THE PROSODY OF MOTIVATION: FIRST RESULTS FROM AN INDOOR CYCLING SCENARIO

Lisa Viktoria Skutella, Luise Süssenbach, Karola Pitsch, Petra Wagner

Center of Excellence Cognitive Interaction Technology (CITEC), Universität Bielefeld
Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld
skutella@uni-bielefeld.de

Abstract: Intuitively, prosody seems to play a key role in what constitutes a „motivational speaking style". In this study, the phenomenon of motivational prosody is studied within dyadic interactions between an instructor and a trainee during a session of an indoor cycling class. This scenario is characterized by a strong usage of multimodal resources, a predominance of short coordinative instructions and evaluative comments expressed both verbally and non-verbally [1, 2]. These studies indicated that prosody fulfills an important function in the instructor's motivational function, but its form has not been examined thoroughly. In our exploratory empirical study on motivational contexts, we investigated to what degree instructors report on their usage of prosody as motivational means and examined their prosodic features in indoor cycling situations. As accompanying music is likely to serve as external anchor for rhythmic-prosodic structuring, only training sessions without music were investigated. Both the interviews and the prosodic analyses revealed several characteristics: (i) A high frequency of *prominent, usually accented words,* fulfills a coordinative and informative function, (ii) P*itch* and verbal rhythm mirrors posture and movement tempo, indicating an iconic relationship between pitch and movement. (iii) Rising *boundary tones* indicate the non-terminality of an exercise, falling ones the opposite. (iv) *Rhythmic structure and tempo* is synchronized with the instructor's cycling movements during, but not before and after the individual exercises. (v) Vocal effort and articulatory precision are high during explanation, but are reduced during the subsequent exercises. While having been able to filter out characteristic prosodic patterns, their effectiveness in creating and maintaining motivation remains to be shown.

1 Introduction

Previous studies [1, 2] identified *motivation* as a prerequisite for successfully carrying out physical exercises and have begun to investigate its communicative-interactional dimension and sequential organization in indoor cycling classes. In this vein, the authors have identified a set of multi-modal communicative procedures used by the instructors during such situations, such as counting, visual contact, verbal encouragement or praise. While such procedures are realized as complex 'communicative gestalts' consisting of at least talk, gesture, bodily movement, gaze orientation [3], *prosody* seems to constitute an major motivational factor in such settings. For the case of a bench press exercise [4], prosody, among other aspects, has been found to help structuring exercises and anticipate the amount of upcoming effort.

Starting from videotaped recordings of indoor cycling classes, we use the approach of Conversation Analysis (CA) [5] combined with a multimodal perspective to understand the sequential structures and interactional dynamics of the participants' actions. This way, we are able to isolate sequences of actions between instructor and trainee, focusing on those in which prosody and the successful realisation of an exercise are interdependent. In a second analytical step, we investigate the nature of this link, e.g. building on rich phonetic and phonological annotations, we qualitatively describe which prosodic means are used by the instructor during the exercises.

More specifically, we expect the instructor to use prosodic means in order to

- create a positively dynamic atmosphere during demanding exercises
- illustrate the temporal and spatial structure of the exercise, i.e. when and how the trainee is expected to move
- call attention to crucial or new information
- provide feedback on performance

Studies on prosody in live sports commenting [6, 7] have described the usage of prosodic means. One of their key insights was the strong usage of a globally raised f0 in order to express excitement. Given the task difference between *commenting on* sports events and *motivating others to perform* sports, it is unlikely that these results can be directly transferred to our scenario. Therefore, we refrain from considering a limited set of prosodic factors. Rather, we took into account a wide range of potential features in our analysis, including local and global intonation, articulation precision, prominence, tempo and rhythm.

Still, certain expectations concerning potentially relevant prosodic cues can be formulated: It is likely that a positively charged atmosphere is created by the use of prosodic markers that are commonly perceived as signalling positive emotions, representing a high degree of activation in a dimensional approach to emotion [8, 9]. It is also possible, that the prosodic structure, expressed by boundary placement and rhythmic grouping, parallels the structure of the exercises, i.e. its beginnings, ends and individual movements. In line with theories of prosodic information structure [10, 11], crucial or novel aspects of the instructions are likely to be produced with an increased amount of prominence, typically expressed by various types of pitch accents. This can aid calling the listener's attention and ensure his or her comprehension.

A special focus of our investigation lies in the question of whether prosody is used in an iconic fashion, i.e. whether pitch patterns mirror the direction of either ongoing or upcoming body movements as part of the exercise, or whether the temporal structure of the prosody parallels the movements of the exercise. This expectation is motivated by many findings of a strong prosody-gesture linkage which can be seen as extending to the whole body [12 and references therein] as well as the idea of embodied cognition, arguing that cognitive and verbal processing is intimately linked to our own bodily experience [13, 14].

2 Methods

2.1 Qualitative Interviews

Prior to subsequent analyses, we carried out qualitative interviews with three professional indoor cycling instructors. In the first part of the interview, they were asked to report on how they plan an indoor cycling session, what type of music they prefer and which are their most important aims and intentions in a session. In a second part, they reflected on their own communication in the role of an instructor, specifically in comparison with other types of everyday communication. They were also explicitly asked to describe the communicative tools they normally use, also based on authentic indoor cycling session videos.

2.2 Scenario and Data

Indoor cycling is a fitness program for groups focussing on endurance training. In natural interaction, an indoor cycling instructor and a group of normally 10-15 participants are cycling in a gym on so-called stationary indoor bikes during approximately one hour while being accompanied by music, typically expressing a notable and constant beat. Both the group and the instructor practise a set of different types of exercises on the bikes. Most indoor cycling sessions are based on a circular process and consist of three phases: warm-up, main part (different cycling exercises) and cool-down. During the main part, exercises consist of

different speeds directly linked to the tempi of the different songs and provoke different muscle and cardio loads. Instructors tend to describe the course metaphorically as a real bike tour through mountains and valleys while mountains are seen as the exercises separated by short pauses, the valleys.

One typical type of exercise is a so-called *jump*. (cf. Figure 1) Jumps exhibit a constant change between a sitting and standing position during cycling. The instructor determines both the point in time of the changes and the duration of jumps. While cycling, everyone stands up and leaves the saddle, stays up for a while and sits down again. This procedure is repeated several times and the duration of jumps, or the time participants stay away from the saddle, can differ a lot. The exercise is very dynamic and highly demanding for the participants and needs to be accompanied by clear instructions.

Figure 1: The laboratory indoor cycling scenario. The participants are engaged in a typical *jump* exercise.

As part of the interdisciplinary research project „Sozi-Rob“ [15] and in addition to studies of natural indoor cycling interactions (4 sessions), a laboratory indoor cycling scenario was built (3 instructors, 5 sessions each). Instead of the natural scenario, this consisted of a 1:1 situation of a professional indoor cycling instructor and one trainee. In one of these laboratory sessions, there was no accompanying music. This setting enables us to examine whether instructors use prosodic communicative strategies despite the lack of music, i.e. we can confirm or falsify the independence of “music-like” interaction strategies and the presence of music itself. Also, we ensured an acoustic quality reasonably suited for phonetic analysis. All subsequent analyses are based on three interactions of this laboratory setting of indoor cycling, focusing on the analysis of *jumps*. In total, 42 short jumps, and 34 long jumps taken across sessions, i.e. all three instructors were analysed.

2.3 Conversation Analysis (CA)

We used the CA approach to define sequences of action and combined it with fine-grained rich phonetic analyses to mark the coherency between prosody analysis, interaction strategies and motivation (cf. [16, 17, 18] for similar methodological approaches). Using the annotation tool ELAN, we identified interaction sequences during which the participants’ utterances and actions co-occur, thus revealing underlying interdependencies of verbal and nonverbal communicative resources. Based on the participants’ ability to follow, follow with delay or fail to follow the instructor’s announcements, we drew conclusions with respect to the success of the used prosodic strategies and analysed the rich prosodic structure in a subsequent step.

The individual jump exercises are usually initiated by a brief introduction (a), followed by the jumps themselves (b), and a termination (c):

Example 1:

(a) oke. einmal handposition ZWEI. und dann auf EINS. Das heißt wir gehen TIEF.
okay, for now hand position TWO. And then on ONE. That means we're going LOW

(b) hoch, tief, hoch, tief…
high, low, high, low

(c) und wir bleiben TIEF.
And we're staying LOW.

2.4 Prosodic Annotation

For the prosodic analysis, we decided to combine interaction analysis and a phonetic-phonological annotation. That way, we investigated not only specific prosodic patterns in the scenario under investigation but also analyse their interaction with the ongoing communication during successful, i.e. sufficiently motivating, interactions. We therefore combined a GToBI annotation [19] with a GAT2-based transcription [20].

This approach enabled us to gain knowledge of both detailed local intonation patterns marking boundaries and accents and more global prosodic patterns such as global intonation contours, but also additional parameters like loudness, prominence, pauses and speech rate, indicating dialogue actions of turn taking and repairs.

For specific phenomena, the original annotation schemata were modified or extended. For example, standard GToBI, allows no down-stepped sequences of low tones, but such cases occurred in our material. Therefore, our annotation should be regarded as descriptive phonetic rather than phonological.

Furthermore, even in the lack of music, our material contained events exhibiting a regular rhythmic structure, which was potentially meaningful for the ongoing communication. We therefore decided to complement our annotation with a description of the verbalised rhythmic pattern based on traditional musical notation (cf. Figure 3).

In order to compare verbal prosody and body motions, characteristic movements (steps, upward and downward movements during jumps) were annotated. It is important to notice that due to the cycling scenario, the bodily motions are constrained, e.g. unlike verbal prosody and unlike other types of interaction, "accentuation" cannot be easily expressed using gesture or body movements. Thus, the expected parallelisms between verbal prosody and bodily motion are constrained to issues of timing, speed and movement direction or position (up = standing, down = sitting).

3 Results

3.1 Interview Analysis

All three instructors suggested to frequently use linguistic strategies to structure their training by using particular words and phrases, together with typical voice qualities like shouting or whispering or prosodic means. They also pointed out that their communication during sessions is strongly influenced by their own personality and their individually preferred way to fulfill their role of an instructor in front of people. From this we can expect a lot of individual variation in the way that "the prosody of motivation" is performed.

Nevertheless, two of the professional instructors underlined that the communicative style used during indoor cycling is appropriate in indoor cycling contexts only. They unanimously described this interaction as "acting" with the intention to gain and secure the participants' attention by taking recourse to various roles – mostly shifting back and forth between a "strict" and an "entertaining, funny" instructor. Their main tool to perform these roles is

"playing with their voice". This points at least to the possibility of a certain prosodic style that is typical for the observed interaction and partly independent of the instructor's personality.

3.2 Typical Prosodic Patterns

Our prosodic analyses revealed various reoccurring prosodic patterns. We believe these prosodic patterns to be likely to serve the function of enhancing motivation and support success in executing individual exercises. Whether this functional relationship holds, would need to be examined in a subsequent analysis.

3.2.1 Accentuation and prominence

When comparing with other types of communication, the instructors' prosody contains a high number of prominent accents, albeit of various prosodic shapes (cf. Figure 2). Accents are always combined with a strong lengthening and locally raised intensity, i.e. they are made very prominent. Unlike common assumptions on "stress clash avoidance", they often occur in immediate precedence, probably because words and utterances tend to be short, and are frequently separated by pauses. Also unlike common assumptions about accent-word alignment, accents sometimes occur on conjunctions or particles, when these are used in order to signal the initiation of an exercise part:

Example 2: und HOCH (H* H-%) UND (H*) wir bleiben TIEF (H+L* L-%)
and HIGH (H H-%) AND (H*) we're staying LOW (L* L-%)*

In the cast majority of cases, accents accompany crucial parts of an exercise or prepare its initiation, sometimes using unusual lexical stress placements ("achTUNG", engl.: attention). That is, in line with common assumptions about prosodic information focus, accents appear to fulfill their role of highlighting new or relevant information, but do so, independently of their shape: Contrary to [10], new elements can be signaled with various types of high or low accents (cf. 3.2.2).

3.2.2 Prosodic iconicity in intonation

During jumps, locally raised intonation, often leading to high tones, mostly accompanies upward movements, while locally falling intonation leading to low tones, mostly co-occurs with downward movements (cf. Example 2 and Figure 1). Movements in pitch are often accompanied by a raised intensity on the word indicating the target position, e.g. "down-UP (H*)" or "up-DOWN (L*)". A level intonation can co-occur with both types of positions, but usually accompanies the maintenance of either a sitting of standing position.

A second case of intonation based prosodic iconicity can be seen in the global increase or decrease in pitch relative to the simultaneous tempo or heart rate. This finding duplicates Trouvain and Barry's [6] results on horse race commentaries, where pitch was find to mirror excitement and "perceived" tempo.

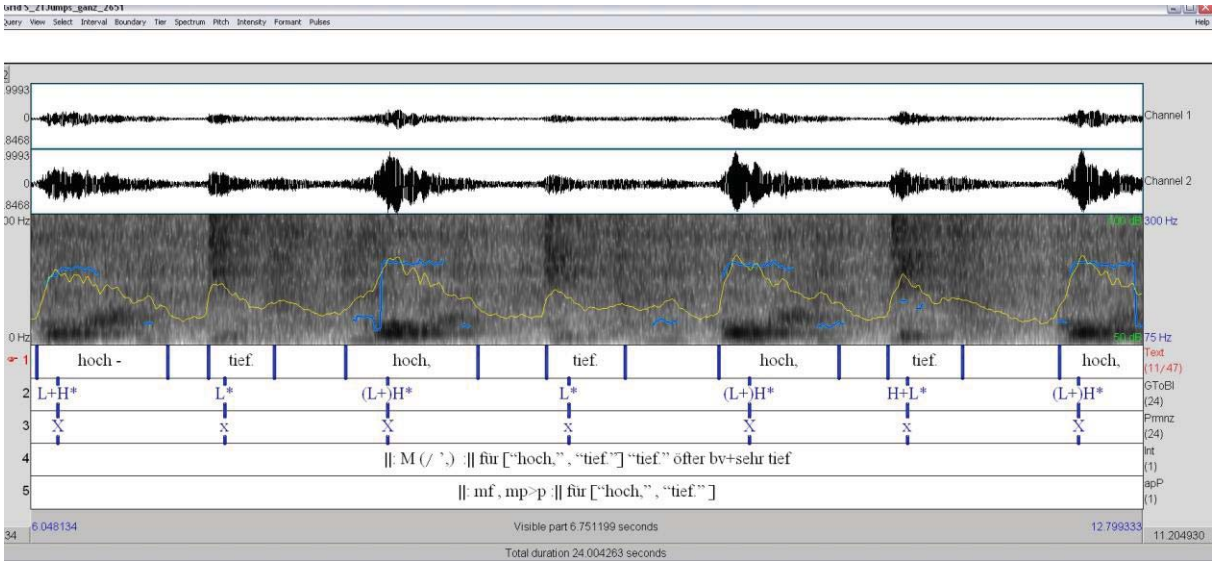

Figure 2: Accentuation and prosodic iconicity ("hoch" = upward movement synchronized with high accents; "tief" = downward movement synchronized with low accents)

3.2.3 Prosodic iconicity in rhythmic structure

Another case of prosodic iconicity can be seen in rhythmically structured pauses (cf. Figure 3) where pauses are used to synchronize verbal expression with the individual steps of the cycling movements and are strongly constrained by those in terms of duration. Rhythmic structure and tempo is synchronized with the individual cycling movements during, but not before and after the individual exercises. Apart from rhythmic pauses, the verbal instructions are sometimes rhythmically structured likewise: "less rhythmic" prosody is found in introductory and termination phases before and after the exercise. Partly rhythmic prosody can be found in utterances where a stepping example or a certain movement is explained but where it is not part of a real ongoing exercise. Clearly rhythmically structured prosody can be found during ongoing jump exercises. The instructors accompany the jumps with words like "up" or "down" and the steps with "left" or "right" in order to indicate the stepping foot. That way, they control both motion and tempo of the exercise performance.

3.2.4 Phrasing and global intonation

New or important information is not only signalled by pitch accents but often combined with a globally rising contour, while globally falling contours or boundary tones indicate the termination of an exercise. Semantic pauses separate semantically coherent utterance units while rhythmic pauses (cf. 3.2.3) are used to establish a synchrony between cycling steps and verbal explanations. Unlike rhythmic pauses, semantic pauses can differ in duration and are often independent of the simultaneous cycling movements.

3.2.5 Articulatory precision, tempo and voice quality

Vocal effort and articulatory precision are high in phases of explanation, but reduced during the subsequent exercise itself. Articulation is generally clear but can differ strongly in speed. Due to the demands of rhythmic synchrony of movement, it is sometimes necessary for the instructors to squeeze a lot of information into a relatively short interval. Rhythmic synchrony thus overrides the imperative of clear, accented articulation of new information. In general, both voice quality and pitch register varies much, often apparently serving the purpose to gain and maintain the participant's attention.

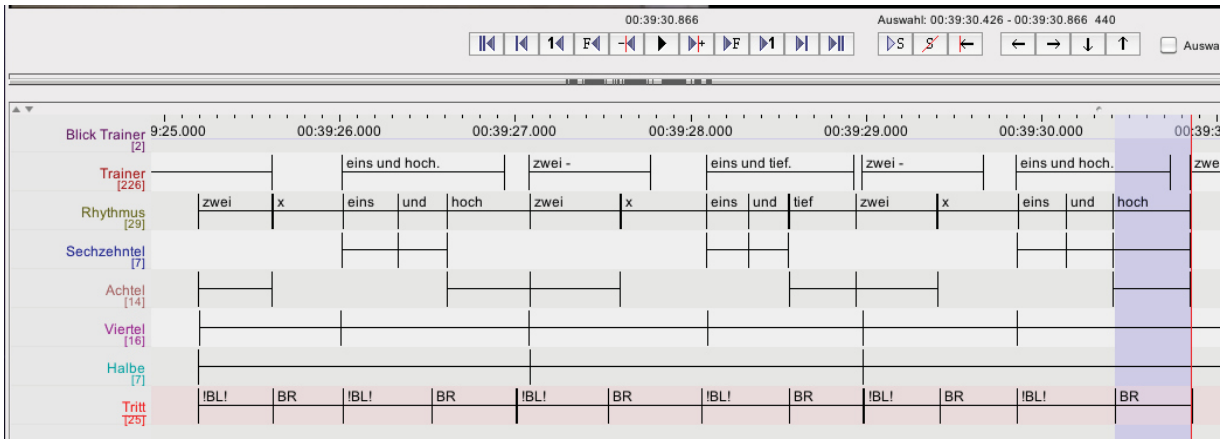

Figure 3: Synchrony and rhythmicity of verbal instructions. Tiers depict verbal instructions ("Trainer"), beats of verbal instructions ("Rhythmus"), a musical analysis ("Sechzentel" = 1/16, "Achtel" = 1/8, "Viertel" = 1/4, "Halbe"= 1/2,) and left and right cycling movements („Tritt").

4 Discussion

We find that our instructors are aware of their usage of prosodic means to structure indoor cycling training sessions (based on interviews), and have been able to isolate a set of prosodic means serving this purpose (based on videotaped recordings of indoor cycling sessions). We find that prosodic phrasing and pause construction serves to structure information into semantic packages that participants can understand easily.

New information is very systematically marked by prominence lending prosodic cues such as pitch accentuation, duration and intensity. It is interesting that information focus is marked very consistently and we find that this might be a scenario specific feature. However, unlike in [10], no clear-cut relationship between accent shape and the amount of "givenness" was found. On the contrary, this type of accentual shaping was completely overridden by the strong iconic usage of pitch, which systematically followed the directions of exercise movements. Also, the movements of the cycling exercise were synchronised with the rhythmic shape of the verbal instructions during exercises. This outcome furthermore points towards an embodied relationship between prosody and simultaneous bodily action.

Although the prosodic structure leaves much room for individual variation, all instructors unanimously took recourse to the prosodic devices described above. Apparently, prosody serves as a tool to manipulate and shape the exercise, to gain attention and maintain motivation. However, the effectiveness of this strategy remains to be shown. As our laboratory scenario clearly differed from authentic indoor cycling sessions that take place in groups and are accompanied by music, it is unclear whether our results are still valid "in the wild". Therefore, a comparison with a more natural scenario would be an important next step. With more participants, we expect a wider range of prosodic designs, more interruptions and more changes between the special prosodic style found in jump exercises and "normal conversation". In such an investigation, the impact of music needs to be attended to. Music can be expected to lighten the "workload" of the instructors' as there is less necessity to use prosody for rhythmic structuring. Rhythmic iconicity may thus become less clear in the presence of music, while the iconic usage of pitch should persist. Besides, the general song structure should support both the instructor and the participants as it might deliver useful cues to anticipate the beginning or termination or termination of an exercise. This may potentially influence boundary marking or accent placement, as they would become less crucial.

5 Acknowledgments

L. Süssenbach acknowledges the support by the German Aerospace Center (DLR) - support code 50RA1023 - with funds from the Federal Ministry of Economics and Technology

(BMWi) due to resolution of the German Bundestag. Karola Pitsch acknowledges the financial support form the Volkswagen foundation (Dilthey fellowship).

6 References

[1] Süssenbach, L. & Pitsch, K. (2011). Interactional Coordination and Alignment: Gestures in Indoor Cycling Courses. *GESPIN Gesture and Speech in Interaction*, Bielefeld.

[2] Süssenbach L (2011). Motivation im Sport: Eine konversationsanalytische Rekonstruktion interaktiver Motivationsverfahren im "Indoor Cycling". Masterarbeit, Universität Bielefeld, Germany, (*unpublished*).

[3] Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10), 1489-1522.

[4] Schmitt, R. (1990): Bodybuilding: Eine spezifische Form geselliger Vergesellschaftung. In W. Kleine & W. Fritsch (Eds): Sport und Geselligkeit. *Beiträge zu einer Theorie von Geselligkeit im Sport*. Meyer & Meyer Verlag, Aachen, 153-195.

[5] Sidnell, J. & Stivers, T. (Ed.). The Handbook of Conversation Analysis. (2013). The handbook of conversation analysis. Chichester, West Sussex, UK: Wiley-Blackwell

[6] Trouvain, J. & Barry, W.J. (2000). The prosody of excitement in horse race commentaries. *ISCA Workshop in Speech and Emotion*, Belfast, 86–91.

[7] Kern, F. (2010): *Speaking dramatically. The prosody of live radio commentary of football matches,* in: Barth-Weingarten, Dagmar et al. (eds.), Prosody in interaction, Amsterdam, 217-238 (= Studies in discourse and grammar; 23).

[8] Kehrein, R. (2002). *Prosodie und Emotionen*. Tübingen, Niemeyer.

[9] Schröder, Marc (2003). Speech and Emotion Research: An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis. Dissertation, *PHONUS 7*, Research Report of the Institute of Phonetics, Saarland University.

[10] Baumann, S. & A. Riester (2013). Coreference, Lexical Givenness and Prosody in German. In J. Hartmann, J. Radó & S. Winkler (Eds.). *[Special Issue "Information Structure Triggers"]*. Lingua 136, 16-37.

[11] Kügler, F. & Féry, C. (2008). Pitch accent scaling on given, new and focused constituents in German, *Journal of Phonetics*, 36(4), 680-703.

[12] Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication* 57, 209–232.

[13] Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press, Chicago, IL.

[14] McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, IL.

[15] Süssenbach, L., Riether, N., Berger, I., Lütkebohle, I., Kummert, F., Schneider, S., Pitsch, K., "A robot as fitness instructor: Towards an interactive motivation model" (in prep.)

[16] Kelly, J., Local, J. 1989. On the use of general phonetic techniques in handling conversational material. In: Roger, D., Bull, P., (eds), *Conversation: An Interdisciplinary Perspective*. Clevedon: Multilingual Matters 197–212.

[17] Couper-Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In *Prosody in conversation: Interactional studies* (Vol. 12, pp. 11-56). Cambridge University Press.

[18] Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, *46*(1), 8 – 23

[19] Grice, M. & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, 191, 267–298.

[20] Selting, M. et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT2). *Online-Zeitschrift zur verbalen Interaktion*, 10, 353–402.