Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Speaking loudly reduces flexibility and variability in the prosodic marking of focus types

Simon Roessig, Lena Pagel, Doris Mücke

University of Cologne, Germany
mail@simonroessig.de

Abstract

Modulations of F0 height and movement magnitude are used to mark focus and differentiate between focus types. Similar F0 changes have been described for loud speech. In this production study, we investigate the interplay of speaking style (habitual vs. loud speech) and focus type (broad vs. corrective focus) by analyzing characteristics of nuclear pitch accents in German. Our study reveals that the prosodic system becomes less flexible in the differentiation of focus types in loud speech. While nuclear accents of falling and rising types occur in habitual speech, there are only rising accents in loud speech. In loud speech, the differentiation between broad and corrective focus may be maintained by a gradual difference in the magnitude of the pitch accent rise excursions, but this difference is weaker than in habitual speech. Interestingly, the speaking styles are characterized by different variability profiles within and across speakers: In loud speech, the productions across speakers become more similar and speakers tend to be more consistent in their individual patterns. We discuss the findings in the light of physiological explanations and suggest that they may exemplify how communicative demands and production constraints interact in shaping prosodic patterns.

Index Terms: speech production, prosody, intonation, pitch accents, loud speech, variation

1. Introduction

Loud speech is not only characterized by higher intensities but also affects, among other phonetic dimensions, fundamental frequency (F0). Most studies have reported higher F0 means on the *global level* of sentence production (English: [1]; German: [2]; French: [3]; Swedish: [4]; Finnish: [5]). In addition, it has been found that loud speech exhibits larger F0 movements or increased F0 fluctuations [1], [6] (although the opposite has also been reported [4]). Higher F0 means and larger F0 movements have additionally been attributed to closely related speaking styles, such as clear speech, e.g., [7], [8], and Lombard speech, e.g., [9], [10].

Crucially, F0 is one of the primary parameters of prosodic prominence. In contrast to loud speech, prosodic prominence affects F0 on or around a single syllable, i.e., on a *local level.* In West-Germanic languages, F0 height and movement magnitude are tightly connected to the placement of pitch accents (unaccented vs. accented) but also to the degree of prominence inherent to different pitch accent types and their realizations. As such, F0 is strongly modulated in the expression of focus structure: Nuclear pitch accents in narrow or corrective focus are characterized by higher F0 targets and larger F0 movements than nuclear pitch accents in broad focus (English: [11]; German: [12], [13], [14]). In general, different pitch accent types have been found to be associated with certain focus types, although there is no one-to-one relation between pitch accent type and focus type. In German, the proportion of falling pitch accent types, such as H+L* decreases from broad to corrective focus and gives way to a larger proportion of rising accents such as L+H* [13] (see also [15] for L+H* in corrective focus in English). Consequently, while clear distributional preferences or tendencies can be attested, there remains a fair deal of variation in the marking of focus types, and the mapping can be described as probabilistic [16]. In addition, continuous modifications towards higher F0 means and larger F0 excursions within one accent category in corrective focus have been reported [14].

Since both loud speech and the differentiation of focus types seem to modulate F0 in similar ways, the question arises: Can focus types be expressed in loud speech to the same extent as in habitual speech? In particular, does the fact that loud speech *scales F0 up* on the global level prevent the *gradation of prominence* for focus types on the local level? This question addresses the distributions of accent types (proportions of falling and rising accents) as well as the continuous modifications (higher F0 means, larger F0 excursions).

In the present study, we explore the interplay of loud speech and the marking of focus types by nuclear pitch accents in German. We quantify the direction of the accentual F0 movement (as a categorical distinction: falling vs. rising) and the magnitude of this movement (as a quantity to assess continuous difference) with the same measure. Our results suggest that the prosodic flexibility for focus marking is reduced in loud speech: We show that the differentiation between broad focus and corrective focus is decreased in loud speech compared to habitual speech. In addition, we capture the degree of variability across and within speakers. We report that the realization of nuclear pitch accents becomes less variable in loud speech, not only between but also within focus types. We investigate whether the latter finding can be attributed to an increased *across-speaker* consistency (i.e., speakers in the sample become more similar) or increased *within-speaker* consistency (i.e., the productions of one speaker become more similar) and conclude that both contribute to the reduction of variability. In light of research suggesting a strong link of intensity and F0, we suggest that our results may show an interplay of communicative demands and production constraints. We follow this line of thought to conclude that the interplay would affect phonological and phonetic aspects of prosody.

2. Methods

Eight native speakers of German were recorded acoustically using a headset microphone and with electromagnetic

10.21437/SpeechProsody.2022-102

articulography simultaneously. The articulatory data are not reported here. Four of the speakers identified as female, four as male. They were between 23 and 27 years old (mean: 24.75; SD: 1.39) and had not received special training in phonetics, phonology or speech analysis. The speakers were seated in front of a screen and fulfilled a game-like task, which included the interaction with a virtual avatar. The game was designed as an interactive web application. Speakers were told they went to see a soccer match with their friend Marie. Marie had forgotten her glasses and the speakers were tasked with helping her by describing the scenery on the soccer field. Marie was displayed as the virtual avatar and her questions were auditorily presented to the participants.

The questions were pre-recorded by a female native German speaker in habitual and loud speech. The answers to these questions were the target sentences displayed on the screen. As the indirect object, the target sentences included one out of six German-sounding pseudo words, which were introduced as names of the soccer players: Nimami /niˈmaːmi/, Libami /liˈbaːmi/, Sibabi /ziˈbaːbi/, Nabima /naˈbiːma/, Labiba /laˈbiːba/, and Samima /zaˈmiːma/.

The target sentences were elicited in two different focus conditions. The avatar's question elicited the target word to be either in *broad focus* or in *corrective focus*, as illustrated in Table 1. In these examples, the focus domain is marked by square brackets and the subscript letter F. Each target word was produced in each focus condition twice. The order of the utterances was pseudo-randomized and filler trials were inserted.

Table 1: *Examples for question-answer pairs*

Focus type	Example
broad	Q: Was passiert gerade? *What is going on?* A: [Carlotta spielt Nabima zu.]$_F$ *Carlotta passes the ball to Nabima.*
corrective	Q: Spielt Carlotta Holly zu? *Does Carlotta pass the ball to Holly?* A: Carlotta spielt [Nabima]$_F$ zu. *Carlotta passes the ball to Nabima.*

During the first half of the experiment, the utterances were produced in a *habitual speaking style*, in the second half in a *loud speaking style*. Participants were told that during the half-time break, the atmosphere in the stadium heated up and there was a lot of noise. They were asked to speak very loudly so that their interlocutor Marie could still understand them. No actual background noise was played as to avoid Lombard effects, but speakers reported they could adapt their speaking style nevertheless.

For the present data set, 384 utterances were recorded, of which 381 were included in the analysis (one production was excluded due to technical issues, two productions were excluded because no reliable F0 track could be obtained). The data were forced-aligned using the Montreal Forced Aligner [17] and hand-corrected in Praat [18].

The F0 contours were labelled in Praat using the following annotation scheme: First, listening to the whole utterance, it was confirmed that the nuclear accent was placed on the indirect object. Second, listening to the sequence of the pre-accentual and accentual syllable (two syllables window), it was decided whether the accentual movement was rising or falling. The beginning and end of the movement were marked within the two-syllable window *plus* the post-accentual syllable. The latter syllable was included because rising movements may reach their peaks shortly after the accented syllable. For falling movements, the end point of the fall was sometimes difficult to determine, since the boundary tone of the contour was low and the contour was hence falling throughout the following syllables. In cases of doubt, the end of the fall was marked at the midpoint of the accented vowel.

The *tonal onglide* was measured as the distance in semitones between the beginning and the end of the accentual movement: $12 \times \log_2\left(\frac{E}{B}\right)$, where E is the F0 at the end of the movement and B is the F0 at the beginning of the movement. The measure is illustrated in Figure 1, where the black box represents the accented syllable (σ*) and the red arrows indicate the tonal onglide. For rising accents, the onglide is positive; for falling accents, the onglide is negative. For the calculation of the F0 track, we used the default *To Pitch* function of Praat with a range of 75 to 700 Hz.

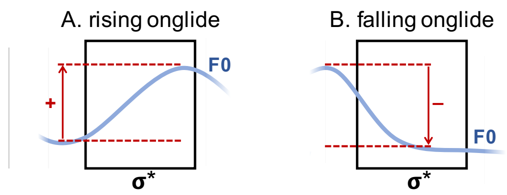

Figure 1: *Onglide measure*

To check whether our data are able to reproduce previous results on F0 in loud speech, we also calculated the *mean F0* over the whole utterance. While this may also take spurious F0 points in voiceless segments and stretches of creaky voice into account, the calculation of the mean is rather robust and represents a reasonable approximation to the general level of F0. To improve the measure, we used Praat's more detailed *To Pitch (ac)* function with a conservative value for *voicing threshold* of 0.8 (otherwise default values with the same range as above).

3. Results

Before moving to the central part of the study – the tonal onglide patterns of the nuclear pitch accents – we briefly report the mean F0 values on the global level of utterance production. Our data is able to confirm the general pattern of increased F0 means found by many previous studies. While the mean F0 is 210 Hz for female and 131 Hz for male speakers in habitual speech, it increases to 375 Hz for female and 227 Hz for male subjects in loud speech. (Note that these values represent the average of mean values calculated for each utterance.)

The distributions of tonal onglide values are shown in Figure 2 for the two speaking styles (habitual / loud) and focus types (broad / corrective). The figure presents the data in half violin plots (right side) and dot plots (left side). The half violin visualizes the distributions as a kernel density plot. The dot plot is similar to a histogram, as it stacks data points falling into one bin. This left side of the plot increases the transparency of the data presentation in showing each data point. The thick black dot indicates the mean of the distribution. In addition, Table 2 summarizes the proportion of falling (negative) and rising (positive) onglides. The table also comprises two mean values for each style-focus combination. The “Mean overall” column contains the mean of all onglides of this style-focus combination (and corresponds to the thick black dot in Figure 2). The “Mean rises” column contains the mean of the rising onglides only, i.e., the means of the sub-distributions where the onglide is above zero.

Both Figure 2 and Table 2 clearly show that falls *and* rises occur in habitual speech. In this speaking style, we see a reduction in the proportion of falls from broad to corrective focus, which is partially responsible for the increase of the overall mean from broad to corrective focus. However, as Table 2 shows, even when only rising accents are considered, there is still a considerable difference between broad and corrective focus: Tonal onglides of rises become larger in corrective focus (broad: 3.27 st; corrective: 4.10 st; difference: 0.83 st). In loud speech, there are only rising accents in our data. The differentiation of broad vs. corrective focus is, however, encoded gradually in the magnitude with larger onglides for corrective focus compared to broad focus (broad: 3.69 st; corrective: 4.26 st; difference: 0.57 st). These data show that broad and corrective focus are generally more similar in loud speech than in habitual speech.

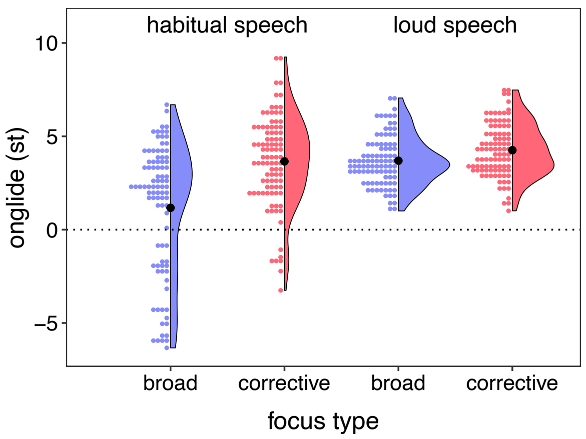

Figure 2: *Distributions of onglides by style and focus*

Table 2: *Overview of onglides by style and focus*

Style	Focus type	Falls	Rises	Mean overall	Mean rises
habitual	broad	29.5%	70.5%	1.25 st	3.27 st
habitual	corr.	7.4%	92.6%	3.66 st	4.10 st
loud	broad	0%	100%	3.69 st	3.69 st
loud	corr.	0%	100%	4.26 st	4.26 st

We fitted a Bayesian multilevel model to the subset of *rising* onglides with brms [19] with focus type, speaking style, and the style-focus interaction as fixed effects. We used random intercepts for subject and target word as well as by-subject and by-word random slopes for focus type, speaking style, and their interaction. The intercept prior was a normal distribution with a mean of 1 and standard deviation of 10; the regression coefficient priors were normal distributions with a mean of zero and standard deviation of 10. We are interested in the estimate β of the difference between the focus types (corrective minus broad focus). For habitual speech, the model provides strong evidence that the onglide magnitudes are higher in corrective than in broad focus: The probability that the difference β between corrective and broad focus is positive is estimated to be 1 ($\beta = 1.00$, CI = [0.48, 1.53], $P(\beta > 0) = 1$). For loud speech, the model estimates larger onglides for corrective than for broad focus but the probability that the difference β is positive is only 0.91 ($\beta = 0.57$, CI = [-0.17, 1.31], $P(\beta > 0) = 0.91$).

The speaking styles are characterized by very different variability profiles reflected in the shape of the distributions and the standard deviations: Broad focus has a standard deviation of 3.49 st in habitual speech, while it has a standard deviation of 1.32 st loud speech; corrective focus has a standard deviation of 2.42 st in habitual speech, while it has a standard deviation of 1.42 st loud speech (see Figure 3).

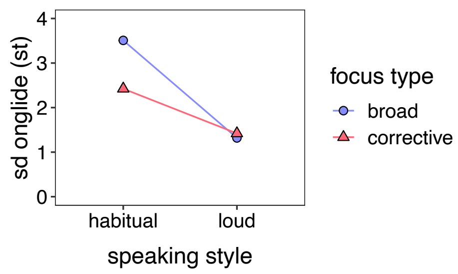

Figure 3: *Standard deviations by style and focus*

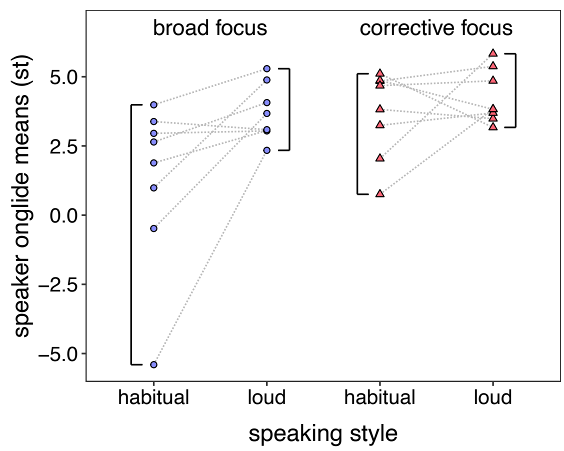

Figure 4: *Speaker onglide means. Each point corresponds to the onglide mean of one speaker.*

These findings may be explained in terms of variability across and within speakers: First, speakers could become more similar to each other when speaking loudly (i.e., decrease in *across-speakers* variability). Second, each speaker could become less variable in their individual productions (i.e., decrease in *within-speaker* variability). Figure 4 shows the mean onglides of each speaker. The dotted line connects the means of each speaker in a focus type across styles. The square brackets visualize the range of mean values across all speakers. The ranges *shrink* from habitual to loud speech. This trend is very clear for broad focus, but it can also be attested for corrective focus. Consequently, in both focus types, the variability *across* speakers decreases considerably from habitual speech to loud speech. Figure 5 visualizes the standard deviations for all speakers separately. Each speaker shows a downtrend in standard deviation from habitual to loud speech, with the exception of speaker S5 in broad focus. Interestingly, this speaker uses only falling accents for broad focus and only rising accents for corrective focus in habitual speech. The results for the vast majority of cases indicate that the variability *within* speakers decreases in loud speech.

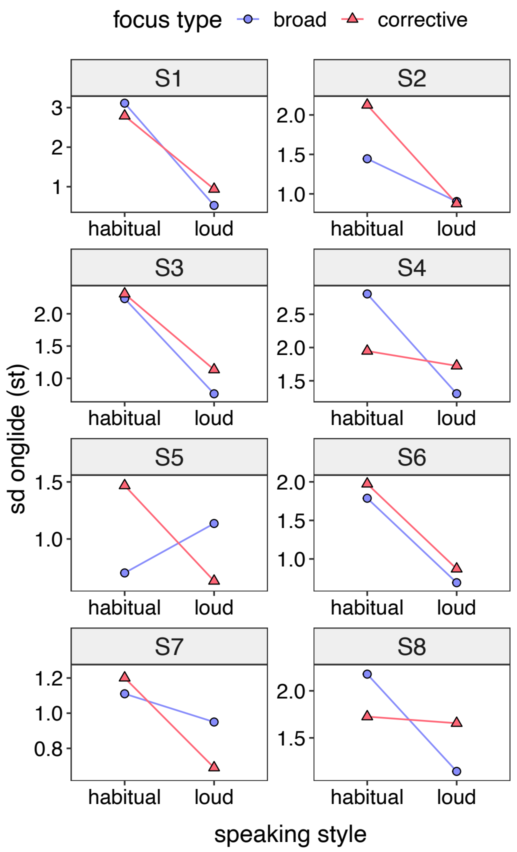

Figure 5: *Standard deviations of individual speakers*

4. Discussion

This study explored how a major function of intonation, namely the marking of focus structure, is realized across two speaking styles: habitual and loud speech. In a first step, the study was able to reproduce previous findings on increased mean F0 in loud speech on the global level of the utterance. Furthermore, it revealed differences in the prosodic patterns of focus marking between the speaking styles. In habitual speech, there are two aspects of focus type differentiation that go hand in hand. Broad and corrective focus are characterized by different proportions of falling and rising accent types with more falls in broad focus. In addition, the rising accents have larger magnitudes in corrective focus than in broad focus. Both the change in the distribution of accent types (increasing proportion of positive onglides) and phonetic realization (larger rises) are reflected in overall higher onglide values in corrective focus. In loud speech, exclusively rising accents are found. Hence, only the continuous aspects of scaling the rising accent remains. However, even this difference between the focus types was found to be weaker. The prosodic system seems to lose some of its flexibility used in the gradation of focus types to the overall 'scaling up' for loud speech.

We found that the distributions of tonal onglides exhibit a lower degree of variability in loud speech than in habitual speech (in both focus types). This lower variability in loud speech can be attributed to more consistency in the patterns *across* and *within speakers*. The trend towards more consistent patterns in loud speech, in particular more rising onglides, may be attributed to the fact that rising accents (in particular L+H*) entail more prosodic prominence [15]. The preference for rising accents is already apparent from broad to corrective focus in habitual speech. We thus see a potential conflict of focus marking and loud speech: Since speaking loudly eliminates falling accents and leaves exclusively rising accents in broad focus, the probabilistic mappings of accent types to focus types (with more falls in broad focus) is neutralized. The differentiation of focus types is thus left to the continuous means only.

We suggest that these results reflect an interaction between physiological traits of speech production and the prosodic system. Reasons for the tendency towards rising onglides and less differentiation in rise magnitude in loud speech may be sought in the physiological relation between F0 and intensity. Increased subglottal pressure leads to increased vocal fold vibration rate and intensity at the same time, by altering the transglottal pressure gradient [20]. As such, an increase in F0 may be interpreted as a passive consequence of the increased subglottal pressure needed for raising the sound pressure [22] – although the relation between F0 and intensity may be non-linear [23]. The reduced F0 variability in loud speech may thus be caused by the need to raise the subglottal pressure close to a limit. The requirement to raise the intensity influences the possible accent type choices, so that falling accents become unlikely, since they lead to low F0 in the accented vowel. The same mechanism may also account for the reduction of falling onglides from broad to corrective focus in habitual speech to a lower extent. Our study may also provide evidence in favor of [21]'s proposal of "a shared control parameter" for F0 and intensity. This control parameter may be able to change the probabilities of *phonological* (categorical) pitch accent choices in conjunction with the *phonetic* (continuous) implementation of these accents [14]. It may already be scaled up when speaking loudly (primarily for the production of high intensities), restricting accent category choices and minimizing room for gradual differences between focus types in this speaking style.

In future research, it will be interesting to investigate whether the differentiation between focus types is supported by other means, such as peak alignment or the scaling of the nuclear pitch accent relative to a preceding prenuclear accent. These means of differentiation may be more independent of the relation of F0 and intensity.

5. Conclusions

Data on the marking of information structure in different speaking styles show less differentiation of focus types in loud speech, i.e., the differences in the tonal onglides for broad focus and corrective focus are smaller in loud speech. Overall, loud speech reduces variability in tonal onglides across speakers as well as within speakers. An explanation for this finding may be that the link between intensity and F0 restricts the choice and realization of pitch accents and thus reduces the flexibility of the prosodic system.

6. Acknowledgements

This work was supported by the German Research Foundation (DFG) as part of the SFB1252 "Prominence in Language" (Project-ID 281511265), project A04 "Dynamic modelling of prosodic prominence" at the University of Cologne.

7. References

[1] C. Dromey and L. O. Ramig, “Intentional Changes in Sound Pressure Level and Rate: Their Impact on Measures of Respiration, Phonation, and Articulation,” *Journal of Speech, Language, and Hearing Research*, vol. 41, pp. 1003–1018, 1998, doi: 10.1044/jslhr.4105.1003.

[2] A. Geumann, “Invariance and variability in articulation and acoustics of natural perturbed speech,” in *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)*, vol. 38, P. Hoole, Ed. Munich: Institut für Phonetik und Sprachliche Kommunikation, Ludwig-Maximilians-Universität München, 2001, pp. 265–393.

[3] J.-S. Liénard and M.-G. di Benedetto, “Effect of vocal effort on spectral properties of vowels,” *The Journal of the Acoustical Society of America*, vol. 106, pp. 411–422, 1999, doi: 10.1121/1.428140.

[4] H. Traunmüller and A. Eriksson, “Acoustic effects of variation in vocal effort by men, women, and children,” *The Journal of the Acoustical Society of America*, vol. 107, pp. 3438–3451, 2000, doi: 10.1121/1.429414.

[5] T. Raitio, A. Suni, J. Pohjalainen, M. Airaksinen, M. Vainio, and P. Alku, “Analysis and Synthesis of Shouted Speech,” in *Proceedings of INTERSPEECH, 25-29 August, Lyon, France*, 2013, pp. 1544–1548. doi: 10.21437/Interspeech.2013-391.

[6] C. T. Ishi and T. Kanda, “Prosodic and voice quality analyses of loud speech: differences of hot anger and far-directed speech,” in *SMM19, Workshop on Speech, Music and Mind, 14 September, Vienna, Austria*, 2019, pp. 1–5. doi: 10.21437/smm.2019-1.

[7] S. Köster, “Acoustic-phonetic characteristics of hyperarticulated speech for different speaking styles,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 6-11 June, Toronto, Canada*, 2001, pp. 873–876. doi: 10.1109/icassp.2001.941054.

[8] R. Smiljanić and A. R. Bradlow, “Production and perception of clear speech in Croatian and English,” *The Journal of the Acoustical Society of America*, vol. 118, pp. 1677–1688, 2005, doi: 10.1121/1.2000788.

[9] C. Davis, J. Kim, K. Grauwinkel, and H. Mixdorff, “Lombard speech: Auditory (A), Visual (V) and AV effects,” in *Proceedings of Speech Prosody, 2-5 May, Dresden, Germany*, 2006, pp. 248–252.

[10] L. Folk and F. Schiel, “The Lombard Effect in Spontaneous Dialog Speech,” in *Proceedings of INTERSPEECH, 27-31 August, Florence, Italy*, 2011, pp. 2701–2704. doi: 10.5282/UBM/EPUB.13690.

[11] M. Breen, E. Fedorenko, M. Wagner, and E. Gibson, “Acoustic Correlates of Information Structure,” *Language and Cognitive Processes*, vol. 25, no. 7, pp. 1044–1098, 2010, doi: 10.1080/01690965.2010.504378.

[12] S. Baumann, J. Becker, M. Grice, and D. Mücke, “Tonal and Articulatory Marking of Focus in German,” in *Proceedings of the 16th International Congress of Phonetic Sciences, 6-10 August, Saarbrücken, Germany*, 2007, pp. 1029–1032.

[13] M. Grice, S. Ritter, H. Niemann, and T. B. Roettger, “Integrating the discreteness and continuity of intonational categories,” *Journal of Phonetics*, vol. 64, pp. 90–107, 2017, doi: 10.1016/j.wocn.2017.03.003.

[14] S. Roessig and D. Mücke, “Modeling Dimensions of Prosodic Prominence,” *Frontiers in Communication*, vol. 4, p. art. 4: 1–19, 2019, doi: 10.3389/fcomm.2019.00044.

[15] J. Pierrehumbert and J. Hirschberg, “The Meaning of Intonational Contours in the Interpretation of Discourse,” in *Intentions in Communication*, P. R. Cohen, J. Morgan, and M. E. Pollack, Eds. Cambridge/London: MIT Press, 1990, pp. 271–311. doi: 10.7551/mitpress/3839.003.0016.

[16] C. Kurumada and T. B. Roettger, “Thinking probabilistically in the study of intonational speech prosody,” *WIREs Cognitive Science*, p. e1579, 2021, doi: 10.1002/wcs.1579.

[17] M. McAuliffe, M. Socolof, S. Mihuc, and M. Wagner, “Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi,” in *Proceedings of INTERSPEECH, 20-24 August, Stockholm, Sweden*, 2017, pp. 498–502. doi: 10.21437/INTERSPEECH.2017-1386.

[18] P. Boersma and D. Weenink, “Praat: doing phonetics by computer.” 2021.

[19] P.-C. Bürkner, “Advanced Bayesian Multilevel Modeling with the R Package brms,” *The R Journal*, vol. 10, no. 1, pp. 395–411, 2018.

[20] I. R. Titze, “On the relation between subglottal pressure and fundamental frequency in phonation,” *The Journal of the Acoustical Society of America*, vol. 85, pp. 901–906, 1989, doi: 10.1121/1.397562.

[21] S. Tilsen, “A shared control parameter for f0 and intensity,” in *Proceedings of Speech Prosody, 31 May - 3 June, Boston, USA*, 2016, pp. 1066–1070. doi: 10.21437/speechprosody.2016-219.

[22] P. Gramming, J. Sundberg, S. Ternström, R. Leanderson, and W. H. Perkins, “Relationship between changes in voice pitch and loudness,” *Journal of Voice*, vol. 2, no. 2, pp. 118–126, 1988, doi: 10.1016/S0892-1997(88)80067-5.

[23] M. Garnier, J. Wolfe, N. Henrich, and J. Smith, “Interrelationship between vocal effort and vocal tract acoustics: A pilot study,” in *Proceedings of INTERSPEECH, 22-26 September, Brisbane, Australia*, 2008, pp. 2302–2305. doi: 10.21437/interspeech.2008-588.