AMI COMMUNICATION

AUGMENTED
MULTI–PARTY
INTERACTION
http://www.amiproject.org/

GUIDELINES FOR AMI SPEECH TRANSCRIPTIONS

Jo Moore [a], Melissa Kronenthal [b] and Simone Ashby [b]

AMI COMMUNICATION DX.Y

7 FEBRUARY 2005
VERSION 1.1

AUGMENTED
MULTI–PARTY
INTERACTION
http://www.amiproject.org/

[a] IDIAP, Switzerland
[b] University of Edinburgh

Project Ref. No.	IST FP6-506811
Project Acronym	AMI
Project Full Title	Augmented Multi-Party interaction
Security	Restricted
Date	10-Dec-04
Communication Number	Dx.y
Title	Guidelines for AMI Speech Transcriptions
WP Contributing to Communication	WP3
Task Lead Contributor	Jo Moore, IDIAP, Switzerland
Other Contributors	Steve Renals, UEDIN, UK Thomas Hain, USFD, UK Barbara Peskin, ICSI, USA Jan Cernocky, BRNO, David van Leeuwen, TNO, Netherlands
Author	Jo Moore, IDIAP, Switzerland, Melissa Kronenthal and Simone Ashby, University of Edinburgh.
Keywords	Speech. Transcript. Channeltrans. Meeting. Guideline.
Abstract	This document provides guidelines for the transcription of speech (recorded for the AMI Meeting Corpus) and it has several purposes: to detail the AMI speech transcripts, to provide guidelines for transcribers, to document the different options that were looked at by the project, and to stimulate discussion within the project. The AMI meeting corpus will include 100 hours of meeting data, the majority of which will belong to a series of four meetings, be forty-five minutes long, and have four participants. Speech is recorded using close-talking headsets, lapel microphones, and a microphone array. The speech transcripts will contain time-binned, word-level transcriptions for each speaker (channel), and additional information such as overlaps, interrupted words, restarts, backchannels, and nonverbal events. This guideline applies to speech recorded in AMI's smart meeting rooms, but it is also expected (perhaps with some mapping) to be used by other projects. To minimize the resources necessary to produce useable transcripts of ALL meetings, two levels of transcription are defined, namely low quality (LQ) and high quality (HQ). This document defines these levels in detail and outlines the major components of each.

1 Background to Project

The AMI project targets computer-enhanced multimodal interaction in the context of meetings. AMI is concerned with new multimodal technologies to support human interaction, in the context of smart meeting rooms and remote meeting assistants. The project aims to enhance the value of multimodal meeting recordings and to make human interaction more effective in real time. These goals are being achieved by developing new tools for computer-supported cooperative work and by designing new ways to search and browse meetings as part of an integrated multimodal group communication, captured from a wide range of devices.

In every project related to multimodal interaction, availability of large (annotated) databases is a critical issue. As a consequence, AMI (in collaboration with other EC initiatives, such as NITE) is focused on the collection, management, annotation and sharing of large multimodal meeting recordings (including dozens of gigabytes of audio, video, and XML per meeting) through media file servers (allowing downloading and uploading of information). The project aims to make recorded and annotated multimodal meeting data widely available for the European research community, thereby contributing to the research infrastructure in the field.

The AMI project commenced in January 2004, and will run for three years on the current funding. Within the first eighteen months, the AMI partners expect to collect a corpus of one hundred hours of meeting data. The data will be collected in three smart meeting rooms. The meeting rooms are at IDIAP in Switzerland, the University of Edinburgh in Scotland, and The TNO Human Factors Institute in the Netherlands. Each of these rooms is set up with at least four cameras, twenty-four microphones, special tools that capture participant's handwriting on paper and on the whiteboard, and a device that records which slides are presented using the beamer in the room. All of these data sources need to be annotated, in several different ways to create a complete corpus with a full set on annotations. In this document, we focus on the transcription of speech.

** Please note that it is recommended that you read the entire document

before starting to transcribe. **

a. Where to find documentation

The following sites provide some additional documentation for transcription systems.

Channeltrans

Using ICSI's Channeltrans

http://min.ecn.purdue.edu/~chenl/ALIGN_TOOL/channeltrans.htm

Extensions to Transcriber for Meeting Recorder Transcription

http://www.icsi.berkeley.edu/Speech/mr/channeltrans.html

Tips for using the Extended Version of Transcriber

http://www.icsi.berkeley.edu/Speech/mr/channeltransuse.html

Transcriber

The following link provides a User Manual for the Transcriber software. While some of the information found here will not be useful for the Channeltrans system, (because not all functions were included in the Channeltrans adaptation) it may be helpful with some of the basic functions and understanding the layout etc of Channeltrans.

http://www.etca.fr/CTA/gip/Projets/Transcriber/en/reference.html

b. Where to find details of project

You can find more information on the project at: www.amiproject.org. More information about the different meeting rooms involved in the project can be found at http://www.idiap.ch/~mccowan/Idiap_meeting_room.ppt (IDIAP, Switzerland), and (University of Edinburgh, Scotland) http://homepages.inf.ed.ac.uk/mlincol1/wp2/Edin_meeting_room.ppt. (anything at TNO?)

c. Media File Server

The AMI corpus will be stored and shared using a Multimodal Media File Server. The File server can be found at the following web address: http://mmm.idiap.ch.

d. Transcription Wiki

A Wiki has been created to keep transcription consistent across the sites and to circulate any changes. The Wiki can be modified by anyone with a username and password, and should be visited regularly by all transcribers. The Wiki can be found at http://wiki.idiap.ch/ami/AmiTranscribers.

e. Who to ask for help

A list of people to contact for help – for tools, speech, transcriptions, pay etc.

University of Edinburgh – Melissa Kronenthal, mkronent@inf.ed.ac.uk; Simone Ashby, sashby@inf.ed.ac.uk; Mike Lincoln, mlincol1@inf.ed.ac.uk

ICSI - Barbara Peskin, barbara@ICSI.Berkeley.EDU

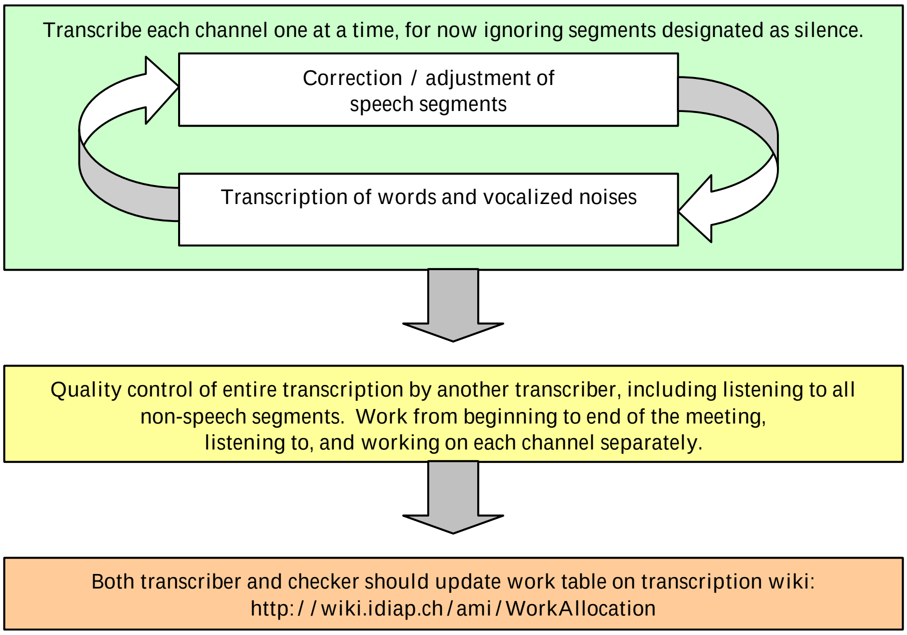
UT - Roeland Ordelman, ordelman@ewi.utwente.nl

BRNO - Jan Cernocky, cernocky@fit.vutbr.cz

2 Transcription Guidelines

The following flow chart outlines briefly the basic steps of the transcription process. Each of these steps is then explained in detail below.

Steps for Creating Transcriptions



Transcribe each channel one at a time, for now ignoring segments designated as silence.

Correction / adjustment of speech segments

Transcription of words and vocalized noises

Quality control of entire transcription by another transcriber, including listening to all non-speech segments. Work from beginning to end of the meeting, listening to, and working on each channel separately.

Both transcriber and checker should update work table on transcription wiki: http://wiki.idiap.ch/ami/WorkAllocation

a. How to Open a Transcription

1. On a Dice machine, open a command window and type ‘trans’. When prompted, specify ‘4’ channels.
2. When you are prompted to open a transcription, navigate to the folder which contains the meeting you have been allocated. The first time you start a transcription, select the file called ‘preTranscript.xml’. You will then be asked to select the audio file (signal) to use: in the same folder as before select the mixed headset channel. You will now see a waveform, one unsegmented and four pre-segmented channels below. **Note that the waveform you see does not change when you change what audio file you are listening to – it displays only the form of the first audio file you opened, which will normally be the mixed channel.*
3. Now add the remaining individual channel audio files: in the Multiwav menu, select ‘add a wav file’. Again in the same folder, choose ‘headset’ or ‘lapel’ 1, whichever you prefer or whichever is there. Go back to the Multiwav menu and repeat this process until all four headset/lapel files

have been loaded. You will come back to this menu any time you want to change the audio file you are listening to and select one of the five wav files there. Before continuing with the transcription, click ‘save as’, and give the transcription you are about to create a unique name. This should look something like: IS1008a.trs (I=Idiap, S=scenario meeting, 1008=meeting set number, a=first meeting of day). Replace the ‘I’ with ‘E’ for Edinburgh, which starts its numbering at 2000, and ‘T’ for TNO, which starts numbering from 3000. Save this file to your home directory. Now every time you want to work on this transcription, open this file first, then navigate to the folder where all the audio is stored in order to load the audio channels.

4. You are now ready to verify which transcription channel (the green bars) has been segmented for which audio channel. The first green bar, by default, is for the mix file and has not been autosegmented. You will not transcribe anything here. For the remaining four channels, determine which audio file it has been segmented for by clicking on segments of speech (the ones that are not labeled with ‘..’), and rotating through the audio files in the multiwav menu until you find the person that is actually speaking in that segment. This green bar will hold only the transcription for this wav file. Repeat with the other green bars, switching between audio files in the multiwav menu as needed until you have matched all the transcription channels with audio channels. It will very likely be that they correspond in their numbers, i.e. the first segmented channel will correspond to headset 1, the second to headset 2, etc., but don’t assume this is the case.

5. There will probably be a section at the beginning of the audio file which consists of no dialogue, or unintelligible dialogue as participants put on their microphones. There is however usually a point where the meeting is officially ‘opened’, and it is from here you should start to transcribe, ignoring everything that has come before it.

b. Segmentation

Each channel of speech (i.e. the audio on each personal microphone) must be broken into small segments, ideally separating speech from non-speech/silence regions and potentially breaking up long speech turns into smaller pieces. This task can be very time consuming, and for this reason we have used an automatic pre-segmentation system to do a first pass on the data. As you listen to the meeting, the pre-segmentation should help you to identify who is speaking, so that you can annotate that section of the meeting on that speaker’s channel.

The original speech signal has been automatically broken into separate segments (i.e. automatically segmented) for speech and non-speech (i.e. sections containing words and silence) for each speaker. For any given speaker, the segments that the computer believes contain silence will be marked by the annotation "..". For the purposes of the LQ segmentation step, these sections of the transcript, that have been marked as containing silence, can be ignored until the quality control step.

The main task here therefore is to check and adjust the boundaries of the segments that contain speech. The segment boundaries should be adjusted to ensure they accurately indicate when the speech starts, when it stops (ensuring not to cut any sounds off at the beginning or end of the segment) and to split large segments (when necessary). Segment boundaries generally correspond to speaker turns, although long stretches of speech uninterrupted by other speakers may need to be split further to obtain manageable pieces. In the AMI transcriptions, speaker segments are indicated by the separate line of transcription for each speaker. Automatically detected segments may need to be broken into smaller segments, or have their boundaries adjusted. (I actually don't think prosodics, beyond pause info, has too much to do with it. Silence/pause is really the main determiner, perhaps followed by syntax – e.g. sentence or phrase boundaries – provided there is a perceptual pause there.) These smaller segments are often called utterances. The speech can be broken into chunks based on pauses within the speech stream, preferably at syntactic boundaries corresponding to sentences or phrases provided there is a perceptual pause there. At first this may seem a little difficult as clear grammatical sentences are not often present in natural speech. However, once you become more experienced doing speech transcriptions the task of where to place segment boundaries becomes clearer and quite natural. The general rule is that if you have to think too long and hard about where to place a segment boundary, you probably don't need to put one in. Likewise, keep in mind that for our purposes segment boundaries are primarily for the benefit of the transcriber to make the task more manageable and do not have a high theoretical importance – so don't stress unduly about these!

As a general rule, no segment should be more than a minute long and most segments will be much shorter than that – generally only a few seconds. If you need to break long segments, try to do so at pauses in the speech stream *to ensure that the speech is not cut off*.

Things to remember about segmentation

1. Each segment should be padded by a small buffer of silence (1/4-1/2 second) on both sides (before and after the speech) *if possible*.
2. Breakpoints should be inserted at natural linguistic points in the utterance such as sentence or phrase boundaries to the extent possible. It is always preferable to break at such locations if the speaker pauses there, but for very long utterances it may be necessary to break wherever the most pronounced pauses occur, whether they are at natural linguistic boundaries or not.
3. An utterance may be up to 1 minute- in length, but most will be much shorter.

4. It is not necessary to insert a separate silence segment for an intra-utterance pause (even if it seems to be of considerable length); however if this has already been done by the automatic presegmentation it is fine to leave it as it is.

Note: the automatic segmentation may have marked several segments of speech/ non-speech at the beginning of the transcript before any dialogue has taken place. This is due to the interference from noises in the room at the beginning of the meeting such as chair scraping etc. These segments should be consolidated into one 'silence' segment before transcription starts.

c. Words and Vocal Noise Tags

In this section we explain the steps and guidelines for transcribing the speech within a given time segment. In general, we break noises that are made using the mouth into two categories, words and Vocal Noise tags.

Most of the speech you encounter can be transcribed into words with standard orthographic representations, such as you would fitrans t nd in a dictionary. But you will also encounter several other types of "verbal" events which will also need to be transcribed. These include words that may not appear in a standard dictionary but that are common in speech, such as reduced forms like "dunno" or "wanna", or acknowledgements like "uh-huh", or pause-fillers like "uh" or "um". In addition, when speech is broken off, there may be word fragments. Finally, there will be vocal sounds, like laughs and coughs and sighs, that may contain communicative value but do not have usual lexical representations. For this transcription effort, we consider all except this last category to be "words" and transcribe them as such, using a standardized set of spellings; the members of the last group are instead transcribed using special tags more fully described below.

It is important to remember that (where possible) the transcription should be an accurate record of what was actually said, and speakers should not be corrected to make their speech more grammatically correct. For example if the participant says "I dunno" this should be transcribed as it is heard, not as "I don't know". In an effort to assist the transcribers, the following list has been developed for common spoken phrases and communicative sounds. This list will be updated as the project grows, you should ask your transcription supervisor for updated versions throughout the project.

Table of common spoken phrases and communicative non-standard words

Please see http://wiki.idiap.ch/ami/RegularizedSpellings for an up-to-date list of these.

Please note: for these 'words' no special notation is necessary.

Vocal tags describe sounds that are made using the mouth (or nose) – but that do not have standard lexical representations. In these transcriptions, we have reduced the number of these which will be annotated to three, each of which has a simple symbolic representation. These will be:

$: laugh

% : cough

# : other prominent vocal noise (including creaky voice, aspiration, yawn, throat clearing, tongue click, etc)

The following text provides an example:

I mean I % really don't know what you are talking about.

In these transcriptions, it is not necessary to transcribe vocal noises in minute detail. The annotator should focus on transcribing mainly those sounds that add information to the annotation or that communicate meaning. For example there are occasions where a cough is used to interrupt another speaker or to indicate disagreement, and these should be annotated.

**It will probably be useful to make a list of these symbols and keep them handy when you are transcribing, perhaps on a sticky note glued to your monitor.*

Important information about transcribing words and vocal noises

1. Transcribe verbatim, without correcting grammatical errors, e.g. "I seen him", "me and him have done this".
2. Standard spoken language should be transcribed as it is spoken, e.g. "gonna" not "going to", "wanna" not "want to", "kinda" not "kind of", etc. See the regularized spelling page on the wiki for more examples.
3. Avoid word abbreviations, i.e. "doctor" not "Dr", and "mountain" not "Mt".
4. Use normal capitalization on proper nouns and at the beginning of sentences.
5. Remember to watch for common spelling confusions like "its" and "it's", "they're" and "there" and "their", "by" and "bye", "of" and "off", "to" and "too", etc.
6. Mispronunciations: if a speaker mispronounces a word and you know what word was intended, transcribe the word as it should be spelled and mark it with an asterisk after the last letter, e.g. spaghetti*. If you do not know what word was intended, transcribe what you hear and mark it with parentheses, e.g. (fligop).
7. Spell out number sequences, e.g. "forty four" not "44".
8. Acronyms should be spelt as they are pronounced, e.g. "NASA" or "U_S_A_".

9. When a letter sequence is used as part of a word, add the inflection after the underscore: “They I_D_ed him”.
10. If a speaker does not finish a word, and you think you know what the word was, you can spell out as much of the word as was pronounced inserting a single dash as the last letter of the word, e.g. “ I think basic- ”.
11. Punctuation should be limited in the corpus. You should only use commas, full stops (periods), and question marks to punctuate a ‘sentence’. You can include dashes in compound words that are traditionally written with them, e.g. ‘passer-by’, but make sure there is no space between the dash and the words on either side of it.
12. If the speaker is cut off or doesn’t finish their sentence, a dash should be inserted at the end of the statement, e.g. “I was going to do that, but then –” (remember, if there is no space between the last word and the dash, this indicates that the word was not finished). **If the utterance ends in a word fragment, it is not necessary to add a separate dash to indicate that the sentence was also left unfinished.*
13. While punctuation should generally follow (simplified) standard English usage, particular care should be taken at two locations: (a) at regions of disfluency, where the speaker interrupts himself to correct or restart or repeat, use a dash (if there is not already a dash from a word fragment), e.g. “I just meant – I mean …”, and (b) at the end of a speech segment, punctuation should be used to make it clear whether the next turn is a continuation of the current one. If the speaker continues with the same utterance, punctuate as you would if there was no break (including potentially having no punctuation at all at the end of the initial turn); if the speaker breaks off and does not continue the sentence in his next turn, then indicate this with a dash.
14. As regards spelling, please note that we are using British standard spelling throughout the transcriptions, i.e. colour vs. color, realise vs. realize.

The following section outlines the practical steps for annotating the speech in transcriptions.

Practical Steps

You should have just adjusted the time boundaries for a given segment. By the time that you have done this, it is likely that you have understood a lot of what is said in that segment. For this reason, it is a good idea to transcribe this segment before moving on to the next one. The following steps will help you to accurately transcribe the speech.

Remember we are interested in a balance between quality, consistency and accuracy. If after listening to a ‘word’ several times you are still not sure what it is, (this may occur *more than you think, particularly because this corpus includes the speech of many people with different accents and native languages) you should mark the word as "(??)" and move on to the next word.*

1 Did you understand what was said?

Yes Go to Question 2

No Listen again, if you think you might know what the word is but are very uncertain, type that word inside parentheses, e.g. (egg). If after the second listen you still don't know what a word is, mark the transcription with a '(??)', i.e. using question marks inside the parentheses, where that word should go. Move on to the next word and start again at Question 1. If a whole sequence of words is unclear, use one set of parentheses for the whole interval. Parentheses should not be used for any purpose other than to indicate uncertainty in these transcriptions. You should in fact mark *anything* you are uncertain about with parentheses, so that your supervisor or transcript checker will easily be able to identify any problem areas.

2 Does the word that was spoken appear in the dictionary?

Yes Type the word as it appears in the dictionary. If the word is spelt differently in different countries you should use your native spelling. *Remember numbers should be spelled out rather than using numerals, e.g. "sixty six" not "66".*

Move on to the next word and start again at Question 1

No Go to Question 3.

3 Did the participant use an acronym or a (group of) letter(s)?

Yes There are two possibilities:

1. If they used an acronym as it is commonly pronounced, you should type the acronym as it normally appears e.g. "We have a meeting room in IDIAP" (probounced eediyap).
2. 2. If the person spells the acronym out as they speak, you should separate the letters with an underscore e.g. "We have a meeting room in T_N_O_". We use a similar notation if a person uses letters to identify a list e.g. " we have two things to do A_ we need to fix the segmentation and B_ we need to type the words".

Move on to the next word and start again at Question 1

No Go to Question 4

4 Did the person communicate real meaning when they talked? For example were they answering a question, trying to interrupt another speaker etc.

Yes Go to Question 5
No Go to Question 7

5 Is the word or phrase common and accepted in spoken English?
Yes Use the AMI regularized spellings list on the transcription wiki: http://wiki.idiap.ch/ami/RegularizedSpellings. Keep in mind that we want to have as few variations as possible so check the list for the correct spelling.
Remember no special tag is required for common spoken language.
Move on to the next word and start again at Question 1
No Go to Question 6

6 Is the sound a word or phrase in a foreign language?
Yes If you can transcribe the sounds, do so, marking them on both ends with the carat sign '^'. If you do not want to hazard a guess, and you are certain it is non-English, mark the sounds as follows: ^(??)^.
Move on to the next word and start again at Question 1
No Go to question 7.

7 Does it seem to be a word fragment or unfinished word?
Yes Transcribe as many of the phonemes as you hear, marking that the word has been cut off with a '-' as the last letter. If you are uncertain about the phonemes you hear, mark the whole thing with parentheses.
No Go to Question 8
8 Can one of the vocal tags be used to describe this sound?
Yes Insert the appropriate symbols.
Move on to the next word and start again at Question 1
No If you still think it is important to note the sound, you may need to create a new vocal noise tag. Otherwise, you could just mark it with question marks in parentheses for review later.
Move on to the next word and start again at Question 1

3. Quality Control

In this stage a second transcriber, or the transcription supervisor, has a chance to listen to each audio file (i.e. for each speaker), working from start to finish, correcting and improving the quality. It is important in this step to start from the very beginning of the meeting and to listen to all segments (including those that have not previously been listened to because they were marked as containing silence) on each channel to ensure nothing has been missed. Segment boundaries should be adjusted, if necessary. The checker should as well use this time to double check the transcriptions for errors, omissions and misspellings. This is also the time to review anything that was marked in parentheses as uncertain by the original transcriber. The file should be saved as a different version that indicates that it has been checked. Once this step is complete the checker should update the work allocation page on the wiki.

4. Checking the transcription into CVS

More details to come.

5. Updating the Wiki

When you have completed a first pass transcription or a check, update the work allocation page on the transcription wiki to include this, including the number of hours it took you to complete the job. This page can be found at http://wiki.idiap.ch/ami/WorkAllocation. Click the ‘Edit Text’ link near the bottom of the page in order to modify the tables.

Appendix

Channeltrans – User hints

The software provides drop-down menus for all of its features. However, some features (such as speaker information and channel IDs) were removed when Transcriber was adapted to Channeltrans. See the Channeltrans documentation for more details.

The following table provides details of the steps that are commonly used when creating a speech transcription.

Task	Notes
Start Channeltrans	For windows - you can either double click on the Icon or go through the start menu – programs – Transcriber – main.tcl. In Unix type 'trans' into a command line. When you first open Channeltrans it will prompt you to indicate the number of speakers in the meeting. For the AMI meeting corpus, this number will generally be 4.
Opening a file Ctrl+O	The system will prompt you to choose a file to open. You should open either the preTranscript to start a new transcription, or a transcription file you have previously worked on (see section 'a' above). The wav file that appears will be the wav file for the first audio file you open.
Segmentation	You will note that the transcription has already been divided into speech / non-speech segments for each channel. The non-speech segments are indicated by ".." . The sections of the audio that the computer thinks contain speech, will be left blank – ready to be transcribed.
Loading individual audio channels	If you would like to load the audio files for each of the individual channels you can then do this using the MULITWAV menu – ADD A WAV FILE. It is recommended that you add the individual files in order i.e. speaker 1, then speaker 2, then speaker 3 and finally speaker 4.
Save – Ctrl+S	You can save the transcript using the menu – FILE – SAVE.
Save As	After each pass of the transcription (see Sections 7&8 for more details) you should save a new version of the transcription. You should label each version Meeting#_v1, Meeting#_v2 etc.
Select a section of the transcription Leftclick on green bar	If you left click your mouse on one of the sections in the green transcription channels, that section of the audio channel will be highlighted. If you play the audio now it will only play the audio for that section.
To select several segments together Shift+leftclick on green bar	If you hold down SHIFT and LEFT CLICK on several consecutive sections for a speaker, all of these sections will be highlighted.

Task	Notes
Change a segment boundary Ctrl+leftclick+drag boundary	If you find that the automatic segmentation boundary needs adjusting (perhaps it cuts off the end of the last word that the person says) then you should hold down CONTROL and LEFTCLICK and DRAG the boundary to its new position. (Perhaps this is how to do it in Windows…? I certainly don't need to hold down the control key in Unix, just click on a boundary and drag.)
To add a section boundary Enter	If the automatic segmentation has missed a segment (for example a section of silence) you should first CLICK on the green channel for that speaker, then position the cursor at the point in the audio where the section break should appear, and then hit ENTER to add a break.
To delete a section boundary Shift+Backspace	If the automatic segmentation has incorrectly split what should be a single segment into two segments, you should remove the border between them. LEFTCLICK in the second segment (on the green bar) and hit SHIFT and BACKSPACE.
Play/ Pause Tab	The TAB key will toggle between play and pause. However, if you have a single segment selected, only that segment will play when you hit TAB. If you then hit TAB again (after it has played that segment), it will play that segment again. (It may be worth noting that you can modify playback options using the appropriate pull-down menu. It is sometimes helpful to switch to continuous playback or pause-and-continue mode, rather than stopping at segment boundaries.)
Quit/ Exit Ctrl+q	After you have saved your transcription you can quit Channeltrans using CONTROL+Q

Annotators may choose to print this summary table and keep it nearby (perhaps attaching it to their computer screen) while they become more familiar with the software.

TASK	SHORTCUT
Open file	Ctrl+o
Save	Ctrl+S
Select section	Leftclick on green bar
Select several segments	Shift+leftclick on green bars
Change segment boundary	Ctrl+leftclick+drag
Add section boundary	Enter
Delete a section boundary	Shift+Backspace
Play/ Pause	Tab
Quit/ Exit	Ctrl-q