



# Investigating phonetic convergence of laughter in conversation

*Bogdan Ludusan, Marin Schröer, Petra Wagner*

Phonetics Workgroup, Faculty of Linguistics and Literary Studies & CITEC, Bielefeld University,  
Germany

{bogdan.ludusan, marin.schroeer, petra.wagner}@uni-bielefeld.de

## Abstract

Laughter is one of the most encountered paralinguistic phenomena in conversation. Similarly to other communicative elements, evidence for laughter convergence, in particular for its temporal distribution and its acoustic marking, has been found between interlocutors. We investigate here whether segmental-level convergence effects, previously observed for speech, may be also found in the case of laughter. Using a corpus of dyadic interactions, we evaluate phonetic convergence of the vocalic part of laughs, by means of distances between the formant values. This was carried out for two proposed measures of convergence: global – at the level of the entire conversation, and local – considering consecutive laughs. Our global measure results reveal that interlocutors converge towards the end of the interaction, compared to its beginning, although important individual variation exists. With respect to the local measure, our findings show a lack of phonetic convergence (or divergence) between conversational partners.

**Index Terms:** entrainment, phonetic convergence, laughter, dialogue

## 1. Introduction

Convergence (or entrainment), a widely encountered phenomenon in human communication, postulates that interlocutors mutually influence each other during their interaction, resulting in dialogue partners becoming more similar across various modalities and linguistic domains. Several theories have been put forward to explain this process, from being an automatic process driven by priming [1], to being driven by intentional choices, to enhance communication (e.g., [2]), or hybrid accounts, combining elements of the previous two types of proposals [3].

Conversational partners have been observed to converge in terms of both the symbolic and the temporal aspects of their communicative form. At the symbolic level, this is accomplished by adapting on modalities not related to the spoken dimension, such as postures and facial expressions [4], but also with respect to speech, by modifying their syntactic structures [5], lexicon [6], or their acoustic-phonetic segmental and prosodic features (e.g. [7, 8, 9]). An increase in the level of temporal coordination can be found, for instance, by adjusting the timing of feedback signals, the temporal distribution of discourse markers, by speech tempo convergence, or by a synchronization of nonverbal behaviours [10, 11, 12, 13].

As mentioned above, convergence has been shown to occur also on the segmental level, being investigated both perceptually and acoustically. The latter type of investigations included vowel and consonantal characteristics alike [14, 15, 16, 17]. For vowels, convergence is usually expressed in the form of changes in the vowels formant structure, whereas consonants show convergence in an array of different features, depending on the con-

sonant produced (e.g., fricatives converging in their centre of gravity) [18].

Convergence effects are not exclusive to verbal phenomena, having been observed also for non-verbal ones, such as conversational fillers [19]. Laughter, a widely encountered phenomenon in spontaneous interaction [20], has been found to fulfill similar roles in conversation as other non-verbal vocalizations, such as hesitations and fillers. Laughter is not restricted only to the expression of emotional display (mirthful laughter), as it also exhibits a strong social dimension (non-mirthful or social laughter), having various social roles in conversation (see [21] for a review).

Evidence for convergence of laughter in conversation has been previously observed [22, 23, 24], with dialogue partners coordinating the distributions of their produced laughter and acoustically realizing their laughter in a more similar way. Investigating three aspects of entrainment, [22] observed a strong positive correlation between the amount of laughter produced by interlocutors and found evidence for temporal alignment (as seen by the many instances of overlapping laughs), as well as for similarity in the phonetic realization of synchronous laughter. These findings were complemented by those of [23], investigating laughter use in dyadic interactions. They found that conversational partners use similar amounts of laughter across different parts of the conversation, with moderate positive correlations being found in three languages. Additionally, this latter study found that laughter also entrains from an acoustic point of view, exhibiting significantly higher agreement in the voicing of consecutive laughter, when compared to non-consecutive laughter. Evidence for laughter convergence has been found not only at the laughter-token level, but also at the speaker turn-level, whereby interlocutors more often marked turn boundaries with laughter [24]. Moreover, this process was found to occur more often in the second half of the interaction than in its first half [24].

These studies revealed that the conversational function of laughter influences its form, by modifying its distribution and acoustic characteristics. However, there has been no previous study investigating whether convergence-induced segmental-level changes occur in laughter. Given that entrainment at the level of phonetic segments has been observed for speech (e.g., [9]), it would be interesting to explore whether this phenomenon occurs also for laughter. With laughter vowel quality displaying a relatively low intra-speaker variability [25], such adaptation could have noticeable effects on how laughter is used in conversation. We investigate here whether interlocutors become more similar, at the segmental level, in their use of laughter, throughout the interaction. An acoustic analysis of dyadic interactions is performed and the similarity between conversation partners is evaluated by means of the Euclidean distance between the first two formant values of the vocalic parts of the produced laughs.

## 2. Materials

### 2.1. Corpus

The DUEL corpus [26], consisting of dyadic interactions, was employed in our analyses. It contains recordings of French, German and Mandarin Chinese pairs of interlocutors discussing three scenarios: Dream Apartment, Film Script and Border Control. The participants were mostly students, with the majority of them being friends or acquaintances. In the first scenario, the participants would discuss the design and furnishing of a shared apartment, considering that they had a large sum of money at their disposal in order to accomplish this. For the second scenario, the interlocutors were asked to come up with a script for a movie based on an embarrassing moment, which can be based on personal experience. In the last scenario, the dyads enacted a conversation between a border control officer and their in-law trying to enter the country, the latter being in an unfavourable situation. As a result of these tasks, a large amount of laughter, both social and mirthful, was produced throughout the recordings. All materials were orthographically transcribed and annotated for speaker turns and conversational phenomena, including for laughter.

We considered here the German sub-part, for which the first scenario was recorded by 9 dyads and the other two scenarios were recorded, one after the other, by 10 different dyads. In order to avoid convergence already having taken place for the latter 10 dyads, only the scenario at the start of the recording session (Film Script) was employed. The analyzed materials consisted of a total of more than 4 hours 20 minutes of recordings, with 14 minutes per conversation, on average.

### 2.2. Data processing

For this study we looked at instances of “typical” laughter, referred to as “voiced song-like” by [27]. We decided to consider only this type of laughs, as they are employed quite often by most speakers in our data. Moreover, due to their rhythmic, voiced, alternations it is relatively easy to discriminate their vocalic and non-vocalic parts in order to analyze them.

We employed a semi-automatic method for determining the vocalic intervals, as follows. In a first step, based on the existing laughter annotations supplied with the corpus, we automatically selected, using Praat [28], all laughter instances that were at least 300 ms long and at least 40% voiced. While this heuristic did not find all typical laughter instances in the dataset, it returned a sizeable amount of them. Subsequently, these automatically obtained laughs were manually-checked and any instances not perceived as typical laughter were removed. Next, the laughter syllable boundaries were determined automatically by means of the “Mark regions by syllables” function provided by the Praat Vocal Toolkit plugin [29] and the obtained syllables hand-corrected. Finally, the vocalic part of each previously determined laugh syllables was automatically annotated utilizing the “Mark vowels in a TextGrid” function of the same Praat plugin and manually corrected. A total of 1,202 data points (laughter syllables) were obtained after this annotation process.

Once the vocalic segments of the laughter syllables were identified, the mean values of the first two formants (F1 and F2) were extracted using Praat. We decided to focus on F1 and F2 values as they have been previously considered in acoustical analyses of laughter [27, 25] and because of their widespread use in studies of phonetic convergence (e.g., [14, 17, 18]). The default settings for formant extraction given by Praat were used. The formant values were then Mel-scale normalized, sim-

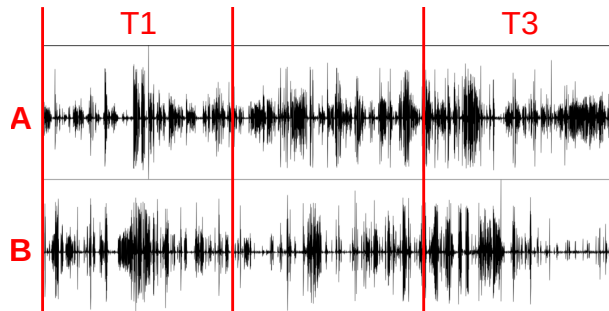


Figure 1: The waveform of a conversation between speakers A and B, illustrating the parts (T1 and T3) considered in our global convergence analysis.

ilarly to other phonetic convergences studies [17], employing the *phonR* [30] R package.

## 3. Methods

We investigated two types of phonetic convergence: global and local. The global convergence describes the degree to which interlocutors are more similar at the end of the conversation, compared to the beginning of the interaction. It is, thus, an overall measure of how the two conversation partners influence each other during their interaction. The second measure, local convergence, characterizes the amount of similarity of laughs produced by the interlocutors, in close temporal proximity. This latter measure is motivated by the fact that evidence supporting the convergence of consecutive laughter along other dimensions (e.g., prosody) has been documented [22, 24].

The similarity between laughs is operationalized at the level of the vocalic part of each laughter syllable, by considering the values of the first and second formants. For laughs having multiple syllables, each syllable represents a different data point. We use a measure of similarity that takes into account the differences between both formants, having been previously employed in convergence studies (e.g., [14]). It is defined as the Euclidean distance between the respective formant values (see Equation 1, where A and B denotes the conversation partners).

$$d_E(A, B) = \sqrt{(F1_A - F1_B)^2 + (F2_A - F2_B)^2} \quad (1)$$

For the global convergence measure, we divide each conversation in three equal parts (see illustration in Figure 1). Then, we consider as belonging to the beginning of the conversation all laughs produced in the first third (T1), and we compare them with the laughs produced at the end of the conversation – all laughs from the last third (T3). The measure was defined at two different levels: dyad- and speaker-level, respectively. At the dyad-level (Equation 2; S denotes speaker, while I interlocutor), the convergence measure compares the Euclidean distance between the speaker and their interlocutor in T1, on the one hand, and the Euclidean distance between the same conversation partners in T3, on the other hand. We also investigated convergence for individual speakers, since interlocutors might show different or even diverging behaviours throughout the conversation. For each speaker, we considered the level of their interlocutor in T1 to be the reference level and we compare the Euclidean distances in T1 and T3 of the speaker with respect to this reference level (see Equation 3). Thus, this measure shows how much each speaker has converged to the level of their interlocutor, as

seen at the beginning of the conversation.

$$global_{dyad} : d_E(S_{T1}, I_{T1}) \iff d_E(S_{T3}, I_{T3}) \quad (2)$$

$$global_{spkr} : d_E(S_{T1}, I_{T1}) \iff d_E(S_{T3}, I_{T1}) \quad (3)$$

The local convergence measure is defined in Equation 4, for each speaker. It compares the Euclidean distance between consecutive laughter pairs and the Euclidean distance of all non-consecutive laughter pairs created by combining one of the two laughs of the consecutive pair with all other laughs (except for its pair) produced by the interlocutor. We defined as consecutive laughs all laughter pairs produced overlapping or within 2 seconds from each other (the start of the second laugh should be within 2 seconds from the end of the first laugh). Also here, two levels were examined, the dyad- and speaker-level. The difference between the two is that while the former level takes into account the contribution of the non-consecutive pairs from both speakers, the per-speaker measure includes only the created non-consecutive laughter pairs of that particular speaker.

$$local : d_E(S_{cons}, I_{cons}) \iff d_E(S_{cons}, I_{non-cons}) \quad (4)$$

The proposed global measure indicates convergence when distances between speakers in T3 are lower than their distances in T1 (or divergence in the opposite case). Similarly, for the local measure, a lower distance between interlocutors for consecutive laughs than for non-consecutive ones suggests convergence. We tested the differences between the obtained values by means of Wilcoxon rank sum tests (considering all distances either at the dyad- or speaker-level). The statistical analyses were performed using the appropriate R [31] functions. Moreover, we used a second method, non-parametric bootstrapping, to validate the findings obtained with the first statistical analysis, since our data was not balanced, with some dyads providing more data points than other dyads. It uses the R *boot* package [32] and performs sampling with replacement from the dataset consisting of the two distances involved in each analyzed measure. The process is repeated for 10,000 times, and at each iteration the mean of each of the two distances is computed. Then, based on the determined 95% confidence intervals, a decision is taken whether the two distances differ significantly (if the intervals do not overlap).

We considered in our analysis of global convergence all dyads which had more than one data point (laughter syllables) for each interlocutor, in each analysed part (T1 and T3). Moreover, each combination of speaker and part (e.g., speaker A in T1 and speaker B in T1) needed to have at least 20 data pairs to be included. Based on these conditions, we were able to examine 6 of the 19 dyads present in the German part of the corpus, having a total of 613 data points in T1 and T3. For the local convergence, we limited ourselves to those dyads producing consecutive laughs, resulting in a set of 10 dyads (the same six as for the global measure, plus four other). They included a total of 426 consecutive laughter syllable pairs and 9,298 non-consecutive laughter syllable pairs.

## 4. Results

The convergence results obtained with the investigated dyad-level measures are presented in Figure 2, for the global convergence on the left side and for the local one on the right side of the figure.

For the global convergence we observed overall higher distances at the beginning of the conversation than at its end, but also important variations between dyads and speakers. We

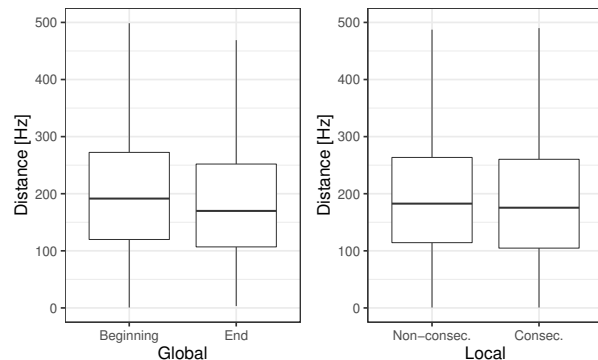


Figure 2: The median, first and third quartile of the Euclidean distances involved in the analysis of the global and local convergence (both computed at the dyad-level). For simplicity, the outlier values were not plotted. A lower distance at the end of the conversation, compared to its start, would indicate global convergence. Similarly, a lower distance for consecutive laughs, compared to non-consecutive ones, would suggest local convergence.

tested the difference between distances in the first and last third of the conversation, on the data pooled from all included dyads/speakers, using Wilcoxon rank sum tests. Both global measures (per-dyad and per-speaker) were found significant ( $p < 2.2e - 16$  and  $p = 4.4e - 06$ , respectively).

Next, we evaluated within-dyad and within-speaker differences between the laughter vowels produced in the first and the last third, by means of Wilcoxon rank sum tests. The obtained results are given in Table 1, with a checkmark (✓) denoting convergence (a significantly lower distance in T3 compared to T1), an x-mark (✗) signifying divergence (a significantly higher distance in T3 than in T1) and a dash (–) representing no significant difference between the two parts of the conversation. Of the six dyads included in our analysis, three of them converged, two diverged and one did not exhibit any change. Also at the level of individual speakers, we see a similarly variable picture, four speakers converging towards the vowel quality of their interlocutors, three diverging and five not showing any difference.

The results obtained by means of bootstrapping largely overlapped with those of the Wilcoxon test, showing a significant difference between dyads, overall. The only diverging results were at the level of dyad “r15” and speaker B of dyad “r13”, for which bootstrapping showed no effect. For the local convergence measure (right side of Figure 2), we noticed a slightly lower difference between the distances involved in its computation, compared to the global convergence measure, suggesting neither convergence nor divergence. This result was confirmed by Wilcoxon rank sum tests, showing no significant difference between differences in formant values of consecutive and non-consecutive laughs ( $p = 0.095$  at both levels).

Examining the local convergence of individual dyads, we observed only one converging dyad, the remaining nine showing no significant difference between consecutive and non-consecutive laughs. At the level of individual speakers, two speakers displayed convergence (those of the dyad which converged), two (belonging to two different dyads) showed divergence, while the rest exhibited no significant difference.

Also for the local convergence measure, the bootstrapping method returned similar results to the Wilcoxon test. The only

Table 1: *Global convergence results, for each of the six dyads included in this analysis, as well as for each speaker within these dyads. A significant result indicates either convergence (✓) or divergence (✗), while non-significant differences are marked with a dash (–).*

Dyad	Result	Speaker	Result
r10	✓	A	–
		B	✓
r11	✓	A	–
		B	✓
r12	✗	A	–
		B	✗
r13	–	A	–
		B	✗
r15	✗	A	✗
		B	–
r19	✓	A	✓
		B	✓

difference was obtained for one of the diverging speakers (according to the Wilcoxon test), which showed an overlap of the bootstrapping confidence intervals (no significant difference).

## 5. Discussion and conclusions

Investigating phonetic convergence of laughter in conversation globally, by comparing the similarity between interlocutors at the beginning and at the end of their conversation, revealed that speakers become more similar to each other in the spectral characteristics of the vocal intervals of their produced laughs. This was observed both at the level of the dyad (speakers being more similar to each other at the end of the interaction), as well as for individual speakers (with speakers becoming, towards the end of the interaction, more similar to the initial level of their interlocutor). However, examining the same phenomenon locally, by studying the similarity of consecutive laughter produced by the conversational partners, our findings revealed no significant effect. Moreover, for both types of convergence measures, we noticed important variation across dyads and speakers, with some of them converging, while other diverging. Comparable results for convergence along various dimensions have been observed, with recent studies having also emphasized the importance of disentrainment for the management of the conversation [33].

Our results partly align with previous findings, showing entrainment effects for laughter, with respect to other dimensions [24]. Our global phonetic convergence measure was found to be significant on the same dataset on which several other types of entrainment, such as temporal and form-related entrainment have been previously observed. However, differently from our results, this latter study reported not only global convergence effects for laughter temporal distribution but also local ones (form-related), by which interlocutors became more similar in the intensity level of consecutive laughs.

This warrants the question why no local phonetic convergence effects were found for laughter, when they have been previously found for other communication elements, including for other dimensions of laughter. First, it might be that interlocutors do not converge along all dimensions of a communication element. We have previously seen, on the same materials, both temporal and form-related laughter entrainment [23, 24]. Here, we observed global convergence effects. As conversa-

tion partners already show entrainment along different dimensions/levels, it might not be necessary to entrain any further. Obviously, further studies, looking at several dimensions of the same phenomenon, are needed for a better understanding of this aspect.

Second, the result might be due to physiological factors, with speakers having less precise control of their articulators during laughter. While some production aspects may still be controlled (see, for instance, the evidence for convergence in voicedness, fundamental frequency level or intensity [22, 23]), others requiring a more fine control of the vocal tract apparatus may not be, due to the tongue being in resting position during laughter [34]. Further indirect support for this explanation comes from studies of the acoustic characteristics of laughter, showing that the quality of the vocalic part of laughs hardly changes within speaker [25]. It could be that this lack of fine control might not allow the speakers to quickly respond to their interlocutor’s level and converge locally with them, but a more gradual change is still possible (as seen by the existence of global convergence effects).

Third, we found no local convergence between dyads, when using distances between the formant values of the vocalic portions of laughs. A recent study [35] suggested that employing a similar method, based on the differences between distances, may, in certain cases, underestimate convergence and that other methods, such as those based on linear mixed effect models may provide a more robust solution. However, while we welcome such methods, for the convergence of the respective models one would require at least an order of magnitude more data than we had here. The limited and unbalanced amount of data we had at our disposal here, or the fact that we had different subsets between the global and the local measures, could have also played a role in the different results we obtained at the two levels. Unfortunately, a higher amount of data is generally not available in the study of several crucial communicative phenomena, including laughter. With the development of larger data sets annotated for laughter, we will be able to confirm our findings on larger data sets, new languages and by employing other methods, as well.

To summarize, our study revealed that interlocutors converge with respect to the values of the first two formants of the vocalic part of laughter, when comparing laughter distances at the end of the conversation with the same measure at the beginning of the interaction, but not locally, in consecutive laughs. Moreover, dyads (and individual speakers) do not show a consistent formant convergence trend in the case of laughter, with some of them exhibiting convergence, while others diverging or showing no change. Besides increasing our understanding of how paralinguistic phenomena are used in human communication, these findings may also have applications for other domains, such as human-machine interaction. A better understanding of all communicative aspects of speech will allow the development of more naturally-interacting systems. In the future, we intend to investigate the possible role of several message-external factors on convergence, such as the gender of the speakers, their age and their familiarity with each other, since these factors have been shown to play a role in the convergence of various communication elements (e.g., [36, 37]).

## 6. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 461442180.

## 7. References

- [1] M. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behavioral and Brain Sciences*, vol. 27, no. 2, pp. 169–190, 2004.
- [2] H. Clark, *Using Language*. Cambridge, UK: Cambridge University Press, 1996.
- [3] P. Wagner, Z. Malisz, B. Inden, and I. Wachsmuth, "Interaction phonology – A temporal co-ordination component enabling representational alignment within a model of communication," in *Advances in Interaction Studies: Vol. 6. Alignment in Communication: Towards a New Theory of Communication*. Benjamins, 2013, pp. 109–132.
- [4] T. Chartrand and J. Bargh, "The chameleon effect: the perception–behavior link and social interaction," *Journal of Personality and Social Psychology*, vol. 76, no. 6, pp. 893–910, 1999.
- [5] H. Branigan, M. Pickering, and A. Cleland, "Syntactic co-ordination in dialogue," *Cognition*, vol. 75, no. 2, pp. B13–B25, 2000.
- [6] S. Brennan and H. Clark, "Conceptual pacts and lexical choice in conversation," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, no. 6, pp. 1482–1493, 1996.
- [7] M. Babel and D. Bulatov, "The role of fundamental frequency in phonetic accommodation," *Language and Speech*, vol. 55, no. 2, pp. 231–248, 2011.
- [8] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proc. of INTERSPEECH*, 2011, pp. 3081–3084.
- [9] J. Pardo, "On phonetic convergence during conversational interaction," *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.
- [10] Z. Malisz, M. Włodarczak, H. Buschmeier, J. Skubisz, S. Kopp, and P. Wagner, "The ALICO corpus: Analysing the active listener," *Language Resources and Evaluation*, vol. 50, no. 2, pp. 411–442, 2016.
- [11] J. Edlund, J. Hirschberg, and M. Heldner, "Pause and gap length in face-to-face interaction," in *Proc. of INTERSPEECH*, 2009, pp. 2779–2782.
- [12] A. Schweitzer and N. Lewandowski, "Convergence of articulation rate in spontaneous speech," in *Proc. of INTERSPEECH*, 2013, pp. 525–528.
- [13] K. Shockley, D. Richardson, and R. Dale, "Conversation and coordinative structures," *Topics in Cognitive Sciences*, vol. 1, pp. 305–319, 2009.
- [14] J. Pardo, A. Urmanche, S. Wilman, and J. Wiener, "Phonetic convergence across multiple measures and model talkers," *Attention, Perception, & Psychophysics*, vol. 79, no. 2, pp. 637–659, 2017.
- [15] H. Quené, R. Orr, and D. Van Leeuwen, "Phonetic similarity of /s/ in native and second language: Individual differences in learning curves," *The Journal of the Acoustical Society of America*, vol. 142, pp. EL519–EL524, 2017.
- [16] I. Gessinger, E. Raveh, S. Le Maguer, B. Möbius, and I. Steiner, "Shadowing synthesized speech-segmental analysis of phonetic convergence," in *Proc. of INTERSPEECH*, 2017, pp. 3797–3801.
- [17] M. Wieling, M. Tiede, T. Rebernik, L. de Jong, A. Braggaar, M. Bartelds, M. Medvedeva, P. Heisterkamp, T. F. Offrede, H. Sekeres, A. Pot, M. van der Ploeg, K. Volkers, and G. Mils, "A novel paradigm to investigate phonetic convergence in interaction," in *Proc. of the 12th International Seminar on Speech Production ISSP2020*, 2020, pp. 1–4.
- [18] M. Wagner, M. Broersma, J. McQueen, S. Dhaene, and K. Lemhöfer, "Phonetic convergence to non-native speech: Acoustic and perceptual evidence," *Journal of Phonetics*, vol. 88, p. 101076, 2021.
- [19] Š. Beňuš, A. Gravano, R. Levitan, S. Levitan, L. Willson, and J. Hirschberg, "Entrainment, dominance and alliance in supreme court hearings," *Knowledge-Based Systems*, vol. 71, pp. 3–14, 2014.
- [20] J. Trouvain and K. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," in *Proc. of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, 2012, pp. 36–39.
- [21] P. Glenn, "Towards a social interactional approach to laughter," in *Laughter in Interaction*. Cambridge University Press, 2003, pp. 7–34.
- [22] J. Trouvain and K. Truong, "Convergence of laughter in conversational speech: effects of quantity, temporal alignment and imitation," in *Proc. of the International Symposium on Imitation and Convergence in Speech*, 2012, pp. 37–38.
- [23] B. Ludusan and P. Wagner, "Laughter dynamics in dyadic conversations," in *Proc. of INTERSPEECH*, 2019, pp. 524–528.
- [24] —, "Laughter entrainment in dyadic interactions: Temporal distribution and form," *Speech Communication*, vol. 136, pp. 42–52, 2022.
- [25] D. Szameitat, C. Darwin, A. Szameitat, D. Wildgruber, A. Sterr, S. Dietrich, and K. Alter, "Formant characteristics of laughter," in *Proc. of the Interdisciplinary Workshop on the Phonetics of Laughter*, 2007, pp. 9–13.
- [26] J. Hough, Y. Tian, L. de Ruiter, S. Betz, S. Kousidis, D. Schlangen, and J. Ginzburg, "DUEL: A multi-lingual multi-modal dialogue corpus for disfluency, exclamations and laughter," in *Proc. of LREC*, 2016, pp. 1784–1788.
- [27] J.-A. Bachorowski, M. Smoski, and M. Owren, "The acoustic features of human laughter," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1581–1597, 2001.
- [28] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott International*, vol. 5, pp. 341–345, 2002.
- [29] R. Corrette, "Praat Vocal Toolkit," Online: <http://www.praatvocaltoolkit.com>, 2012.
- [30] D. McCloy, "phonR: Tools for phoneticians and phonologists. R package version 1.0-7," Online: <https://cran.r-project.org/web/packages/phonR/phonR.pdf>, 2016.
- [31] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: <https://www.R-project.org/>
- [32] A. Canty, "Resampling methods in R: the boot package," *The Newsletter of the R Project Volume*, vol. 2, no. 3, pp. 2–7, 2002.
- [33] J. Pérez, R. Gálvez, and A. Gravano, "Disentrainment may be a positive thing: A novel measure of unsigned acoustic-prosodic synchrony, and its relation to speaker engagement," in *Proc. of INTERSPEECH*, 2016, pp. 1270–1274.
- [34] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotions, qualia, and consciousness*. World Scientific, 2001, pp. 426–443.
- [35] U. Cohen Priva and C. Sanker, "Limitations of difference-in-difference for measuring convergence," *Laboratory Phonology*, vol. 10, no. 1, p. 15, 2019.
- [36] L. Namy, L. Nygaard, and D. Sauerteig, "Gender differences in vocal accommodation: The role of perception," *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 422–432, 2002.
- [37] M. Nilsenová and P. Nolting, "Linguistic adaptation in semi-natural dialogues: age comparison," in *Proc. of the International Conference on Text, Speech and Dialogue*, 2010, pp. 531–538.