CMST 24(1) 29–41 (2018) DOI:10.12921/cmst.2018.0000006

Methods, Tools and Techniques for Multimodal Analysis of Accommodation in Intercultural Communication

Maciej Karpiński, Katarzyna Klessa

Institute of Linguistics, Faculty of Modern Languages and Literatures
Adam Mickiewicz University in Poznań, Poland
E-mail: {maciejk, klessa}@amu.edu.pl

Received: 07 Arpil 2017; revised: 05 November 2017; accepted: 16 January 2018; published online: 31 March 2018

Abstract: The holistic approach to interpersonal communication in dialogue, involving the analysis of multiple sensory modalities and channels, poses a serious challenge not only in terms of research techniques and methods but also from the viewpoint of data infrastructure. In the Borderland project, the process of communicative accommodation is studied in young people in the intercultural context of the Polish-German boundary region. In order to collect, annotate, and analyse research material, a new technical and analytic infrastructure has been developed. Centred around a data management system, it incorporates well-known annotation and transcription tools along with custom-designed transcription and annotation tagsets and procedures. Data and metadata formats are designed to enable further corpus-based analyses of accommodation-related processes, mostly in the paralinguistic domain of prosody and gestures. Data formats used in the project ensure wide interchangeability and usage of almost any analytic software. Previously tested methods of quantitative accommodation analysis are adjusted, supplemented with new custom procedures, and applied to each channel under study as well as to the cross-modal (e.g., prosody - gesture) accommodation processes.

Key words: communicative accommodation, intercultural communication, multimodal linguistic resources, multimodal data and metadata management

I. INTRODUCTION

Mutual, interactive adjustment of dialogue parties is one of the most important aspects of spoken dialogue [1]. Dialogue participants adjust mutually to each other in a range of complex and often unconscious processes. Their results can be observed in the behaviour of communicating humans: in language use, including lexical [2], syntactic and stylistic [3] or pragmatic choices, in speech prosody ([4–6]), gestural behaviour (hand gesture, body posture, head movements, etc. e.g., [7, 8], body posture [9], facial expression), as well as in their physiological parameters. It can also be observed between modalities (e.g., gestures and prosody, see [10, 11]). The degree and direction of adjustment may depend on a number of factors, including individual traits of the participants, their psychological and physiological condition, situation context, eye contact, as well as culture-related factors [12].

Due to communicative accommodation, behaviour of dialogue parties may change in some typical ways (cf. [13]):

1. increasing similarity (e.g., while talking, people start to lean towards each other or use similar gestures, both start to speak in a higher voice or louder, and the objective measures of their behaviour are increasingly similar)
2. decreasing similarity (one person speaks increasingly louder in the course of conversation, unlike the other one who speaks increasingly quieter – objective measures of their behaviour are more and more divergent)
3. keeping a constant relation (correlation) (e.g., one person starts to speak faster and the other does the same

> but they never achieve the same speech rate: objective measures of their behaviour remain in similar proportions; alignment does not necessarily involve symmetry [14].

Mutual accommodation influences the flow of dialogue and the shape of each conversational contribution as well as its perception. It also has some prognostic, diagnostic and applicative value as a predictor of a communicative success [15–19].

Among first attempts towards defining the process of accommodation and incorporating it into the linguistic theory of communication was the Speech Accommodation Theory and later the Communication Accommodation Theory ([20–22]). Influential publications in the 1980s and 1990s confirmed that language is a collaborative, interactional activity and should be studied as such ([23, 24]). Building on some earlier contributions (e.g. [4, 25, 26]), as well as the above mentioned works of Giles and Smith, Pickering and Garrod [27] proposed a new conceptual frame for understanding the interactive aspect of dialogue. They distinguished between "co-ordination" and "alignment", the former referring to the physical aspects of communication while the latter operating on the level of mental representations. Alignment was also conceptualised as based on priming ([27]; but cf. [14] or [28]). Simultaneously, other notions and terms were introduced by the research community to refer to related phenomena (e.g., entrainment, synchrony, convergence) which sometimes obscured the key issue under study rather than helped to describe it more precisely (see e.g., [13]).

Quantitative analysis of accommodation processes is a highly demanding task that involves a range of technological and methodological issues, starting with data collection and management, through workflow management, data analysis and interpretation. Managing the workflow in the course of multimodal data collection, annotation and analysis is difficult not only by reason of the corpus sizes but also their internal complexity, which is a natural consequence of the corpus design criteria. These problems may become especially complex and bothersome in the case of intercultural studies that involve representations of different languages or dialects, diverse recording scenarios resulting in a variety of speaking registers, not to mention idiosyncratic differences. The factors coming into play in the case of multimodal analysis of accommodation in intercultural communication are often extracted by means of at least several different software tools operating on the same sets of data (e.g., speech and gesture annotation tools, metadata forms, feature extraction and analysis tools). Consequently, one of the important tasks that emerge is to achieve the assumed goals and co-ordinate the workflow in a way that ensures data consistency while preventing data loss. Satisfying such requirements may pose a significant challenge without specialized management software and ensuring sufficient interoperability support between the tools used (cf. also [29]). Data analysis methods, on the other hand, require simultaneously an individual approach to each modality and communication channel, determined by its peculiar character, while on the other hand, they are expected to provide results which may be compared across modalities and channels, and – going even further – ensure the possibility of cross-modal phenomena exploration.

Laboratories and research groups work on their own solutions for multimodal communication data collection and management, mostly in the context of particular projects or integrate existing tools (e.g., [30]). Others provide more flexible approaches. For example, Red Hen Lab consortium provides comprehensive organisational and technical infrastructure for massive dataset management, annotation and exploration [31]. Another example is the MOCA-A system for multimodal oral corpora administration and analysis, developed by Daniel Alcón López [32], that integrates data management and analysis including motion tracking.

In Section 2 of this contribution, we provide a concise description of the design and implementation of the Borderland multimodal corpus of Polish and German speech. Section 3 reports on the methods, techniques and solutions applied (as newly defined or adjusted) in the course of the corpus development and analysis. In Section 4, selected dimensions of communicative accommodation relevant to the Borderland project data are distinguished and discussed in the context of instrumental analyses of their manifestations as related to speech prosody, gesture, lexical alignment, as well as cross-modal interactions. Section 5 provides a conclusion and proposes selected applications of the results of the studies on the multimodal accommodation in intercultural settings referred to in the present paper.

II. BORDERLAND PROJECT: ACCOMMODATION IN INTERCULTURAL SETTINGS

Intercultural communication has become a widely studied and discussed topic with growing human mobility (e.g., affordable flights), state boundaries opening and rapidly increasing migration processes. Not only language issues but also cultural differences may strongly influence and often impede the process of intercultural communication. The ability to analyse, diagnose and deal with such communication problems becomes indispensable for an increasing range of officers and volunteers engaged in managing or supporting immigrants.

The aim of the Borderland project is to document and analyse the phenomena related to interpersonal communication in the region of Słubice (Poland) and Frankfurt (Oder), Germany. The main research focus is on paralinguistic phenomena (prosody and gestures) and accommodation within

the communication process together with the relevant communication strategies. Frankfurt (Oder) and Słubice, sometimes collectively referred to as "Słubfurt", constitute an investigation and testing ground for fieldwork analyses that represents an exceptionally close co-existence of cultures, languages, and social groups connected by their common history. The project goals include the studies of the dynamics of the processes of mutual accommodation in the paralinguistic domain, especially in prosody and gestures. The studies are based on a new corpus of audiovisual recordings, designed and collected for the Borderland project. Its major part includes collaborative and competitive task-oriented dialogues between pupils of secondary schools in Frankfurt (Oder) and Słubice.

This choice was motivated by the fact that most of those young people did not have an opportunity to stay away from their home towns for a longer period of time and although they were almost on the verge of communicative maturity, their communicative experience was limited to close neighbourhood. Three categories of pairs are recorded: Polish-Polish, German-German, and German-Polish. Dialogue scenarios were extensively tested and adjusted before their application in the project so that they engage dialogue parties and evoke spontaneous communicative behaviour with possibly rich gesturing. First scrutiny of the corpus confirms that these aims have been achieved.

The design of the recording setup was a crucial part of the fieldwork preparation. The aim was to achieve possibly high quality in adverse acoustic conditions while keeping the recording procedure as non-invasive as possible (e.g., avoiding head-on microphones or camcorders directly facing speakers). Each dialogue session was recorded using two digital HD camcorders (Sony NEX-FS700R Super 35), one facing each speaker, as well as an independent portable digital audio recorder (Roland R26) with two condenser microphones (AKG C4000) with almost flat frequency response in the range 50Hz – 1kHz. As a result, each session data comprises two HD video 25 fps streams and two 44.1Hz/16bit audio streams. For further processing, video was downgraded to 640 x 480 frame size. All the files were synchronised and annotated using multilayer annotation specifications as described below.

III. QUANTIFICATION AND DESCRIPTION OF COMMUNICATIVE BEHAVIOUR: METHODS, TECHNIQUES AND SOLUTIONS

III. 1. Fundamental issues of annotation of communicative behaviour

The notions of transcription (orthographic, phonetic or phonemic) and annotation of speech are closely related. In both cases, the underlying idea is to categorise human behaviour as realisations of linguistic units, e.g. phones, syllables, words, phrases, or units of interaction. It is important that annotation operates on a specific level of abstraction. For example, an actual portion of speech signal can be tagged as a phone ("actual sound of speech") or realisation of a phoneme (an abstract category). Tags may refer almost directly to some physical features of behaviour (e.g., the size of gesture, pitch height) or to its function (e.g., a dialogue act realised in a given utterance). Moreover, based on a certain theoretical ground, annotation may operate on a number of hierarchically structured levels, e.g. gestural phrase (a unit of gesticulation) can be divided into gesture phases and become a part of a gesture unit. The hierarchical structure of units can be embodied in the annotation template and annotation software.

In spite of enormous progress in the field of speech and gesture recognition and automated transcription or annotation, in the practice of communication studies, much of this work is done by hand, by trained annotators. Even if automated segmentation or transcription is employed, it often requires detailed proofreading and validation. In case of expert (manual) annotation, its results are often cross-verified by comparing annotation of the same stretch of behaviour made by two or more annotators. In case of smaller corpora, it is possible to have the entire material annotated by two or more annotators. With larger corpora, often only a sample of the entire corpus is annotated by multiple annotators. Annotations can be tested for similarity using a number of methods [33], including the commonly applied *kappa* test [34] which is also available as a ready-to-use function of several speech annotation tools such as ELAN [35], Annotation Pro [36] or SPPAS [37].

III. 2. Transcription and annotation

III. 2. 1. Tagsets and procedures

The description of multimodal recordings collected in the course of the Borderland project consists in providing time-aligned transcriptions and annotations of both speech and gesture. The respective tagsets and procedures have been designed with a view to deliver information on both linguistic and paralinguistic features of the communicative events.

The procedure of the annotation of audio data assumes manual transcription and segmentation on the phrase level as the first step. The basic assumptions for phrase boundary location are in accordance with the ones postulated by Karpiński et al. [38]. The next step is the expert verification of the initial transcription provided by each of the students. This step involves discussing and clarifying all problematic issues with the transcriber. Subsequent files are checked at random. As a result, each audio file available per pair of interlocutors includes five layers: one orthographic transcription layer per speaker, one layer including paralinguistic or non-linguistic label per speaker, and one common layer for

additional comments by the transcriber (also including notes on background noises). The two transcription layers (the orthographic and paralinguistic one) are the starting point for each speaker. Paralinguistic phenomena such as laughter, loud breathing, or filled pauses, are transcribed on a layer independent of the orthographic transcript in order to facilitate further automatic grapheme-to-phoneme conversion as well as segmentation into phones. All layers are time-aligned thus enabling inspection of (co-)occurrence of both linguistic and paralinguistic communicative events in the utterances of each speaker separately and as related to the interlocutor's performance. The labels defined to mark non-well formed utterances, paralinguistic or non-linguistic events in speech are used to indicate:

- incomprehensible utterance or utterance fragments
- transcription of utterances about which the transcriber is uncertain
- filled pauses, "fillers" or hesitation markers (the present labelling scheme includes the information about the event position and, where possible, the closest approximated transcription of the fillers)
- cough
- laughter
- sighs
- breaths
- groans.

Gestural behaviour is a complex, multidimensional process, difficult to track and represent formally and symbolically. Until recently, precise recording of gestures (e.g., their shape, range, speed) was possible only using expensive motion capture equipment. Although increasingly elaborated, these methods remain not fully reliable and relatively invasive (markers attached to the body, numerous camcorders, high sensitivity to lighting conditions, etc.). Therefore, manual gesture annotation based on video recordings is still widely employed in research.

Gesture annotation in the Borderland project is based on the traditional model of Gesture Phrase [39], [40], [4q] and some available gesture annotation schemes, e.g., [42], [43], and comprises:

- Gesture Unit (including at least one Gesture Phrase)
  - Gesture Phrase
  - [Gesture Function – refers to the entire gesture]
  - [Representation Technique – refers to the entire gesture]
    * Gesture Phase
      · Preparation
      · Pre-stroke Hold
      · Stroke
        - [GestureSize (small, medium, large)]
        - [Gesture location (GestureSpaceVer, GestureSpaceHor)]
        - [Hand shape (OpenPalm, Fist, OneFinger, ManyFingers)]
    * Post-stroke Hold
    * Retraction.

Units in brackets are not part of the standard gesture model. They are additional features of gestures that are known or hypothesized to be involved in interpersonal alignment (e.g., [8]). Additionally, head movements were annotated in a simplified way using a single layer and two potential labels, representing horizontal and vertical movement ("nod" and "shake"). Nevertheless, more detailed schemes for head motion are available and can be applied if necessary (cf. [44], [46]).

In the present project, Gesture Phrase and Gesture Unit segmentation made by annotators is always checked by an expert. Annotations on lower levels are validated by random scrutiny as well as by *kappa* test with the same, randomly selected portion of the material (ca. 10%) annotated by two independent annotators.

III. 2. 2. Tools

The linguistic descriptions and analyses of the Borderland corpus are carried out using ELAN and Annotation Pro. ELAN is a widely used free software for time-aligned video annotation [35]. It supports multi-layer time-aligned annotation and a range of editing and analytic functions as well as the possibility of designing hierarchical user-tailored annotation templates. Template hierarchy can be based on various temporal and logical relations. Moreover, user-defined lexicons can be linked to each type of the annotation layer to facilitate annotation by offering immediate access to the lists of potential labels (Fig. 1).

Nevertheless, ELAN does not display speech signal as a spectrogram and its time resolution as well as segmentation precision is lower than required for a range of phonetic-acoustic analyses of speech. Since one of the assumptions of the present analyses was to inspect the cross-modal interactions using both gesture-based and phonetic-acoustic information, it became necessary to employ a separate, specialised tool for the purpose of creating multilayer annotations reflecting both segmental and suprasegmental levels of utterances. Annotation Pro [36] was selected for annotation and analysis of linguistic and paralinguistic features in speech. Annotation Pro not only enables inspecting configurable spectrogram views of the speech signal for the purposes of annotation tasks, but can be used as a perception experiment tool for

testing hypotheses based both on discrete and continuous rating scales. By default, the results of the perception tests are stored as part of the annotation files, in segments belonging to a standard annotation layer, and can thus be time-aligned with the remaining segments used within the

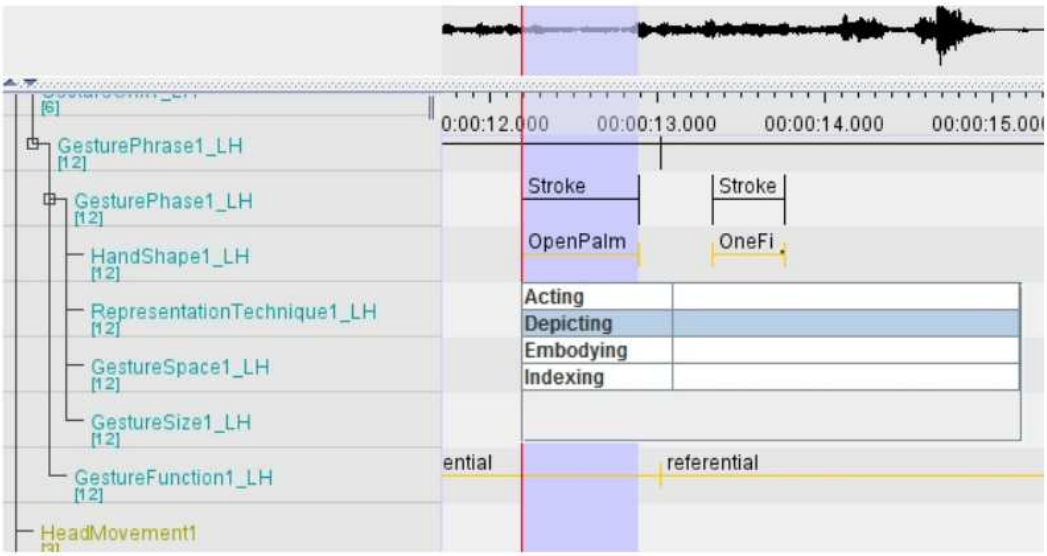


Fig. 1. A subset of gesture annotation layers defined in ELAN
(four independent sets are available: for each hand of each of two dialogue parties)

annotation file. With a view to support data exchange between ELAN and Annotation Pro, and thus enable the analyses of cross-modal accommodation for the Bordlerland resources, the import - export module of Annotation Pro has been extended so that it currently includes the ready-to-use conversion options of .EAF and .ANT formats (the native annotation formats of ELAN and Annotation Pro, respectively).

The automatic phonetic time-alignment procedure is based on using external tools either integrated with Annotation Pro (PolPhone [46]; Salian [47]) or used outside of the present system (SPPAS [37]; WebMAUS [48]). The resulting transcriptions and segmentations are subsequently imported back to Annotation Pro as new annotation layers for the purpose of the final manual verification of the obtained boundary positions and labelling. This step is particularly important in the present context for two main reasons: (1) the conversational, spontaneous character of the recorded speech (numerous disfluencies, incomplete utterances, hesitation markers, fillers or other speaker noises, non-trained speakers), (2) the technical quality of the recordings (features resulting from fieldwork conditions, e.g., background noises of various type). For these reasons, the performance of the automatized procedures is (as expected) more error-prone than in the case of well-formed utterances recorded by professional speakers in laboratory recordings conditions.

III. 3. Corpus MINI: an integrated software solution for multimodal data and annotation management

Processing, annotation, and analysis of the multimodal data collected within the Borderland project requires collaboration of researchers representing various scientific backgrounds, goals, and technical qualifications. In order to support their effective collaboration and satisfy the diversified requirements, a new tool named Corpus Mini has been developed for the purposes of data and annotation management. Corpus Mini is a dedicated annotation database manager which makes it possible to deal with multimodal data and their corresponding metadata in a controlled manner (see also [49]). The system has been designed using client-server methodology and Microsoft SQL Server. The client application (created in Visual Studio .NET WinForms C#) can be installed at any desired number of personal computers connecting to the same central database where all linguistic data and metadata are stored. A vital functional prerequisite for the system was to ensure remote access to the corpus data and possibility to perform the annotation and analyses for both audio and video files even under adverse work conditions, e.g., weaker or unstable Internet connection, older versions of the operating system (the users connected to the database using their own personal computers). A model of the system architecture is shown in Fig.3.

The main features of the Corpus MINI solution include:

- integrated annotation of multimodal data (one or more associated video and audio files) with ELAN and Annotation Pro
- annotation file management
- creating user accounts for database users
- access permissions management
- assigning selected files to particular users
- remote or local access to database

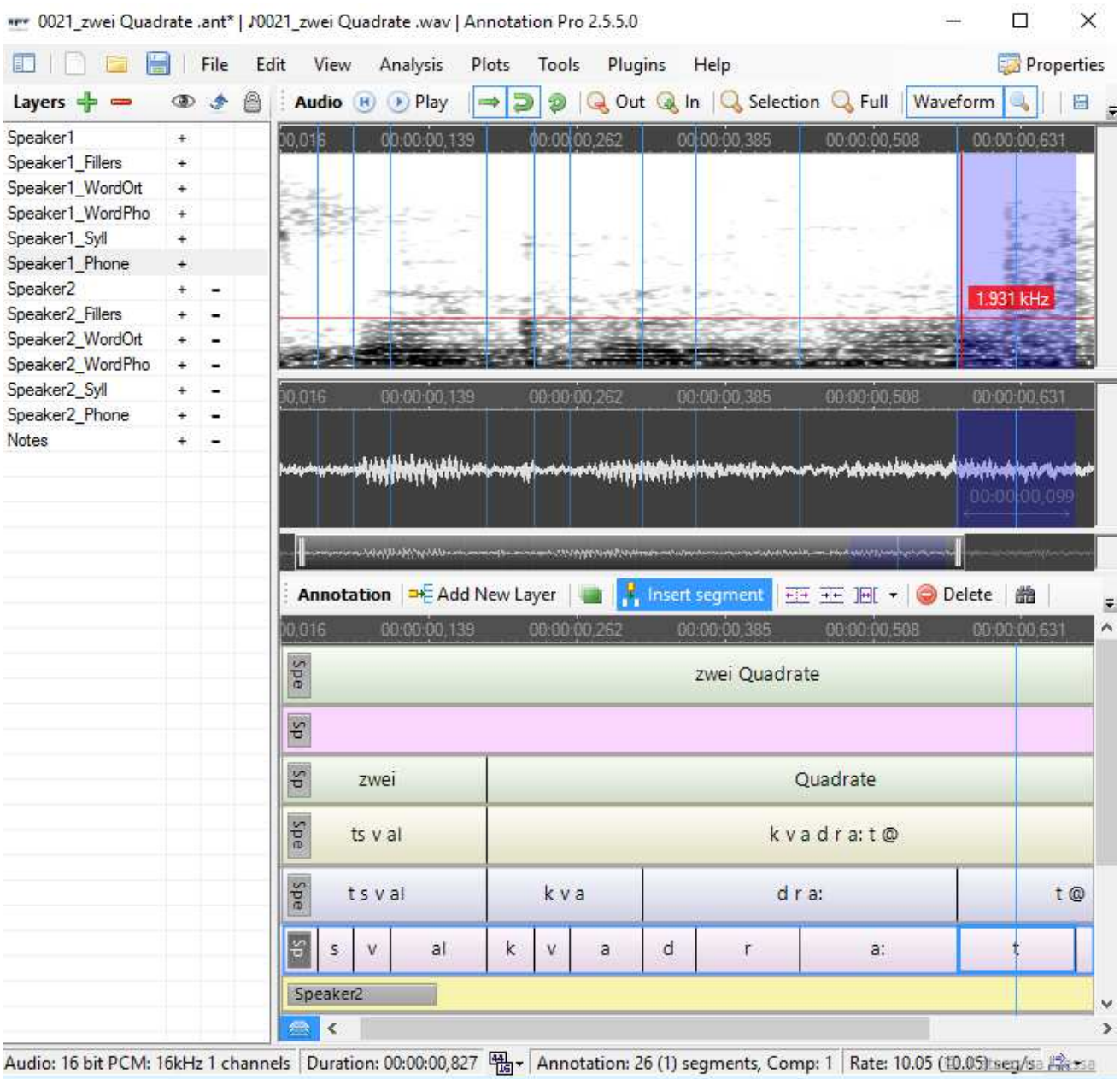


Fig. 2. A sample of multilayer annotation of a German utterance “zwei Quadrate” displayed in Annotation Pro (word, syllable, and phone level annotations and transcriptions)

- automatically blocking data currently used by another user
- basic work statistics
- data searching and filtering
- possibility of attaching additional files related to annotations (e.g., documentation files)
- using centralized data storage on a server
- flexible metadata configuration
- bulk annotation file import
- configurable metadata form
- bulk metadata sheets import.

Corpus Mini helps to supervise annotation and analysis of large corpora of multimodal recordings, restricts simultaneous usage of the same multimodal data by different or unauthorized users, and therefore prevents data loss. Annotation files can be assigned to one or more individual users, and each of the users can only see and access data assigned to their own account. In Corpus Mini, it is also possible to inspect annotation workflow and progress thanks to status flags available in the interface (e.g., 'done', 'accepted', 'locked').

Apart from routine management tasks, the system allows authorized users to flexibly define the types of metadata stored in the database. In the present version of the tool, a metadata sheet composed of up to twenty columns representing desired metadata types can be used. The columns specifications can be freely defined by the user, and can refer to the features of the multimodal data, the speaker information, recording environment, equipment used or any other. The metadata sheets can also be created in external tools (e.g., online forms filled in by the interested parties) and imported from .CSV file formats to the relational database.

IV. DIMENSIONS OF COMMUNICATIVE ACCOMMODATION: ANALYTIC APPROACH

In the Borderland project, several approaches are combined to automatically measure local and global variability in the occurrence of various phenomena during conversation for particular dimensions of accommodation (Sections 4.1 to 4.3) as well as those occurring across modalities (Section 4.4). One of the techniques applied in all the domains and cross-modally is the moving frame method. The method was inspired by the approach of Kousidis et al. [50], [51] assuming the usage of a time frame of a given width to observe communicative convergence in dialogues. Different choices of frame width (e.g., from 10s, 30s, 60s or even longer in exceptional cases) may result from varying distributions of speech chunks over time or different frequency of occurrence of paralinguistic or non-linguistic events. The frame sizes applicable for spontaneous or quasi spontaneous conversational speech usually need to be longer than for read speech or public talks prepared and rehearsed in advance (to avoid an excessive proportion of frames containing no segments in case of utterances accompanied by a significant number of pauses). The moving frame method has been implemented by the present authors as an Annotation Pro plugin (a C# script extending the basic functionality of the programme, and enabling applying automatized procedures to large amounts of data). The initial versions of the plugin was tested and used beforehand to study the variability of speaking rates by [6] and [49]. Considering the multiple dimensions and variables to be relevant in the present context, it was decided to adjust and apply the plugin for studying the rate of occurrence of any types of segments included in the annotation layers. The segments may include not only transcriptions of linguistic and paralinguistic events or gesture labels but also the results of measurements performed basing on the annotations, representing e.g., various rhythm metrics or pitch representations. Consequently, both local and global variability of the features in question can be tracked and analysed **within** and **between** the domains described further in this section, namely:

- the prosodic domain (particularly rhythm and pitch)
- gesture domain (hand gestures and head movement)
- lexical domain.

IV. 1. Prosody: duration (rhythm) and pitch

Components of speech prosody are frequently mentioned as involved in the process of mutual communicative accommodation, e.g., [52]. They involve variability on both segmental and suprasegmental level in the domains of time, frequency and intensity.

Definitions of speech rhythm are often formulated using the notions of base units of rhythm and their alternation, iteration or isochrony. Although the primary association with speech rhythm may be the variability in the domain of time, in fact the nature of the concept is more complex [51], and its correlates are reported to include features from the domains of frequency and, in particular, intensity. The base units vary between the existing approaches and include or combine, e.g., the syllable, the rhythmic stress group (e.g., the foot) with its syllable components or the intervals between vowels and consonants occurring in the spoken utterance [54–59]. The studies of speech rhythm may involve investigation of both local and global variability over time and take into account the relationships between the rhythmic patterns and other features such as speaking rate, e.g., [60, 61].

The temporal dynamics of speech and the corresponding rhythmic patterns have been found to play a significant role in the processes of communicative alignment [4, 50]. The dynamics can be highly affected by the type of speech and communicational context. Spontaneous conversations are characterized by a wide range of discontinuities or distractors managed in quite individual ways by the interlocutors who, however, may still tend to exhibit mutual communicative alignment in the time domain [6]. The discontinuities may impact not only the time domain but also practically all the remaining domains represented in the phonetic-

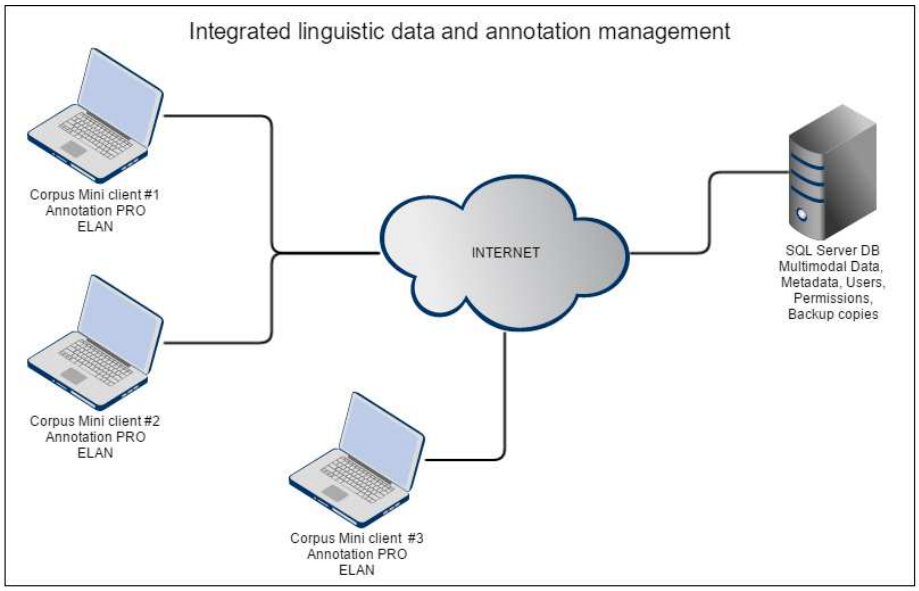


Fig. 3. Integrated linguistic data and annotation management with Corpus MINI

acoustic characteristics of the speech signal, e.g., [62] or [13].

The Borderland speech materials make it possible to extract a number of features for the purposes of measurements and evaluation of the role and significance of speech rhythm in communicative alignment. The basic durational statistics (on the phone, syllable, word and phrase levels) as well as average speaking rates (excluding or including pause times) are accessible directly in Annotation Pro based on the multilayer annotations. Furthermore, automatic extraction of a set of rhythm metric parameters has been implemented in the form C# scripts as Annotation Pro plugins, see also: [6], [63]. Among others, the plugins support measurements of normalized pairwise variability index of syllable durations, nPVI [58], interpausal time group analysis, TGA [64], duration quadrants [66]. All the plugins are publicly available for research purposes at: annotationpro.org/plugins/ or directly from the authors of the present contribution.

Pitch frequency changes in speech signal are believed to come from two major sources: the linguistic system of intonation (and/or tone), and paralinguistic factors that somehow influence many parameters of speech signal [66]. These sources of variability, together with some random and speaker-specific factors, are mostly responsible for the final shape of the “melody of speech”. Linguistic intonation conveys meanings related to linguistic categories, e.g., indicates sentence category (question, statement, etc.) or helps to segment the flow of speech into linguistic units (e.g., phrases). Paralinguistic intonation indicates emotional states of the speaker, his/her attitude towards the topic or the addressee. Paralinguistic intonation is believed to be more universal but still features certain language-specific aspects [67], [68], [69].

Pitch frequency (f_0) is the main correlate of the perceived speech sound height. It is most often extracted from speech signal using self-similarity-based algorithms like the auto-correlation method, e.g., [70] but many other methods are available [71], [72]. As the relation between the pitch frequency trace and its mental representation as "the melody of an utterance" is complex, few researchers dare to propose solutions for intonation perception modelling. Prosogram [73] as well as a more recent Polytonia [74] are useful tools that can help to automatize the symbolic transcription of prosody or at least facilitate categorisation decisions made by human annotators.

Pitch frequency and its variability measures for both speakers can be compared in a moving time-window to find correlation as well as its direction (e.g. increase or decrease) for specific stages of dialogue. A range of different frame and step sizes are used to get better results on different types of phenomena [6], [11]. Using the moving frame approach one can also easily analyse other pitch-related parameters like local and global pitch variability, range, as well as directions of their changes. Results for each pair of speakers are analysed for covariability correlation measures as well as regression analysis (including multiple regression wherever required).

IV. 2. Gesture: hand gestures and head movement

The importance of body movement in communication is no longer questioned [75–77]. Hand gesture, head movements, facial expression, and body position changes are often communicatively intentional and, if not, they may still be communicatively relevant. They are often used not only to convey meaning to the addressee but also facilitate formulation of an utterance and self-expression in general [78–

81]. Gestures are used both as "stand-alone" conversational contributions (dialogue turns) and as a means of feedback. They do not form a homogeneous category. Some of them may occur in the absence of speech and bear "independent" meaning codified in the context of a given culture (e.g., emblems), while some may function only as parallel to speech. Hand movements synchronised with speech and not bearing meanings of their own but rather somehow enhancing the meaning of the spoken utterance, are often referred to as "gesticulation" (Kendon's continuum, see: [82]).

All categories of gestures may be involved in the processes of accommodation. The studies carried out so far have shown that some of their parameters (properties, features) of gestures are more "prone to entrainment" than others (e.g., [8] see also: [11]). Nevertheless, this issue definitely requires a deeper insight and further exploration.

Gesture description based on multiple layers (as proposed in 3.2) introduces more complexity into the process of annotation but facilitates further analysis. One may operate independently on selected types of units (e.g. gesture phrases or phases) as well as their selected features (e.g., representation mode, gesture size and location in the gesture space). All or some of the features may also be treated jointly as a bunch of variables.

Among possible approaches to gestural accommodation analysis, two seem to represent major tendencies. In the first case, the frequency of temporal co-occurrences of similar gestures (or similar features of gestures) can be considered as a measure of accommodation: one speaker adopts gestures of the other, and (potentially) vice-versa. One may hypothesize that certain gestural features may occur in gesture sequences due to the process of entrainment more often than by chance. For example, large gestures performed by one speaker may be followed by large gestures by the other, and this may mean that the feature of size was subject to accommodation. Similar hypotheses can be put forward regarding gesture categories. For example, using more pointing gestures by one speaker may result in increased usage of this category of gestures in the other. This kind of analysis can be achieved using time-window based techniques mentioned above as gesture annotations from ELAN can be imported into Annotation Pro and processed using similar plugins.

Another approach is focused on the structural similarities in gesture usage and may be considered as a type of syntactic accommodation analysis. Although there are strong arguments on the possibility of "gestural syntax", the assumption that gestures may tend to occur in certain sequences seems to be uncontroversial. Accordingly, it seems reasonable to use n-gram-based approach to look for certain more frequent gesture sequences and then track how their occurrence changes in the course of dialogue. This can be achieved using ELAN itself as it offers n-gram analysis. Moreover, with ELAN it is possible to look for multi-layer n-grams, which means n-grams consisting of annotations from different layers. This option enables in-depth analysis of gesture structure that encompasses a range of gestural features and categories.

IV. 3. Lexical alignment

Word usage may be considered as a "linguistic" parameter as words themselves are linguistic units. Nevertheless, their choice and distribution in utterances, dialogue exchanges and entire communication events, is obviously only partially determined by the properties of a given language. Individual speaking style and stylistic devices may result in producing lexically different utterances with the intention to communicate similar meanings. Frequent usage of certain words (or avoiding them) may be regarded as "indexical", i.e. unwillingly and unconsciously revealing some features of the speaker (social background, education, profession, and others) [83, 84].

A range of methods can be employed to estimate lexical alignment and its fluctuations in dialogue. In the Borderland project, the approach described by Figiel [86] with some extensions and modifications is employed. The orthographic transcriptions of the utterances by each dialogue partner are tokenized (segmented into wordforms), and the wordforms from both speakers are joined in a single sequence in the order of their occurrence in time. They are lemmatised, i.e. converted into their basic, "dictionary" forms using one of freely available lemmatisers for the Polish and German, and analysed using frame-based methods: one based on a fixed-size time window and the second on a window containing a fixed number of wordforms from the two dialogue partners. In the first analysis, temporal distribution of the uttered words may be taken into account while in the second one, time distance is abstracted from and only sequentiality is analysed. Further, occurrences of words are counted for each subsequent time- or quantity-frame and compared regarding their numbers within respective moving frames for the two interlocutors.

IV. 4. Cross-modal alignment

With the availability of multimodal data in the well-organised form of multi-layer synchronised annotation created with interoperable tools, it becomes possible to explore accommodation and alignment between modalities using the same or adequately adjusted methods as for intra-modal processes. While speech and gesticulation show a strong tendency to synchronise in a given speaker (e.g., [76, 86]), the relation between dialogue flow within these two modalities is significantly less obvious and requires detailed exploration. Among pioneering attempts in this area is the study of prosody, gesture and gaze co-ordination conducted with the participation of the present authors [11] in which the time window approach was supplemented with multiple regression analysis. It provides better control over all the independent variables, accounting also for their cross-correlations [86]. Available results are promising but they can only be-

come significant with larger datasets that are expected to be ready to use only in the final stage of the Borderland project.

V. CONCLUSIONS

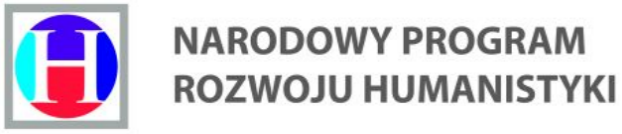
The analysis of multimodal accommodation in interpersonal communication involves a number of serious challenges related to the nature of the data as well as to research techniques and methods of dealing with a number of sensory modalities and channels. In the present text, the process of data collection, management and analysis in the Borderland project is shown, including the application of widely used annotation and transcription tools, adjusting their interoperability as well as design of an entirely new system that supports remote data management, processing and analyses, ensuring safety and fluent work flow.

Speech transcription and gesture annotation systems are tailored to the requirements of further automatized corpus-based analyses of accommodation using time-windows as well as other approaches.

The results of the studies on the multimodal accommodation in intercultural settings may find numerous applications wherever managing and supporting intercultural communication is of importance. Social and health services, education, or security alike are in need of scientifically obtained knowledge on the mechanisms of conversational alignment that can help to guide a range of communication and negotiation processes. Moreover, this knowledge can be implemented in computer-based conversational systems in order to let them tune to the users and facilitate the process of communication, taking into account entrainment-related parameters and cultural background of the user, e.g., [88].

Acknowledgements

This work was funded by the National Programme for the Progress in Humanities (NPRH12H 13 0524 82, *Language of the boundaries – the boundaries of language. Paralinguistic aspects of intercultural communication*).



References

[1] M. Karpiński, *New challenges in psycholinguistics: Interactivity and alignment in interpersonal communication*, Lingua Posnaniensis, vol. 54, issue 1, pp. 97-106 (2014).

[2] S.E. Brennan, H.H. Clark, *Conceptual pacts and lexical choice in conversation*, Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, pp. 1482–93 (1996).

[3] H.P. Branigan, M.P. Pickering, A. Cleland, *Syntactic coordination in dialogue*, Cognition, 75, pp. 13-25 (2000).

[4] R.L. *Street Speech convergence and evaluation in fact-finding interviews*, Human Communication Research, 11(2), pp. 139–169 (1984).

[5] S. Kousidis, *A Study of Accomodation of Prosodic and Temporal Features in Spoken Dialogues in View of Speech Technology Applications*, Doctoral Thesis. Dublin Institute of Technology (2010).

[6] M. Karpiński, K. Klessa, A. Czoska, *Local and global convergence in the temporal domain in Polish task-oriented dialogue*, Proceedings of the 7th Speech Prosody Conference, 20-23 May 2014, Dublin, Ireland. ISSN: 2333-2042, pp. 743-747 (2014).

[7] F. Hahn, H. Rieser, *Explaining Speech Gesture Alignment in MM Dialogue Using Gesture Typology*, [In:] P.Lupowski, M. Purver (Eds.), *Aspects of Semantics and Pragmatics of Dialogue*. SemDial 2010, pp. 99–111 (2010).

[8] K. Bergmann, S. Kopp, *Gestural alignment in natural dialogue*, [In:] R.P. Cooper, D. Peebles, N. Miyake, (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (CogSci 2012), pp. 1326–1331. Austin, T X: Cognitive Science Society (2012).

[9] T.L. Chartrand, J.A. Bargh *The chameleon effect: The perception-behavior link and social interaction*, J. Pers. Soc. Psychol. **76**, pp. 893–910, (1999).

[10] B. Oben, G. Brône, *Explaining interactive alignment: A multimodal and multifactorial account*, Journal of Pragmatics **104**, pp. 32-51 (2016).

[11] A. Czoska, K. Klessa, M. Karpiński, E. Nowikow-Jarmołowicz, *Prosody and gesture in dialogue: Cross-modal interactions*, Proceedings of 4th Gesture and Speech in Interaction (GESPIN) Conference, Nantes, France, pp. 83-88 (2015).

[12] D.A. Cai, J.I. Rodriguez Adjusting to Cultural Differences: The Intercultural Adaptation Model. Intercultural Communication Studies VI: 2 (1996).

[13] R. Levitan, J. Hirschberg, *Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions*, Proceedings of Interspeech 2011, Florence, Italy, August 2011, pp. 3081–3084 (2011).

[14] R.M. Krauss, J.S. Pardo, *Is alignment always the result of automatic priming?*. Behavioral and Brain Sciences, **27**(02), 203-204 (2004).

[15] D.G. Hewett, B.M. Watson, C. Gallois, M. Ward, B.A. Leggett, *Intergroup communication between hospital doctors: implications for quality of patient care*. Social science medicine, **69**(12), pp. 1732-1740 (2009).

[16] R. Porzel, A. Scheffler, R. Malaka, `How entrainment increases dialogical efficiency`, Proceedings of Workshop on Effective Multimodal Dialogue Interfaces, Sydney 2006.

[17] M.J. Pickering, S. Garrod, *Toward a mechanistic psychology of dialogue*, Behavioral and Brain Sciences **27**, pp. 169–226 (2004).

[18] D. Reitter, J.D. Moore, *Predicting success in dialogue*, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 808–815 (2007).

[19] F. Ramseyer, W. Tschaecher, *Nonverbal synchrony or random coincidence? How to tell the difference*, [In:] A. Esposito, ed., COST 2102 International Training School, Heidelberg: Springer Verlag, 2009, pp. 182–196 (2009).

[20] H. Giles, D.M. Taylor, R.Y. Bourhis, *Towards a theory of interpersonal accommodation through language: Some Canadian data*, Language in Society **2**, pp. 177–192 (1973).

[21] H. Giles, P. Smith, *Accommodation theory: Optimal levels of convergence*, [In:] H. Giles, R.N. St. Clair, eds., Language and Social Psychology. Baltimore: University Park Press, pp. 45-65 (1979).

[22] H. Giles, N. Coupland, J. Coupland, *Accommodation theory: Communication, context, and consequence. Contexts of accommodation*, Developments in applied sociolinguistics, pp. 1-68 (1991).

[23] E.A. Schegloff, *Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences*, [In:] D. Tannen (Ed.) *Analyzing Discourse: Text and Talk*, pp. 71–93. Georgetown: Georgetown University Press (1982).

[24] H.H Clark, *Using language*. Cambridge: CUP (1996).

[25] S. Garrod, A. Anderson, *Saying what you mean in dialogue: A study in conceptual and semantic co-ordination*, Cognition **27**(2), 1987, pp. 181-218 (1987).

[26] F.J. Bernieri, R. Rosenthal, *Interpersonal coordination: Behaviour matching and interactional synchrony*, [In:] R. Feldman, B. Rimé, eds., *Fundamentals of non-verbal behaviour*, pp. 401–432. New York: Cambridge University Press (1991).

[27] M.J. Pickering, S. Garrod, *Alignment as the basis for successful communication*, Research on Language, Computation **4**(2), 203-228 (2006).

[28] M. Kaschak, A. Glenberg, *Interactive alignment: Priming or memory retrieval?*, Behavioral and Brain Sciences **27**(02), 201-202 (2004).

[29] N. Ide, J. Pustejovsky, *What does interoperability mean, anyway? Toward an operational definition of interoperability for language technology*. Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, China (2010).

[30] S. Bonacchi, M. Mela, *Practical remarks about the interoperability of the computer programmes Folker, ELAN and Praat for transcription and multimodal linguistic annotation from the user's point of view*, Journal of Multimodal Communication Studies **2**/2014, pp. 18-29 (2014).

[31] P. Caponetto, *Red Hen Lab: A further step towards multimodality* [report]. Journal of Multimodal Communication Studies **3**(1-2) (2016).

[32] D.A. López, *MOCA-A (Multimodal Oral Corpora Administration)*, on-line: http://www.hpsl.uni-freiburg.de/

[33] K. Peshkov, L. Prévot, *Segmentation evaluation metrics, a comparison grounded on prosodic and discourse units*, [In:] Proceedings of the 9th Language Resources and Evaluation Conference, Reykyavik, Iceland, pp. 321-325 (2014).

[34] J. Cohen, *A coefficient of agreement for nominal scales*, Educational and psychological measurement **20**(1), pp. 37-46 (1960).

[35] P.H. Wittenburg, A. Brugman, A. Russel, Klassmann, H. Sloetjes, *ELAN: a professional framework for multimodality research*. Proceedings of the 5th Language Resources and Evaluation Conference, Genoa, Italy, pp. 1556-1559 (2006).

[36] K. Klessa, Karpiński, M., A. Wagner, *Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features*, [In:] B. Bigi, D. Hirst, eds.. Proceedings of TRASP (Tools and Resources for the Analysis of Speech Prosody), Aix-en-Provence, s. 51-54. Aix-Marseille Université. ISBN 978-2-7466-6443-2 (2013).

[37] B. Bigi, *SPPAS-Multi-lingual approaches to the automatic annotation of speech*, [In:] K. Klessa, B. Bigi, eds., The Phonetician - International Society of Phonetic Sciences, Vol. **111-112**, pp. 55-69 (2015).

[38] M. Karpiński, J. Kleśta, E. Baranowska, K. Francuzik (Klessa), *Interphrase Pause Realization Rules for the Purpose of High Quality Polish Speech Synthesis*, Speech Analysis, Synthesis and Recognition (SASR), Szczyrk, AGH Kraków, pp. 85-89 (2005).

[39] A. Kendon, *Some relationships between body motion and speech*, Studies in dyadic communication **7**(177), p. 90 (1972).

[40] S. Kita, *The temporal relationship between gesture and speech: A study of Japanese-English bilinguals*. MS, Department of Psychology, University of Chicago, 90, 91-94 (1990).

[41] D. Mc Neill, *Gesture and Thought*. Chicago: University of Chicago Press 2007.

[42] J. Bressem, *Transcription systems for gestures, speech, prosody, postures, gaze*, [In:] C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. Mc Neill, S. Teßendorf, eds., *Body-Language-Communication: An international Handbook on Multimodality in Human interaction*, Berlin, Boston: De Gruyter: Mouton (2013).

[43] M. Karpiński, E. Jarmołowicz-Nowikow, A. Czoska, *Gesture annotation scheme development and application for entrainment analysis in task-oriented dialogues in diverse cultures*. Proceedings of GESPIN 2015 Conference, Nantes, France, pp. 161-166 (2015).

[44] S.Z. Kousidis, P. Malisz, Wagner, D. Schlangen, *Exploring annotation of head gesture forms in spontaneous human interaction*, [In:] Proceedings of the Tilburg Gesture Meeting (TiGeR 2013) (2013).

[45] Z. Malisz, M. Karpiński, *Multimodal aspects of positive and negative responses in Polish task-oriented dialogues*, [In:] Proceedings of Speech Prosody (2010).

[46] G. Demenko, M. Wypych, E. Baranowska, *Implementation of grapheme-to-phoneme rules and extended SAMPA alphabet in Polish text-to-speech synthesis*, Speech and Language Technology **7**(17) (2003).

[47] M. Szymański, S. Grocholewski, *Transcription-based automatic segmentation of speech*, Proceedings of 2nd Language, Technology Conference, Poznań, pp. 11-15 (2005).

[48] T. Kisler, U.D. Reichel, F. Schiel, C. Draxler, B. Jackl, N. Pörner, *BAS Speech Science Web Services-an update of current developments*. Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, Paper ID 668 (2016).

[49] K. Klessa, *Speech annotation mining with Annotation Pro plugins*, Wydawnictwo Rys, Poznań. ISBN 978-83-65483-20-1 (2016).

[50] S. Kousidis, *A Study of Accomodation of Prosodic and Temporal Features in Spoken Dialogues in View of Speech Technology Applications*, Doctoral Thesis. Dublin Institute of Technology (2010).

[51] S. Kousidis, D. Dorran, Y. Wang, B. Vaughan, C. Cullen, D. Campbell, C. McDonnell, E. Coyle, *Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues*, Proceedings of Interspeech 2008, pp. 1692–1695 (2008).

[52] S.D. Farley, S.M. Hughes, J.N. LaFayette, *People Will Know We Are in Love: Evidence of Differences Between Vocal Samples Directed Toward Lovers and Friends*, Journal of Nonverbal Behavior, Ausgabe 3/2013 (2013).

[53] A. Cutler, *Linguistic rhythm and speech segmentation*, [In:] J. Sundberg, L. Nord, R. Carlson, eds., *Music, language, speech and brain*, Macmillan Education UK, pp. 157-166 (1991).

[54] K.L. Pike, *The Intonation of American English*, University of Michigan Press, Ann Arbor (1945).

[55] W. Jassem, D.R. Hill, I.H. Witten, *Isochrony in English speech: its statistical validity and linguistic relevance*, [In:] D. Gibbon, H. Richter, eds., *Intonation, accent and rhythm*. Studies in Discourse Phonology **8**, pp. 203–225 (1984).

[56] F. Ramus, M. Nespor, J. Mehler, *Correlates of linguistic rhythm in the speech signal*, Cognition **73**(3), 265-292 (1999).

[57] P.A. Barbosa, *Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production*, Speech Prosody 2002, International Conference (2002).

[58] E. Grabe, E.L. Low, *Durational variability in speech and the rhythm class hypothesis*, Papers in laboratory phonology **7**, pp. 515-546 (2002).

[59] V. A. Dellwo, Fourcin, E. Abberton, *Rhythmical classification of languages based on voice parameters*. Proceedings of the XVIth ICPhS, Saarbrücken, pp. 1129-1132 (2007).

[60] P. Wagner, *The rhythm of language and speech: Constraining factors, models, metrics and applications*, Germany: Habilitationsschrift, University of Bonn (2008).

[61] V. Dellwo, *Rhythm and speech rate: A variation coefficient for* ΔC. Language and language-processing, pp. 231-241 (2006).

[62] B. Vaughan, *Prosodic synchrony in co-operative task-based dialogues: A measure of agreement and disagreement*, Proceedings of Interspeech 2011, Florence, Italy, August 2011, pp. 1865–1867 (2011).

[63] K. Klessa, D. Gibbon, *Annotation Pro+ TGA: automation of speech timing analysis. Annotation Pro + TGA: automation of speech timing analysis*, Proceedings of the 9th Language Resources and Evaluation Conference, Reykjavik, Iceland, pp. 1499-1505 (2014).

[64] D. Gibbon, *TGA: a web tool for Time Group Analysis. Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*, Aix-en-Provence, France, pp. 66-69 (2013).

[65] P. Wagner, *Visualizing levels of rhythmic organization*, [In:] XVIth International Congress of the Phonetic Sciences, Saarbrücken, 6-10 August 2007 (2007).

[66] C. Gussenhoven, *Phonology of Tone and Intonation*, Cambridge: CUP 2004.

[67] A. Chen, C. Gussenhoven, T. Rietveld, *Language-Specificity in the Perception of Paralinguistic Intonational Meaning*. Language and Speech 47(4), pp. 311–349 (2004).

[68] A. Chen, *Universal and Language-specic Perception of Paralinguistic Intonational Meaning*, PhD Thesis, Radboud Universiteit Nijmegen (2005).

[69] M. Karpinski, *The Boundaries of Language: Dealing with Paralinguistic Features*, Lingua Posnaniensis, vol. LIV (2)/2012. The Poznań Society for the Advancement of the Arts and Sciences, pp. 37-54. PL ISSN 0079-4740, ISBN 978-83-7654-252-2 (2012).

[70] L. Rabiner, *On the use of autocorrelation analysis for pitch detection*, IEEE transactions on acoustics, speech, and signal processing, **25**(1), pp. 24-33 (1977).

[71] L. Rabiner, M. Cheng, A. Rosenberg, C. McGonegal, *A comparative performance study of several pitch detection algorithms*. IEEE Transactions on Acoustics, Speech, and Signal Processing, **24**(5), pp. 399-418 (1976).

[72] W. Hess, *Pitch determination of speech signals: algorithms and devices*. Berlin – Heidelberg – New York: Springer-Verlag, 1983.

[73] P. Mertens, *The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model*. In Speech Prosody 2004, Nara, Japan, pp. 549-552 (2004).

[74] P. Mertens, *Polytonia: a system for the automatic transcription of tonal aspects in speech corpora*, Journal of Speech Sciences, **4**(2), pp. 17-57 (2014).

[75] A. Kendon, *Do gestures communicate? A review. Research on Language and Social Interaction* **27**(3), pp. 175-200 (1994).

[76] McD. Neill, *Hand and Mind. What gestures reveal about thought*, The University of Chicago Press (1995).

[77] A. Kendon, *Gesture. Visible Action as Utterance*, Cambridge: CUP 2005.

[78] R.M. Krauss, R.A. Dushay, Y. Chen, F. Rauscher, *The Communicative value of conversational hand gestures*, Journal of Experimental Social Psychology, **31**, pp. 533-552 (1995).

[79] R.M. Krauss, *The role of speech-related arm/hand gestures in word retrival*, [In:] R. Campbell, L. Messing, eds., Gesture, Speech and Sign, 93-116, Oxford University Press (2001).

[80] J. Bavelas, J. Gerwing, Ch. Sutton, D. Prevost, *Gesturing on the telephone: Independent effects of dialogue and visibility*, Journal of Memory and Language, **5**(2), pp. 495-520 (2007).

[81] Jarmołowicz-E. Nowikow, Karpiński, M. Communicative intentions behind pointing gestures in task-oriented dialogues. [in:] P. Z. Wagner, C. Malisz, Kirchhof (Eds.) Proceedings of GESPIN 2011: Gesture and Speech in Interaction Conference (2011).

[82] McD. Neill, *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press (1992).

[83] J. Gumperz, *Contextualization and Understanding*, [In:] A. Duranti, Ch. Goodwin, eds. *Rethinking Context: Language as an Interactive Phenomenon.Cambridge*: CUP, pp. 229-252 (1992).

[84] M. Silverstein, *Shifters, Linguistic Categories, and Cultural Description*, [In:] K. Basso, H. Selby, eds. *Meaning in Anthropology*. Albuquerque: University of New Mexico Press (1976).

[85] D. Figiel, *Lexical alignment in task-oriented dialogue*. Unpublished MA thesis supervised by M. Karpiński, Institute of Linguistics, AMU in Poznań (2017).

[86] F. Cummins, R.F. Port, *Rhythmic commonalities between hand gestures and speech*, [In:] *Proceedings of the eighteenth meeting of the Cognitive Science Society*, pp. 415-419 (1996).

[87] A.F. Hayes, *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*, Guilford Press (2013).

[88] K. Bergmann, H.P. Branigan, S. Kopp, *Exploring the alignment space – lexical and gestural alignment with real and virtual humans*, Frontiers in ICT (2015).

Maciej Karpiński – psycholinguist and phonetician at the Institute of Linguistics, Adam Mickiewicz University in Poznań. His scientific interests include linguistic and paralinguistic aspects of spoken dialogue, speech prosody, multimodal communication (including gestures and facial expression), and relations between speech and music. He has led research projects devoted to interpersonal communication and contributed to the development of numerous speech and multimodal corpora, including PoInt, Pol'n'Asia, Paralingua, DiaGest and Diagest2. With his team, he initiated and co-organised a series of international conferences confessed to gestures and speech in interaction (GESPIN), and established the first Polish laboratory focused on multimodal communication analysis. With collaborators from Warsaw University, he publishes a journal confessed to multimodal communication (International Journal for Multimodal Communication Studies). His current research is focused on the process of communicative accommodation as well as the perception of speech and speech prosody.

Katarzyna Klessa – (PhD habil.) phonetician, Assistant Professor at the Department of Phonetics, Institute of Linguistics, Adam Mickiewicz University in Poznań, Poland. Her research interests focus on the analysis and development of spoken language resources, especially with application to speech prosody, particularly timing and rhythm phenomena across speaking styles, also for disordered speech. She initiated the design and development of Annotation Pro, a freely available software tool (annotationpro.org) for annotation of linguistic and paralinguistic features in speech, as well as basic perception tests. Her recent work related to speech corpora includes the design and development of Paralingua corpus of conversational and affective speech in Polish or NeuroPerKog multi-lingual stimuli corpus of adult-directed and infant-directed speech for EEG and eye-tracting studies aimed at the investigation of phonemic hearing and speech development in babies. Another track of her research interests are language archives and dissemination of knowledge about endangered or under-resourced languages (e.g., inne-jezyki.amu.edu.pl, languagesindanger.eu).