THE JOURNAL OF SOCIAL POLICY STUDIES
ЖУРНАЛ ИССЛЕДОВАНИЙ СОЦИАЛЬНОЙ ПОЛИТИКИ

Jeff Higginbotham, Kayla Conway, Antara Satchidanand

RECORDING AND TRANSCRIBING INTERACTIONS OF INDIVIDUALS USING AUGMENTATIVE COMMUNICATION TECHNOLOGIES

The purpose of this article is to provide the reader with tools and recommendations for collecting data and making microanalytic transcriptions of interaction involving people using Augmentative Communication Technologies (ACTs). This is of interest for clinicians, as well as anyone else engaged in video-based microanalysis of technology mediated interaction in other contexts. The information presented here has particular relevance to young researchers developing their own methodologies, and experienced scientists interested in social interaction research in ACTs or as well as other digital communication technologies. Tools and methods for recording social interactions to support microanalysis by making unobtrusive recordings of naturally occurring or task-driven social interactions while minimizing recording-related distractions which could alter the authenticity of the social interaction are discussed. Recommendations for the needed functionality of video and audio recording equipment are made with tips for how to capture actions that are important to the research question as opposed to capturing 'generally usable' video. In addition, tips for processing video and managing video data are outlined, including how to develop optimally functional naming conventions for stored videos, how and where to store video data (i.e. use of external hard drives, compressing videos for storage) and syncing multiple videos, offering different views of a single interaction (i.e. syncing footage of the overall interaction with footage of the device display). Finally, tools and strategies for transcription are discussed including a brief description of the

Jeff Higginbotham– Ph.D., Communication and Assistive Device Laboratory, University at Buffalo, New York, USA. Email: cdsjeff@buffalo.edu

Kayla Conway– B.A., Department of Communicative Disorders and Sciences, University at Buffalo, New York, USA. Email: kconway5@buffalo.edu

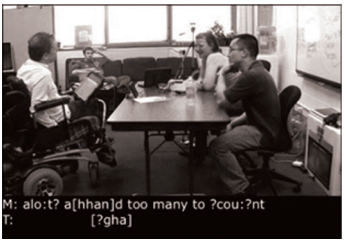
Antara Satchidanand– M.A., Communication and Assistive Device Laboratory, University at Buffalo, New York, USA. Email: satchida@buffalo.edu

role transcription plays in analysis, a suggested framework for how transcription might proceed through multiple passes, each focused on a different aspect of communication, transcription software options along with discussion of specific features that aide transcription. In addition, special issues that arise in transcribing interactions involving ACTs are addressed.

Keywords: microanalysis, conversation analysis, augmentative communication, alternative communication, observation methods, field methods

DOI: 10.17323/727-0634-2021-19-4-601-618

Introduction



4.	T:	((starts typing))
5.	M:	in this lab=er
6.	M:	er wirking with je:ff er
		(2.3)
7.	M:	help(ed) with (=tha) experiments here
8.	T:	^ ((sends massage))
9.	T:	╪ ı alot ı
10.	M:	alo:t? a[hhan] d too many to ↑cou:↓nt
11.	T:	[·gha]
12.	T:	hhe:n ·hyea::

Figure 1. Subtitled Video and Parallel Transcript

This figure is copyrighted by the authors (Higginbotham, Engelke 2013) and used with permission.

The above image and transcript (Figure 1) are a microanalytic representation of an interaction between an individual with Complex Communication Needs (CCN) due to amyotrophic lateral sclerosis (ALS) and their typically-abled partner. Individuals with CCN may rely on both high-tech communication aids, like the computerized speech-generating device (SGD) in the interaction above, and low tech options, such as letter and word boards. Collectively, these devices can be referred to as Augmentative Communication Technologies (ACTs). Social interactions involving those with CCN using ACTs are complex and oftentimes problematic, and the study of these interactions requires research methods that can address the nuances of these problems.

Social interactions involving users of ACTs depart from those involving only typically-abled individuals due to the unique physical and cognitive characteristics of the individual with CCN, the overt influence of the particular ACT on interaction, and the adaptive actions taken by the interactants to maintain social order during conversation. For example, ACTs require the aided speaker to take the time to compose and display their utterances by typing and/or selecting items on either a word/letter board or a computerized device. The resulting time gap, called *composition delay,* can be on the order of seconds to minutes and creates the need for conversants to deploy a variety of adaptive tactics in order to stay 'in time' with one another in the ongoing interaction, which supports the understandability of their conversational contributions (Bloch, Barnes

2020; Clarke, Wilkinson 2009; Higginbotham et al. 2007; Fulcher-Rood, Higginbotham 2019). Microanalysis is one methodology that can be used to examine issues of interaction, like timing and sequencing, in conversation.

As discussed by Bull (2002) microanalysis is a sort of social behaviour microscope, providing the researcher with the ability to carefully observe and analyse segments of social interaction. Used in many fields, microanalytic techniques have been applied by conversation analysts and microethnographers to uncover the basic structures of social interaction such as turn-taking and repair. Microanalysis involves repeated inspection of audio and video recordings of social interactions in real-time. When applied to rehabilitation research, the microanalytic approach allows researchers to account for how an ACT is used on a moment-to-moment basis in interaction. Results of these analyses can provide the foundation for new technology designs based upon real-world, contextualized uses of ACTs. This article provides the reader with tools and recommendations for collecting data and making microanalytic transcriptions of interaction involving people using ACTs, as well the broader area of video-based microanalysis. These recommendations are based on the first author's forty plus years of experience doing microanalytic work involving ACTs use.

Recording Social Interactions

A primary goal for video recording to support microanalysis is to make unobtrusive recordings of naturally occurring or task-driven social interactions, minimizing recording-related distractions which could alter the authenticity of the social interaction. Capturing social interactions including individuals using ACTs may mean recording in a wide variety of physical and social contexts that pose various technical challenges. We often record video in homes where rooms are darkened to support ACT eye-tracking technologies disrupted by bright lights. Recording in school classrooms, clinics, or offices, with a television or other activity in the background can create challenges to recording clear audio, especially of dysarthric or otherwise idiosyncratic speech. Appropriate audio and video recording equipment and strategies are needed to produce recordings suitable for microanalytic research purposes. The following is a set of recommended features for video recording devices and tactics used for capturing video and audio data of interactions including ACTs in naturalistic contexts.

Video Cameras and Microphones

Choosing the right video camera can be critical to one's success in making an analysable video. A video camera should be lightweight, easy to use, and responsive to a variety of distance, lighting, and sound conditions. When choosing a camera, we recommend purchasing one with the following features. (1) Can record across a wide range of lighting conditions, especially low light, without the use of an external lighting source. (2) Records directly to an SD or microSD

memory card using an MP4 or MOV video format. Do not depend on a camera's internal memory as you will inevitably find that your camera has run out of memory just as a particularly important interaction arises. (3) Can use high capacity memory cards Secure Digital High Capacity (SDHC) or Secure Digital eXtended Capacity (SDXC) of thirty-two gigabytes or more in order to record extended high-quality videos. (4) Can use high performance memory cards (UHS class speed) in order to record HD and 4k video formats. This ensures frame-accurate video recordings and enables rapid video transfer from the SD card to a computer. (5) Records in the MP4 and MOV video formats as they are used by most video transcription software. (6) Accepts a wide angle lens to capture the breadth of the environment and/or use the camera effectively in space constrained situations. (7) Includes an audio input port in order to attach an external wireless microphone. This can be useful in noisy situations or when you need to get the best quality sound to aid transcription of device output and impaired speech.

Two additional tips: We advise the researcher to bring a small digital tape recorder as a backup recorder, or for a low cost alternative to a wireless microphone, especially if participants are moving around a lot (e.g. going shopping), or to provide another audio recording source. We also advise that you use a tripod for more stationary work, but one which does not add too much weight to your equipment bag. We are currently using the lightweight tripods that can be easily set up and taken down and adjusted to accommodate a wide range of heights and angles.

Making a video recording[1]

The researcher's primary goal should be to capture actions that are important to the research question as opposed to capturing 'generally usable' video. This will likely mean using two or more cameras to capture the participants' actions from a couple of different angles. For example, if researchers are interested in how ACTs and participants' bodies are used during turn exchanges, cameras should be arranged to accurately assess gaze direction, gestures and activities involving content selection from the ACT. Accounting for the participants' physical context that cannot be captured by the video camera can be accomplished by taking a series of still shots of the room(s) in which the recording is being made. When recording in a single setting in which the participants do not move around very much, it is generally helpful for the main camera to be focused on the participants at a distance and angle that permits the researcher to view the participants' hands and upper torsos and heads in order to record most of the relevant body-based communication activities (i.e. see Figure 2). When the lighting is suboptimal (i.e. either too bright or too dark) the quality of the video can be improved by adjusting the camera's features (e.g. focus, zoom,

[1] This paper does not discuss the important tasks associated with planning and organizing field observations or the development of observation protocols. We refer the reader to Goodwin and Cekaite (2018) and Ochs et al. (2006) for a discussion of these topics.

contrast, white balance, video and audio scenes). We recommend learning about all the camera's features before moving on to make field recordings.

Figure 2. View of Video Setting (with ACT screen superimposed)

This figure is copyrighted by the authors (Higginbotham, Conway & Satchidadand, 2021) and used with permission.

If the researcher is interested in recording the ACT use itself, then it is advisable to use a separate camera for recording the ACT screen display. One tactic is to use a lower cost camera, but it may be better to have multiple cameras that are the same model, which ensures using the same video formats facilitating video processing and use with transcription software. Although we recommend using dedicated video cameras, high quality video can also be obtained with smartphones or tablets. Because of their size, smartphones and even smaller mini cameras may be less obtrusive when participants are moving about, particularly in public places. When possible, we would recommend mounting the smartphone or tablet on a tripod, using one of the many available cell phone holders. Again, separate microphones are recommended.

Finally, if multiple cameras are being used, the videos will need to be synchronized at some point, either during the processing of the videos or through the transcription software. In either case, it is important to record a discrete signal at the beginning of each new recording, using either a clicker or a light flash that can be picked up by each recording camera. This discrete sound or light will serve to align the videos at the same time point. More about the alignment process in the next section.

Processing Audio and Video for Analysis

Once the video recording is made, it often needs to be processed in order to save disk space, stream it over a server or internet, or to combine separate videos into a single one for viewing.

Naming/Identifying Video Files

When making videos with multiple cameras and/or for multiple participants in a study, it is essential to be able to correctly identify each video file. Developing appropriate file naming conventions is an essential part of maintaining an organized video database. In our lab, we have developed a fairly robust set of file naming conventions which can handle both naturalistic fieldwork as well as structured tasks. Figure 3 displays a typical filename used in our lab. The filename starts with the most general category, followed by successively more granular classifications. Each category label is separated by an

underscore (i.e., "_") that provides a visual separation making the filenames much easier to read. Note that these naming conventions should be used for all media related to the specific video series, including the transcript and sound files and additional notes. The categories we use include:

– Project Name – The name of the project
– Investigator (optional) – Truncated name of the current investigator working on the project.
– Participant – Name of the focal participant (e.g. Thad, NL, S 01). Numbered ID's should start with a letter (s = subject, d = dyad) followed by a 2 digit number.
– Task – Name of the primary research task (e.g. NAR, MAP) or activity or context (e.g. DIN(ner), HOME) in which the video was made.
– Condition (optional) – Experimental manipulation other than Task (e.g. output mode, visibility of the ACT display screen to partner).
– Trial refers to the sequential position of the video if the task is repeated.
– Video Part – Sometimes when the file size of the video reaches its maximum size, the camera will generate a new video. Video Part refers to the position of the file in the video series.
– Camera Focus – Identifies what the camera is capturing (e.g., participants, device) in a multi-camera setup.

Organizing Video Files on a Hard Drive or Server

From our experience, there are a few things to point out about storing videos. First, pay attention to the operational requirements of the transcription software you are planning to use. For example, when working with the ELAN program, we found that in order to maintain the links between the transcript and media files, it was essential to keep all the related files in the same folder. Depending on the complexity of the project, individual folders may be required for each task or recording session.

Video Conversion to Save Space

When recording high quality video, the researcher may notice that their video files tend to be large, sometimes exceeding two gigabytes in size. Large video files may produce noticeable delays when viewing them with transcription software, particularly if they are located on a server and streamed remotely to the researcher's computer. To overcome these constraints we recommend converting videos to the High Efficiency Video Coding Format (HEVC) also known as (H.265), which will reduce the size of the video file up to 50% or more. This will shorten delays in loading video files, and increase the ability of the software to start, stop and frameshift the video without noticeable delay during analysis. We use Movavi Video Converter Movavi for OSX. Handbrake, an open source multi-platform software application, is another video conversion option.

Synchronizing Videos for Transcription

There are basically two ways to display multiple videos when doing transcription. First, the transcription software may support the display of multiple videos (e.g., Elan, Transana). This requires the researcher to do the synchronization within the software itself. The videos will then appear side to side, with the secondary video (e.g., the ACT) displayed at the same size as the video of the interactants. The second way is to embed the secondary video within the primary video screen, commonly known as Picture-in-Picture (PiP). The PiP strategy has the advantage of not having to sync videos in the app, which can be difficult at times. It does significantly restrict the size of the secondary video. However, one can still view the original video independently if the need arises. In both cases, the videos need to be synchronized: either by the transcription software or using video editing software. Figure 2 provides an example of superimposing the device display using the PiP strategy. This article will not deal with synchronizing videos using Elan or Transana but will instead focus on creating a PiP video.

Appropriate software is necessary in order to combine two or more videos into a single video frame, you need software editing capable of doing so. In our lab, we use Camtasia to do most of our video editing. To produce a PiP video, it is essential that the two videos be synced in exact time with one another.

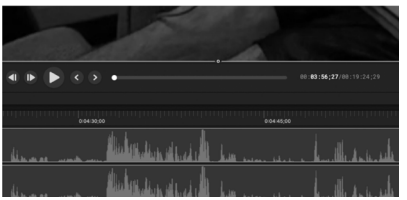


Figure 3. Synchronized Audio Tracks Using Camtasia Software

This figure is copyrighted by the authors (Higginbotham, Conway & Satchidadand, 2021) and used with permission.

The simplest way to do so is to identify a particular landmark (i.e., peaks and valleys) on the audio tracks of both videos, then to align them by moving one video file until the waveforms line up with the other. This can be accomplished by matching the audio waveforms in terms of their peaks and valleys. As mentioned above, using a 'clicker' designed for animal training is an inexpensive way to provide a sound spike and does not call your subject's attention to the fact that recording has begun. It is important that the videos are synced within a 1/10 of a second or less, so that the participant movements and sounds appear at the same time as they do on the video of the ACT. Figure 3 provides a screenshot of Camtasia with synchronized audio and video tracks.

Transcription Strategies and Tools

Role of Transcription in Analysis

In contrast to other forms of transcription which are used to categorize and count different linguistic phenomena, microanalytic transcription plays a central role in describing the talk in interaction phenomenon under study

through a textual – graphical analysis. Transcription in the microanalytic sense is interpretive, representing the orientations of the analyst to the data. The decisions around what phenomena to annotate, what should be in the foreground versus background, the details or granularity of the transcription, and the graphical-textual format of the transcription itself all create this orientation. Traditionally in conversation analysis, transcription has been thought to be a secondary source to the video or audio recording. Reflecting this importance placed on primary data, the recent focus on multimodal analysis has shifted the content of transcripts away from strictly spoken contributions to include embodied modalities.

Stages of Transcription

Higginbotham and Engelke (2013) discuss microanalysis as a multistage process, with each stage resulting in a finer grained, more focused representation of the phenomena being analysed.

Timecode	Primary Participants	Gloss	Notable Events
[00:17:04.04]	M-T-(F)	M asks T who the girl is who is with him, points	Pointing
[00:17:13.06]	T-M-(F)	T types	
[00:17:24.00]	M-T-(F)	M asks T string of questions, sometimes outpacing his answers	Multiple Qs to T
[00:18:22.25]	F-T-(M)	F asks T if they have seen each other before. T vocalizes then types	T Vocalizes
[00:18:46.27]	T-M-F	M talks to F while T composes response to F's question	
[00:18:51.19]	T-F-(M)	T respons. Starts second utterance. While typing second utterance M&F wait and stare at InTra display	Gaze at device
[00:19:23.16]	M-J-(F)-(T)	M asks J about the InTra display (why does it cut off)	Talk about InTra

Figure 4. Example of a Time Sequenced Interaction Catalogue

This figure is copyrighted by the authors (Higginbotham, Engelke 2013) and used with permission.

Cataloguing Data. If there are a number of videos, or the phenomena of interest occurs frequently in the video, one may want to go through your dataset taking observational 'snapshots' by systematically documenting your observations at a given time interval (e.g., every minute) or when you observe a particular phenomenon of interest, other-initiated repair request (see Figure 4). In our own work, we often task undergraduate students with reviewing videos and making systematic observations for each minute of video observed. They report on high-level phenomena: basic activity, the gist of the conversation, if the ACT user starts to compose an utterance, or if they spot any conversation troubles. We keep the trouble category broad, providing them with guidelines, such as to note when the communication partner produces an utterance with an upward intonation. This may also be a good time to report how interpretable the participants' actions are to the student transcriber, as interpretability is frequently a source of problematic communication that may require detailed analysis to unravel. These initial observations should be made using a simple transcription program with timestamps and notes that can be inserted so that specific stretches of the interac-

tion can be located later on. In fact, we sometimes use the initial observations as headers to contextualize a more detailed transcription.

Broad transcription. Broad transcription focuses on representing the words and some of the participants' physical actions in transcript form. It captures the participants' exact contributions in the correct sequential order. It also provides the opportunity for the researcher to repeatedly review the video material, bringing them closer to the phenomena being investigated. Figure 5 from Higginbotham and Engelke (2013) provides an example of broad and detailed transcriptions. We recommend that researchers complete separate passes of transcription to capture various types of information, first transcribing the participants' speech (spoken, SGD generated), one speaker at a time, then transcribing their physical actions.

Time	Line	Speaker	Transcription
1.0	01	M	*how many times uh*
	02	M	*how many times have you been here*
4.0	03	T	((starts typing))
4.5	04	M	*in this lab or working with Jeff*
	05	M	*to help with the experiments here*
	06	T	((pushes button to send message))
13.0	07	T	‡ |alot|
14.5	08	M	*alot((laughing))too many to count*
	09	M	((T looking at M)((vocalization))

Broad Transcription

Time	Line	Speaker	Transcription
01.3	01	M	*how many times eaah::*
02.8	02	M	*how many times have you been here*
			(.)
04.0	03	T	((starts typing))
04.6	04	M	*in this lab=er*
06.1	05	M	*er working with je:ff er*
			(2.3)
10.8	06	M	*help(ed) with(=tha) experiments here*
	07	T	^ ((sends message))
12.8	08	T	‡ |alot|
14.5	09	M	*alo:t? a[hhan]d too many to ↑cou:↓nt*
	10	T	*[·gha]*

Detailed Transcription

Figure 5. Example of Broad and Detailed Transcription
(Adapted from Higginbotham, Engelke 2013)

Detailed Transcription for Analysis and Presentation. At some point in the analytic process, the researcher will need to develop highly detailed transcriptions of at least a portion of the materials that are of analytic interest. At this level of transcription, the performative aspects of the interaction are transcribed. This includes the ways that utterances are produced (e.g., pronunciation, intonation, word prolongations, gaps and pauses within and between utterances), non-spoken behaviour (e.g. head movements, gaze, gestures) and the sequencing of talk and behaviour between interactants (Figure 5). The most widely used transcription conventions for capturing the performative aspects of talk were developed by Gail Jefferson (1983, 2004). Unlike prescriptive transcription systems, Jefferson's conventions have been adapted by many researchers throughout the years to deal with problems specific to their area of inquiry.

Figure 1 presents another important form of re-representation of the transcript for analytic purposes. Using InqScribe's captioning feature, the analyst can merge

the relevant portions of the text transcript into the video, preserving their attentional focus and facilitating the examination of the multimodal aspects of interaction through repeated replaying of the video (Higginbotham, Engelke 2013).

Transcription and ACT

Table 1

Transcription Notation

The transcription notation presented here utilizes the conversation analysis transcription conventions proposed by Gail Jefferson (2004). Additional notational forms are used to depict aspects of interaction not covered by existing conventions (e.g., AAC device sounds, text displayed on the screen), as well as symbols used to depict multiimodal aspects of social interaction.

Notation	*Definition*
(tha)	Text within single parenthesis indicates the analyst's best guess at sound production. (xxx) is used to depict unintelligible sounds.
oh:	A colon indicates a prolongation or extension of the sound or syllable it follows. More colons prolong the stretch.
↑cou:↓nt	Marked rise or fall in intonation is indicated by upward and downward pointing arrows immediately prior to the rise or fall.
,.!?	Punctuation indicates utterance level pitch inflexion, not grammar.
=	An equal sign marks where there is no gap or break within an utterance or between adjacent utterances.
((head shake))	Text in double parentheses indicate a gloss or description of nonverbal actions.
roller coaster [(xxx) the ship...[((nod))	Square brackets sure where to utterances or actions overlap one another.
(.) (2.3)	A period and parentheses indicates an interval of 1/10 of a second or less in the stream of talk. A number indicates the length, in seconds and tenths of a second of a pause and talk or duration of non-vocal activity.
°good°	Degree signs indicate a passage of talk which is quieter than the surrounding talk.
>mashed potatoes<	Talk presented in between > < symbols indicates that the speech rate was increased or rushed.
»I«	P1 and P2's Gaze directed toward one another
«I»	P1 and P2's Gaze directed away from one another
⊞|↑↓ →◎	P1's gaze directed toward SGD, P2's gaze directed upward and downward, then focued on another object, person or event, then unfocused gaze.

Men in Black	
|Men in Black|	Alternate ways of indicating SGD produced speech
MEN IN BLACK	
»	continue from past/into the future
^	Placement of letter, synthesized speech output or gesture relative to ongoing talk of other interactant
[ ]	Onset &/or offset of vocalization or synthesized speech relative to ongoing talk of other interactant
((head shake))	Onset &/or offset of head movements, gestures and other activity relative to ongoing talk of other interactant
nods.....	Continuation of gesture or other action
- - · ·	Movement: steady/transition

Augmentative communication poses several challenges to the default set of Jefferson's conventions, as many analysts need to account for the multimodal performance of utterances, the use of ACTs by both the aided speaker and their conversation partner, as well as the operation of the machine by these interactants. We have provided a glossary of transcript conventions that we use for our work and tailored for this article. It should be emphasized at this point that neither Jefferson's or our transcript conventions are 'set in stone.' On the contrary, they should be used, abandoned and/or adapted to serve the researcher's theories and analytic interests. Table 1 provides a set of transcription symbols used in our lab to describe interactions involving ACTs.

Extract 2

	01	C	*seven an a half* (1.0) [*near:ly*]
	02	J	[*hu*]
→	03	C	*how* (.) *now you ask me a question*
	04		(1.8)
→	05	J	(0.7) * (1.7) * (2.3) * (1.4) * "*how*" (5.0) * m - (1.1) * u - (1.4) * c - (0.4) * h (1.2) * "*how much*" (0.6) * (0.4) * (1.5) * (1.0) * * (1.8) *
	06	C	*how* [*much what*
	07	J	[*
	08	J	(0.5) * (1.6) * (1.5) * (1.2)
	09	J	[*
	10	C	[*ho:w m:uch w:hat*
→	11	J	(0.3) "*old*"
	12		(1.3)
	13	C	*how* [*much hold* ((sniffs))
	14	J	[*
→	15	J	* "*am I*" (0.6) * ["*how old am I*"
	16	C	[*how old*
	17		(1.2)
→	18	C	*eight*
	19		(6.6)
→	20	J	"*yes*"
	21	C	[*yea:::h h.*
	22		[((punches air))

Figure 6. Detailed Transcript: (Clark, Wilkinson)

Figure 6 shows a detailed text-based transcript from Clarke and Wilkinson (2009). In this excerpt illustrates an interaction between two young classmates – one of whom uses scanning access to operate his SGD. The authors use a number of conventions typical of the conversation analysis literature, including (1) vertical organization of turns, (2) brackets to show overlaps between participants, (3) gaps between actions in seconds and tenths of a second, (4) use of colons to display speech

prolongations, and (5) double parentheses for behavioural descriptions. The authors also used an asterisk to depict the device's audible feedback ('bleep') for each selection made. It should be noted that the italics used in the display of spoken speech are a transcript format requirement for the journal in which it was published, and not typically used for microanalysis manuscripts in other journals.

Higginbotham and Koroschetz (in prep) take a more deliberate approach to representing ACT device use within a multimodal interaction framework (Figure 7[1]). This excerpt illustrates the interactions between two women (R, S) during a prolonged message composition by S. In order to focus on the multimodal interactions surrounding the composition of S's message, 'I wonder how Rita and her family are,' S's head gestures (Sh), vocalizations (Sv), and ACT input (Sc) are vertically aligned with R's speech and organized through the use of a light grey bracket. Here, the authors represented S's vocalizations through inventive spelling and the use of the International Phonetic Alphabet and intonation symbols (up down arrow). The relative positioning of letter and word selections are aligned with R's speech using the carat (^) symbol. Brackets within each concurrent transcription segment serve to align S's vocalizations with R's utterances. To the left of the transcript, time is organized vertically to represent the relationship of S's composition efforts throughout the interaction. In order to keep the transcripts relatively easy to read and understand, gaze behaviour was represented to the side of each multimodal transcript segment to provide the reader a summary of the participant's concurrent gaze behaviour during the talk event. To the far left, time stamps representing the beginning of each transcript segment are used for the more traditional line numbers.

Higginbotham and Koroschetz's approach to multimodal representation maintains many features of traditional CA transcription while representing the coordination of S's vocalizations, gestures and ACT actions with R's speech. From this transcript one can observe how S was able to interact with R in real time using her gaze, voice and head gestures, all the while composing a topically different utterance with her device. The transcript also displays R's struggles understanding S's ACT-produced utterance at 13:03, probably related to the preceding two-minute, eight second composition delay occurring before S was able to issue her utterance.

In Figure 8, Savolainen et al. (2019) utilized the multimodal framework proposed by Mondada (2014) to transcribe communication book use during an interaction between a young boy with cerebral palsy and his speech-language therapist. Line seven shows the coordination of G's gaze with T's talk, with gaze shifts (marked with a +) on both T's talk line as well as J & T's gesture lines. On line eight, there is no talk by T. Instead, the top line becomes a behavioural timeline, with the onset and offset of each multimodal behaviour (e.g., ^,@,*,) marked in their order of occurrence on the top time-line and aligned with one

[1] See ps.hse.ru/article/view/13616/13347

```
7  T:  merkkiä nii. ei          + häiritse noi äänet.
       a symbol so. does not + disturb those voices.
   jG  follows                 + book                  → 9
8  J:  ⊢2.2⊣ ^ ⊢1.7⊣ @ ⊢1.3⊣ + ⊢1.3⊣ + ⊢3.2⊣ @* ⊢2.1⊣ ^
   jG  to book                 + his side + book       →
   tG                                          * book  → 11
   jP      ^ hand moves                                ^
   V                @ closing door, steps      @
9  J:  # ⊢1.5⊣ °+ ⊢0.2⊣ # ^ ⊢0.3⊣ ° ⊢0.1⊣ +§ ⊢1.0⊣ § ⊢0.6⊣ ^
   tA  comes back to sit                      §    § moves →
   jG  book    +                                + book  →
   jP  # points            # ^hand moves                 ^
   tS        ° MINÄ                  °
                I
```

Figure 8. Multimodal Transcription (from Savolainen et al. 2020)

the subordinate lines. In this transcript, the underline specifies the interval in which the particular behaviour is operative. This approach to multimodal representation has the advantage of depicting sequential relations on a single timeline, with accompanying content on the subordinate lines below.

Doak's (2019) analysis of student-teacher interaction provides an alternative to text-based transcription, by representing an interaction using a sequence of photographic images taken two to three seconds apart, representing the salient communication activities occurring between a teacher and her student (Figure 9). Here, spoken and signed communication are represented through 'speech balloons' accompanied by a description of the activities presented in the caption below each photographic frame. The refactoring of the transcript provides an important visualization of the participation frame, providing the reader with a richer, independent view of the communication context including the coordination of visual attention between the teacher and focus student, in which the non-spoken activity is no less privileged than the teacher's speech.

Figure 9. Example of a Graphical Hybrid Transcription (Doak 2019)

Higginbotham et al. (in prep) were interested in describing an instance of sequential misalignment taking place over a 0.2 second interval. In this extract (Figure 10), K an individual with paralysis due to ALS raises his eyes in response to J's question but, prior to raising his eyes, J shifts her gaze to his SGD, thereby missing his response and setting up a prolonged sequence of misunderstandings (see Engelke, Higginbotham 2013). Like Savolainen et al. (2019), Higginbotham et al. (in prep) adapt Mondada's multimodal transcription methods to accommodate this fine grained analysis of four 4 seconds of interaction. With an eye towards simplicity, the transcription symbols were selected to be readily interpretable (e.g. ↑ for gaze upward,–– for steady gaze, for gaze transition, initials for person being looked at). Unlike Savolainen et al. (2019), the transcript

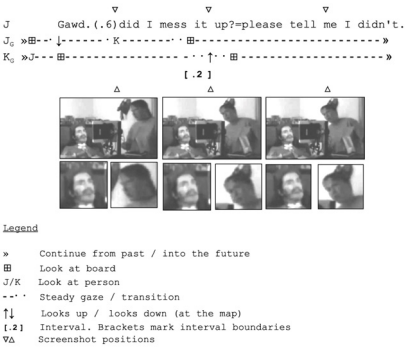
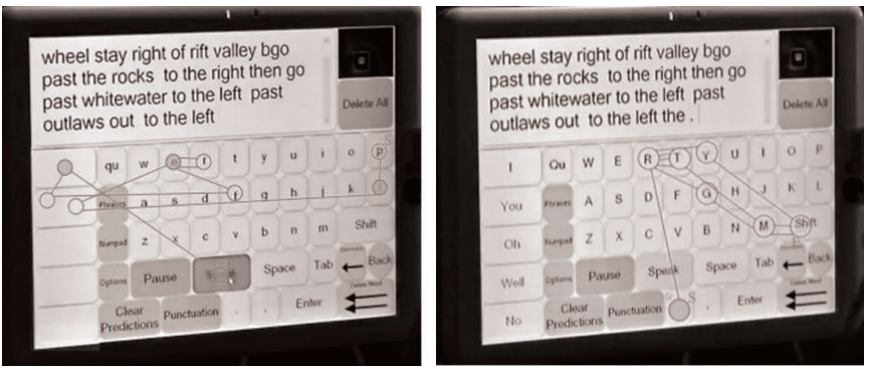


Figure 10: Example of Gaze Shifts in Multimodal Augmented Interaction

This figure is copyrighted by the authors (Higginbotham, Satchidanand, and Conway, in prep) and used with permission.

symbols were aligned, but not replicated in the first line of talk, which in this case simplified the visualization of the interaction.

Representing the ACT Device Use in Microanalytic Transcription. In most interaction studies focusing on ACT use, the speech and sounds produced by the device are the objects of transcription, whereas the interaction between the aided interactant and their device is often ignored or only indirectly reflected in general descriptions (e.g., looks at device) or in graphic representations of the interaction. However, the aided interactant spends much of their focus time using the device and its operation should be an important source of investigation, particularly if we are interested in making improvements in ACTs. In a current investigation, Higginbotham et al. (in preparation) examines the aided conversant's interaction with the device itself. The first example (Figure 11) traces cursor movements while the aided speaker struggles to construct a message using eye tracking. In this situation, the eye gaze system has become miscalibrated. The relatively large number of keys highlighted before a selection is made evidences the individual's difficulty navigating to the target key. By representing the aided speaker's extended attempt to make a letter selection, we can begin to account for the action of their partner, who begins looking at the device display screen.



Composed word "left": 10.6 seconds.
10 highlighted buttons, 4 selections, 1 word

Lost calibration: 9.4 seconds,
12 highlighted buttons, 1 selection, 0 words

Figure 11. Eye Tracking Traces on SGD Display

Transcription Tools

Video-based transcription software to analyze talk-in-interaction is in wide use nowadays. There are a variety of programs, each with their own specific set of features to facilitate different transcription orientations (e.g., interviews, conversation analysis). For microanalysis, transcription software is essential, but to appropriately support microanalysis, it needs to provide the researcher with the ability to locate, view and repeat viewings of specific portions of the video at different speeds, and with the ability to advance and rewind in very precise increments, 1/10th of a second or less is preferable.

Table 2 provides several descriptions of the different transcription software applications available, some of which are free of charge. Each of these packages differs with respect to providing a variety of features that address different microanalytic needs of the researcher. Depending upon one's research objectives, here are some basic features to look for.

Table 2

Transcription Software Features

	Level	*Operating System*	*Transcrip Orientation*	*Multiple videos*	*Waveform*	*Temporal Precision*[1]	*Caption Support*	*Keyboard Shortcuts*
oTranscribe	Simple	Web	Vertical	No	Yes	1 sec	No	Yes
InqScribe	Simple	OSX, Windows	Vertical	No	No	frame	Yes	Extensive
Annotation Transcriber	Simple	OSX	Vertical	No	Yes	frame	Yes	Yes
F4 Transcript	Simple	OSX, Windows	Vertical	No	Yes	1/10th sec	No	Extensive
Clan	Complex	OSX, Windows Linux	Vertical	No	Yes	frame	No	Extensive
ELAN	Complex	OSX, Windows Linux	Horizontal	Yes	Yes	frame	Yes	Yes
Transana	Complex	OSX, Windows	Vertical	Yes	Yes	frame	No	Yes

[1] All transcription programs use timestamps. Programs that use a frame-level temporal resolution (PAL=25 fps / NTSC=30 fps) can display timestamps in a millisecond format.

Controlling the Video. The ability to navigate through a video using keystroke combinations is an important time-saving and attention-focusing feature for the serious microanalyst. These video navigation functions include: (1) start and stop, (2) move forward and backward at different playback speeds, (3) jump forward and backwards at different time intervals, (4) shift back and forth in small time increments, and (5) replay selected portions of video with ease. Again, in our experience, a tenth of a second is the minimal resolution needed for accurate transcription of social interaction.

Figure 12. Example of a Simple Transcription Application (F4)

Timestamps. Each transcription application listed in Table 1 utilizes timestamps. When inserted into a transcript, a timestamp links that portion of the transcript with its associated video location. When clicked, the specific time on the timestamp will be cued up on the video, allowing the researcher to quickly navigate between different parts of a video and to easily repeat viewing of specific stretches of interaction. As shown in Figure 12, simpler software packages display the inserted timestamp in the transcript area itself. More complex video systems such as Elan and Transana selectively hide the timestamps, either by reducing them to a single icon, or in Elan's case, they appear in a timeline ruler above the transcription window. Some tools like f4 and Elan are able to calculate duration based on the interval between adjacent time stamps.

Acoustic Waveform. The visual representation of sound as an acoustic waveform allows the researcher to accurately locate the beginning and end of speaking behaviours which can also be used, coupled with the video image, to locate the beginning and end points of other behaviour. Similar to timestamping, clicking on the waveform should immediately reposition the video to the specific temporal location on the video. This allows for the rapid and precise positioning of video for review or documentation with a timestamp.

Video Display. The ability to change the size of the video image is important. With the exception of oTranscribe, all the transcription tools listed in Table 1 allow for resizing as well as viewing the video in a separate window. As mentioned earlier both Transana and Elan provide the ability to display multiple synchronized videos.

Transcription App as a Workspace

In addition to using it for transcription, we often use our transcription software as a workspace, to record a variety of measurements, make notes, test hypotheses, etc. This works well when we are able to call up multiple windows, using the same video. All the simpler systems support this approach by default. Both Elan and Transana provide support in the comments and notes windows, but not on the basic transcription screen.

Transcript Coding. All approaches provide for text-based tagging or coding, but only Clan, Transana and Elan provide analysis tools. A discussion

of coding with these tools goes beyond the scope of this paper but are included on their websites and in their user manuals.

Conclusion

Performing video-based fieldwork involving individuals who use ACTs may require a variety of techniques specific to capturing important aspects of their interactions both with other participants and with the ACT. This article provides a toolkit for researchers in multimodal social interaction analysis involving the interplay between individuals with complex communication needs, their partners and their ACTs.

New researchers can reduce much of the trial-and-error in setting up their studies by using the information provided in this paper. For more experienced researchers, this toolkit may provide a framework for extending prior research to include individuals with CCN who use ACTs or the techniques presented here might be used to explore joint-social interactions involving other digitally mediated technologies. There are many other excellent resources that provide more comprehensive approaches to the various aspects of this topic that we have raised here including Hepburn and Bolden's (2013, 2017) work on transcription and conversation analysis, Och et al.'s (2016) chapter on microethnography, and Higginbotham and Engelke's (2013) primer on microanalysis and ACTs.

The use of microanalysis in studying interactions involving ACTs is particularly pertinent to informing the design of new, more conversant communication technologies. SGD's current design inhibits their effectiveness for engaging in in-time, social interaction and disrupts use of embodied modalities of communication, especially gaze. Microanalysis serves as a powerful tool for revealing the interactional specifics associated with conversation involving ACTs including the problems that occur, like composition delay, and the adaptive strategies employed by interactants in their attempts to stay in-time during conversation.

References

Bloch S., Barnes S. (2020) Dysarthria and Other-initiated Repair in Everyday Conversation. *Clinical Linguistics & Phonetics*, 34 (10–11): 977–997.

Bull P. (2002) *Communication under the Microscope: The Theory and Practice of Microanalysis*. New York: Routledge.

Clarke M., Wilkinson R. (2009) The Collaborative Construction of Non-serious Episodes of Interaction by Non-speaking Children with Cerebral Palsy and Their Peers. *Clinical Linguistics & Phonetics*, 23 (8): 583–597.

Doak L. (2019) 'But I'd Rather Have Raisins!': Exploring a Hybridized Approach to Multimodal Interaction in the Case of a Minimally Verbal Child with Autism. *Qualitative Research*, 19 (1): 30–54.

Engelke C. R., Higginbotham D. J. (2013) Looking to Speak: On the Temporality of Misalignment in Interaction Involving an Augmented Communicator Using Eye-gaze Technology. *Journal of Interactional Research in Communication Disorders*, 4 (1): 95–122.

Fulcher-Rood K., Higginbotham J. (2019) Interacting with Persons Who Have ALS: Time, Media, Modality, and Collaboration via Speech Generating Devices. *Topics in Language Disorders*, 39 (4): 370–388.

Goodwin M. H., Cekaite A. (2018) Embodied Family Choreography: Practices of Control, Care, and Mundane Creativity. *Journal of the Society for Psychological Anthropology,* 47 (3): e1-e3.

Hepburn A., Bolden G. B. (2013) The Conversation Analytic Approach to Transcription. In: J. Sidnell, T. Stivers (eds.) *The Handbook of Conversation Analysis*. Oxford: John Wiley & Sons: 57–76.

Hepburn A., Bolden G. B. (2017) *Transcribing for Social Research*. London: SAGE.

Higginbotham D. J., Engelke C. R. (2013) A Primer for Doing Talk-in-interaction Research in Augmentative and Alternative Communication. *Augmentative and Alternative Communication*, 29 (1): 3–19.

Higginbotham J., Koroschetz J. (in prep) *The Impact of Composition Delay: Inserted Utterances & Sequential Misalignment*. Communicative Disorders and Sciences, University at Buffalo.

Higginbotham D. J., Shane H., Russell S., Caves K. (2007) Access to AAC: Present, Past, and Future. *Augmentative and alternative communication,* 23 (3): 243–257.

Jefferson G. (1983) Issues in the Transcription of Naturally-occurring Talk: Caricature Versus Capturing Pronunciational Particulars. *Tilburg Papers in Language and Literature*, 34 (1): 1–12.

Jefferson G. (2004) Glossary of Transcript Symbols with an Introduction. In: H. G. Lerner (eds.) *Conversation Analysis: Studies from the First Generation*. Amsterdam: John Benjamins: 13–31.

Mondada L. (2014). The Local Constitution of Multimodal Resources for Social Interaction. *Journal of Pragmatics*, (65): 137–156.

Ochs E., Graesch A. P., Mittman A., Bradbury T., Repetti R. (2006) Video Ethnography and Ethnoarchaeological Tracking. In: M. Pitt-Catsouphes, E. E. Kossek, S. Sweet (eds.) *The Work and Family Handbook: Multi-disciplinary Perspectives, Methods, and Approaches*. London: Routledge: 387–409.

Savolainen I., Klippi A., Launonen K. (2019) Coconstructing in Conversations Using a Communication Book. *Journal of Interactional Research in Communication Disorders,* 9 (2): 141–17.