions of f0 changes were not controlled in the present experiment and may explain the variability encountered in the data. Future work will investigate the role of pitch variation at different levels, and will further attempt to link these to other prosodic variables, voice quality and temporal structure.

[23] observed that, in read speech, the difference in pitch range between the Initial and Medial units of a paragraph is greater than between the medial and final units. The present data show similar results. It would be interesting to investigate whether this pitch range 'break' between the initial and median units is indicative of a turn length, therefore may be used to signal turn change.

We have then tested the Hypothesis H2 that the f_0 peak at the beginning of a turn is higher when the number of IPUs in the turn is larger or when the turn is longer. Our results show that the higher the number of speech units in a turn and the longer the turn, the higher the initial f_0 peak height. This corroborates earlier findings on the relation between the initial f_0 peak height and the duration of an utterance [26, 27, 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 [29]. 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 its f_0 initial peak according to the turn length.

Overall, our findings suggest that pitch at the beginning of a turn and the break between the Initial and Median IPUs of a turn may contribute to turn-taking organisation. This means that not only syntactic and pragmatic information but prosody as well, appears to be used in projecting a speaker change.

We believe that both Reaction and Prediction theories can account for the underlying functioning of turn-taking organisation. As explained in [42], "*Redundancy is a well-studied and recurring principle of human language in use on virtually every level, and it is likely that a phenomenon as important as the taking of turns is orchestrated by a number of redundant control methods*" ([42]:566). In this view, we propose that speakers may 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 may react to the lately uttered signals, adapting on-the-fly, by readjusting predictions if needed.

These findings 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 [43]. This strategy however does not really mirror what is usually done by humans. They, indeed, rather than wait for a silence to come, 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 [44].

In a parallel study [41], using the same data, we have shown that the combined discriminative power of functional and intonation labels (derived from speech-chunks immediately preceding pause and gap intervals) allows for differentiating turn-taking from turn-holding (mean classification error of 15%). In the present study, our results suggest that prosodic information at the beginning of a turn may also be a relevant cue 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) may be used to readjust prediction. This will be particularly addressed in our future classification experiments.

5. Acknowledgments

This research is supported by the Science Foundation Ireland Grant 09/IN.1/I2631 (FASTNET). We would like to thank Dr. Brian Vaughan (Dublin Institute of Technology) for providing the DIT Emotional Speech Corpus.

6. References

[1] H. Sacks, E. A. Schegloff, and G. Jefferson, “A simple systematic for the organization of turn-taking in conversation,” *Language*, vol. 50, no. 4, pp. 696–735, 1974.

[2] A. Kendon, “Some functions of gaze-direction in social interaction,” *Acta Psychologica*, vol. 26, pp. 22–63, 1967.

[3] V. Yngve, “On getting a word in edgewise,” in *Chicago Linguistics Society, 6th Meeting*, 1970, pp. 567–578.

[4] S. Duncan, “Some signals and rules for taking speaking turns in conversations.,” *Journal of Personality and Social Psychology*, vol. 23, no. 2, pp. 283–292, 1972.

[5] A. Cutler and M Pearson, “On the analysis of prosodic turn-taking cues,” *Intonation in Discourse*, pp. 139–156, 1986.

[6] C. Ford, B. Fox, and S. Thompson, “Practices in the construction of turns: The itc revisited,” *Pragmatics*, vol. 6:3, pp. 427–454, 1996.

[7] E. Schegloff, “Discourse as an interactional achievement: Some use of ‘uh-huh’ and other things that come between sentences,” *Georgetown University Round Table on Languages and Linguistics, Analyzing discourse: Text and talk*, pp. 71–93, 1982.

[8] J. De Ruiter, H. Mitterer, and N. Enfield, “Projecting the end of a speaker’s turn: A cognitive cornerstone of conversation,” *Language*, pp. 515–535, 2006.

[9] J. Local and J. Kelly, “Projection and silences: Notes on phonetic and conversational structure,” *Human Studies*, vol. 9, no. 2, pp. 185–204, 1986.

[10] C. Ford and S. Thompson, “Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns,” *Studies in Interactional Sociolinguistics*, vol. 13, pp. 134–184, 1996.

[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. 371–388, 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, no. 3-4, 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, no. 2, pp. 251–276, 2003.

[14] P. Lieberman, “Intonation, perception, and language,” *MIT Research Monograph*, 1967.

[15] S. Maeda, *A characterization of American English intonation*, Massachusetts Institute of Technology, 1976.

[16] 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.

[17] R. Collier, “Physiological correlates of intonation patterns,” *The Journal of the Acoustical Society of America*, vol. 58, pp. 249, 1975.

[18] J. Ohala, “Respiratory activity in speech,” in *Speech Production and Speech Modelling*, pp. 23–53. Springer, 1990.

[19] D. Ladd, *Intonational Phonology*, Cambridge University Press, 2008.

[20] I. Lehiste, “Perception of sentence and paragraph boundaries,” *Frontiers of speech communication research*, pp. 191–201, 1979.

[21] G. Bruce, “Textual aspects of prosody in swedish,” *Phonetica*, vol. 39, no. 4-5, pp. 274–287, 1982.

[22] A. Sluijter and J. Terken, “Beyond sentence prosody: Paragraph intonation in dutch,” *Phonetica*, vol. 50, no. 3, pp. 180–188, 1993.

[23] P Nicolas and D. Hirst, “Symbolic coding of higher-level characteristics of fundamental frequency curves,” in *Fourth European Conference on Speech Communication and Technology*, 1995.

[24] H. 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, no. 2, pp. 116–129, 2009.

[25] E. Couper-Kuhlen, “Prosody and sequence organization in English conversation,” *Sound Patterns in Interaction. Amsterdam: John Benjamins*, pp. 335–376, 2004.

[26] K. Snider, “Tone and utterance length in Chumburung: An instrumental study,” *28th Colloquium on African Languages and Linguistics, Leiden*, 1998.

[27] E. Couper-Kuhlen, “Interactional prosody: High onsets in reason-for-the-call turns,” *Language in Society*, vol. 30, no. 1, pp. 29–53, 2001.

[28] A. Rialland, “Anticipatory raising in downstep realization: Evidence for preplanning in tone production,” *Cross-linguistics Studies of Tonal Phenomenon: Tonogenesis, Typology, and Related Topics*, pp. 301–322, 2001.

[29] B. Connell, “Tone, utterance length and f_0 scaling,” in *International symposium on tonal aspects of languages: With emphasis on tone languages*, 2004.

[30] P. Prieto, M. D’Imperio, G. Elordieta, S. Frota, M. Vigário, et al., “Evidence for ‘soft’ preplanning in tonal production: Initial scaling in Romance,” in *Proceedings of Speech Prosody*, 2006, pp. 803–806.

[31] B. Vaughan, *Naturalistic Emotional Speech Corpora with Large Scale Emotional Dimension Ratings*, Ph.D. thesis, Dublin Institute of Technology (DIT), 2011.

[32] M. Heldner, J. Edlund, and J. Hirschberg, “Pitch similarity in the vicinity of backchannels,” in *Proceedings of Interspeech 2010*, 2010, pp. 1–4.

[33] J. Sohn, N. Kim, and W. Sung, “A statistical model-based voice activity detection,” *Signal Processing Letters, IEEE*, vol. 6, no. 1, pp. 1–3, 1999.

[34] M. Wlodarczak and P. Wagner, “Effects of talk-spurt silence boundary thresholds on distribution of gaps and overlaps,” *Proceedings of Interspeech, Lyon, France*, pp. 1434–1437, 2013.

[35] J. Garrido, J.and Llisterri, C. Mota, and A. Ríos, “Prosodic differences in reading style: isolated vs. contextualized sentences,” in *Third European Conference on Speech Communication and Technology*, 1993.

[36] N. Gronnum Thorsen, “Intonation and text in standard danish,” *The Journal of the Acoustical Society of America*, vol. 77, no. 3, pp. 1205–1216, 1985.

[37] P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” 2006.

[38] C. De Looze, *Analyse et interprétation de l’empan temporel des variations prosodiques en français et en anglais contemporain*, Ph.D. thesis, Université de Provence, 2010.

[39] J. Pinheiro, D. Bates, S. DebRoy, and D. Sarkar, “Linear and nonlinear mixed effects models,” *R package version*, vol. 3, pp. 57, 2007.

[40] R. Baayen, *Analyzing linguistic data*, vol. 505, Cambridge University Press Cambridge, UK, 2008.

[41] 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,” *Proceedings of Speech Prosody 2014*, In press.

[42] M. Heldner and J. Edlund, “Pauses, gaps and overlaps in conversations,” *Journal of Phonetics*, vol. 38, no. 4, pp. 555–568, 2010.

[43] A. Raux and M. Eskenazi, “Optimizing endpointing thresholds using dialogue features in a spoken dialogue system,” in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*. Association for Computational Linguistics, 2008, pp. 1–10.

[44] A. Gravano and J. Hirschberg, “Turn-taking cues in task-oriented dialogue,” *Computer Speech & Language*, vol. 25, no. 3, pp. 601–634, 2011.