quarterly report for

Improved Monosyllabic Word Modeling on SWITCHBOARD

submitted by:

J. Hamaker, N. Deshmukh, A. Ganapathiraju, and J. Picone
Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University
Box 9571
413 Simrall, Hardy Road
Mississippi State, Mississippi 39762
Tel: 601-325-3149
Fax: 601-325-3149
email: {hamaker, picone}@isip.msstate.edu

EXECUTIVE SUMMARY

SWITCHBOARD (SWB) Corpus consists of 2430 conversations digitally recorded over long distance telephone lines. The SWB Corpus totals over 240 conversation hours (elapsed time) of data. The average conversation duration is six minutes. The transcriptions contain more than 3 million words of text. The SWB Corpus includes more than 500 adult-aged speakers and covers most major American English dialects. Such impressive statistics make SWB the premier database for telephone bandwidth large vocabulary conversational speech recognition (LVCSR) research. The goal of this project is to resegment the speech data and correct the transcriptions in an effort to significantly advance LVCSR technology.

The Institute for Signal and Information Processing (ISIP) has previously released 1000 SWB conversations with revised segmentations and transcriptions. We also demonstrated that one could obtain a 2% decrease in WER by simply reestimating LVCSR models on the corrected segmentations. This work was presented at the recent Hub-5 Conversational Speech Recognition (LVCSR) Workshop and was met with much support as well as many suggestions for improvement from the speech research community. In the months following the workshop, we have made numerous revisions to our transcription guidelines and procedures to address the issues raised. These improvements include:

- giving strong preference to creating utterances with clear linguistic meaning such as phrases or complete thoughts even when that requires a smaller silence buffer on the utterance;
- stressing the importance of placing boundaries in a region of low energy. Specifically we do not want to place boundaries in large bursts of noise as that corrupts delta features for acoustic modeling;
- training our validators to place more importance on generating clean utterances for acoustic model training and language model training than on following the segmentation rules “to the letter”.

With the changes in transcription conventions has come a profound change in our quality control methods. We have implemented a incremental and multiple pass quality control procedure which provides almost immediate feedback to the validators. This has worked to decrease error rates and increase productivity. We had earlier reported cross-validation of close to 3% WER for a relatively clean utterance. We have tested a new validator on that same conversation using our new procedures and have found that their error rate is less than 1%. This is a substantial improvement over the current LDC transcriptions which have an 8% WER measured under the same conditions.

To this point, we have released 150 of the revised transcriptions and plan to release 100 per week until the beginning of the calendar year. By January 1, we will release the 1000 conversations with what we believe are our final transcriptions conventions in place. We expect that these transcriptions will have an average WER of close to 1%. These 1000 conversations comprise 60% of the conversations used in the WS’97 partition, and 45% of the entire SWB corpus. By April of next year we will be back on track for our December 1999 deadline of delivering the entire set of corrected transcriptions and segmentations with automatic and manual word alignments. All information relevant to the SWB work is located at *http://www.isip.msstate.edu/resources/technology/projects/current/switchboard/*.

TABLE OF CONTENTS

1. ABSTRACT

The SWITCHBOARD resegmentation project (SWB) has gained increased visibility in the speech research community since our last project report. Presentation of our work at the Hub-5 Conversational Speech Recognition (LVCSR) Workshop yielded numerous suggestions from our colleagues. In response to these suggestions and from an analysis of our error performance, we have restructured our transcription and segmentation process to provide a more thorough multi-level review of the data before it is termed “finished”.

As of our last report, we had released 1000 conversations with revised segmentations and transcriptions. Retooling our methods and retrofitting those 1000 conversations has been the focus of our time since October. We believe that we are now producing transcriptions and segmentations which are as accurate as humanly possible and that we will require no future major changes to the transcription conventions. Our new streamlined process has also enabled us to produce data more efficiently and to train validators more quickly. It is our plan that by the end of the year we will have completely updated the 1000 conversations and will again be releasing new data. We estimate that it will take us until March 1999 to get back on track for our goal of completing the project by December 1999.

2. INTRODUCTION

The SWITCHBOARD Corpus [1] [2] has become critical to the success of state-of-the-art LVCSR systems. Using this data, however, has not been without its share of drawbacks. SWB was a great example of the trials and tribulations of database work, in that the quality of the data suffered from a lack of understanding of the problem. Word-level transcription of SWB is difficult, and conventions associated with such transcriptions are highly controversial and often application dependent. By 1998, the quality of the SWB transcriptions for LVCSR was recognized to be less than ideal, and many years of small projects attempting to correct the transcriptions had taken their toll. Numerous versions of the SWB Corpus were floating around; few of these improved transcriptions were folded back into the LDC release; and many sites had spent a lot of research time cleaning up a portion of the data in isolation. In February of 1998, ISIP began a project to do a final cleanup of the SWB Corpus, and to organize and integrate all existing resources related to the data into this final release.

In the first six months of this project, we made significant progress in transcribing and resegmenting the corpus by releasing 1000 of the 2430 SWB conversations. We also amassed a large collection of tools and resources for use with the SWB project. Most notable of these are the development of our public-domain segmentation tool [3], the SWB frequently asked questions (FAQ) web-site [4], the SWB educational resources web-site [5], and a comprehensive collection of statistics [6] related to SWB. We continue to maintain a mailing list (*swb@isip.msstate.edu*) which is our point of contact to the research community for resolving subtle transcription issues and communicating progress on our efforts.

From the start of this work, we have solicited feedback from the speech research community at large to insure that the data we are generating will be well-suited for state-of-the-art research and that we are conforming to accepted standards in the community. Through discussions with many participants at the recent Hub-5 LVCSR Workshop we became concerned that there were some

issues that we were not placing enough importance on. The most important of these are 1) the marking of boundaries in noise which can cause severe problems for training acoustic models and 2) maintaining phrase boundaries in the utterances even at the cost of smaller silence buffers for each utterance. We have addressed these issues in the last three months by developing a more extensive set of transcription and segmentation guidelines and using these to do a second pass over the previously released data.

3. CHANGES IN TRANSCRIPTION CONVENTIONS

Our participation in the Hub-5 LVCSR Workshop brought numerous comments from our colleagues in research which have changed our approach to validation and have created specific transcription rules. We have also carried out an intensive review of the previously released data which uncovered problems that had not been expected such as boundaries being placed in noise. A modified transcription guidelines document has been built to disambiguate the problems faced by the validators, thus producing a more consistent set of transcriptions and segmentations. The most current version of our transcription guidelines document is included as Appendix A.

3.1. Marking boundaries near noise or echo

One of the driving points for reformulating our transcription and segmentation conventions was to avoid the problems our validators were having with marking boundaries in the presence of noise or echo. It was our intention to have “clean” utterances where each boundary is in a point of silence, each utterance is buffered by silence, and each utterance contains a meaningful phrase. However, our validators had the false impression that the placement of the boundary was unimportant as long as a 0.5 second silence buffer on either side was maintained. This caused them to place boundaries in large bursts of noise or echo when there was an acceptable low-energy region in close proximity. Boundaries in this location cause corruption of delta features and are contrary to our desires because the utterance starts with noise instead of silence.

Thorough examination of the data also showed that, in their confusion over silence buffers, the validators were often not choosing the best place for a boundary and were corrupting the phrase structure of the conversation. To avoid each of these problems, we created the more detailed segmentation rules shown in Figure 1. We also put each validator through a detailed training session designed to tease out the subtle points of confusion which were disrupting their work. We are now confident that each validator is producing much more accurate segmentations.

3.2. Consistency in capitalization

Before the LVCSR workshop, our convention for capitalization was to capitalize words as they would appear in written text excluding capitalization of words which begin a sentence. This convention included capitalization of the pronoun “I”. A few participants of the workshop expressed concern about a language model’s ability to distinguish the capitalized pronoun “I” from a capital “I” indicating an abbreviated proper name or a capital “I” in a title. To address this concern, we have changed our procedure to use a lowercase “i” when the word is used as a proper noun and a capitalized “I” in other cases.

1. Each utterance should be padded by a nominal 0.5 second buffer of silence on both sides. In general, these silence buffers can range from 0.35 to 0.75 seconds.
2. The boundary can **only** be placed in a "silence" consisting solely of channel noise and background noise. Whenever possible place the boundary in a section with very low energy (visually speaking, this is a flat part of the signal in the segmentation tool)
3. The 0.5 second buffers can contain breath noises, lip smacks, channel pops, and any other non-speech phenomena. However the boundary **can not** be placed in a noise of this sort.
4. No utterance can be longer than 15 seconds. As an utterance approaches 15 seconds in length, the validator is allowed to find a point of segmentation that will generate silence buffers less than 0.5 seconds but not less than 0.1 seconds. If this segmentation can not be found then that utterance should be marked as "NEEDS_REVIEW" in the log file and the validator should send an e-mail to the adjudication team explaining the problem.
5. Every utterance containing only silence must be greater than 1.0 seconds in duration.
6. Whenever possible choose a segmentation that maintains the phrase structure of the conversation. This means that, ideally, we would like every utterance to contain a single phrase. However, due to the nature of the SWB data, we realize that this is not always possible. **Note**: The previous instructions take precedence over this one.
7. The end of the preceding utterance coincides with the start of the next utterance. Hence all data is accounted for. Segmentation essentially involves placing a boundary between two utterances.
8. Consider a stretch of silence which has small amplitude noises embedded in it as a silence only utterance - do not mark the noise and do not segment the noises into separate utterances. However, if a noise has a particularly high amplitude, then segment it into its own utterance.

Figure 1. Detailed segmentation rules that explicitly cover cases of boundaries in noise and echo. These rules also stress the need for segments that follow the phrase structure of the conversation.

3.3. Marking asides

A situation that occurs relatively infrequently in SWB is when one of the two speakers in the conversation talks to a person in the background. In the past, this may have been transcribed as [noise], as part of the normal transcription, or, worse, not transcribed at all. This could have dire consequences for training or testing a system since the acoustics for these "asides" would be on par with the conversational acoustics. Also, these asides will often carry over into the conversation between the two primary speakers. For this reason, our initial inclination was to simply transcribe the words which were intelligible as part of an utterance. However, on the advice of Dr. William Fisher at NIST, we adopted their practice of transcribing the parts of the conversation spoken as asides between the markups "<b_aside>" and "<e_aside>". An example is shown in Figure 2.

2264A-0040 and i'm kind of like you i wish they would do that because so at least i knew somebody you know was getting the money out of it you know that i[t]- was gonna use it for good so
2264B-0035 **<b_aside>** *what's the matter sweetie you need to wash your hands maybe Paw-Paw can help you sure* **<e_aside>** sorry
2264A-0043 [noise] sounds like you have a little one just like i [laughter-do]
2264B-0037 she's uh she'll be two in July

Figure 2. Example of an exchange where speaker B talks to a child in the background. The details of the aside are separated from the primary conversation by the <b_aside> and <e_aside> markers. In this case, the aside becomes the topic of the primary conversation. For a speech understanding system, having the transcription of the aside may be extremely important for understanding the remainder of the conversation.

4. AN EFFICIENT WORKFLOW WITH IMPROVED QUALITY CONTROL

A good portion of the past three months has been spent in tightening the quality control on the data produced by our validators. As we began to examine the "completed" data very closely we found that our quality control measures were allowing certain problems to slip through — chief among these was the placing of boundaries in bursts of noise. We have combatted these difficulties by implementing a more strenuous, multi-layered quality control system, a detailed battery of quality control scripts and continued monitoring of relative performance through cross-validation tests. All of these have resulted in a much cleaner set of utterances and a more efficient workflow.

4.1. SWB data control flow

Formerly, one validator would both segment and transcribe the data, and, when a large portion of data was ready for release, the project manager would run a small number of quality control scripts to verify the validators work. This approach was flawed in three respects. First, the data was only reviewed in large chunks (on the order of 100 conversations) which meant that the same type of error may have been propagated through a large number of conversations before being corrected. Second, subtle problems with the data were not being found because there was limited oversight from the more experienced members of our group. Lastly, the validators had difficulty focusing on both the segmentation and transcription because the number of issues involved in each is substantial. We have addressed each of these issues by putting the workflow demonstrated in Figure 3 into place.

The first benefit this new process provides is an increased review of the data. Each conversation is now completely reviewed by two different validators. One validator only resegments the data — building a set of utterances which match the specifications of our transcription guidelines. The other validator concentrates on making transcription corrections and makes note of any segmentations that are questionable so they can later be reviewed by the project manager. This segmental approach has worked well because the validators are able to focus on a single task rather than balancing both segmentation and transcription. We are also able to bring new validators up to speed more quickly by having them focus on the relatively simple task of segmentation while our more experienced validators work with transcriptions. In the past it might have taken us up to three weeks before we could put a validator in production mode. Now that time is reduced to a matter of days.

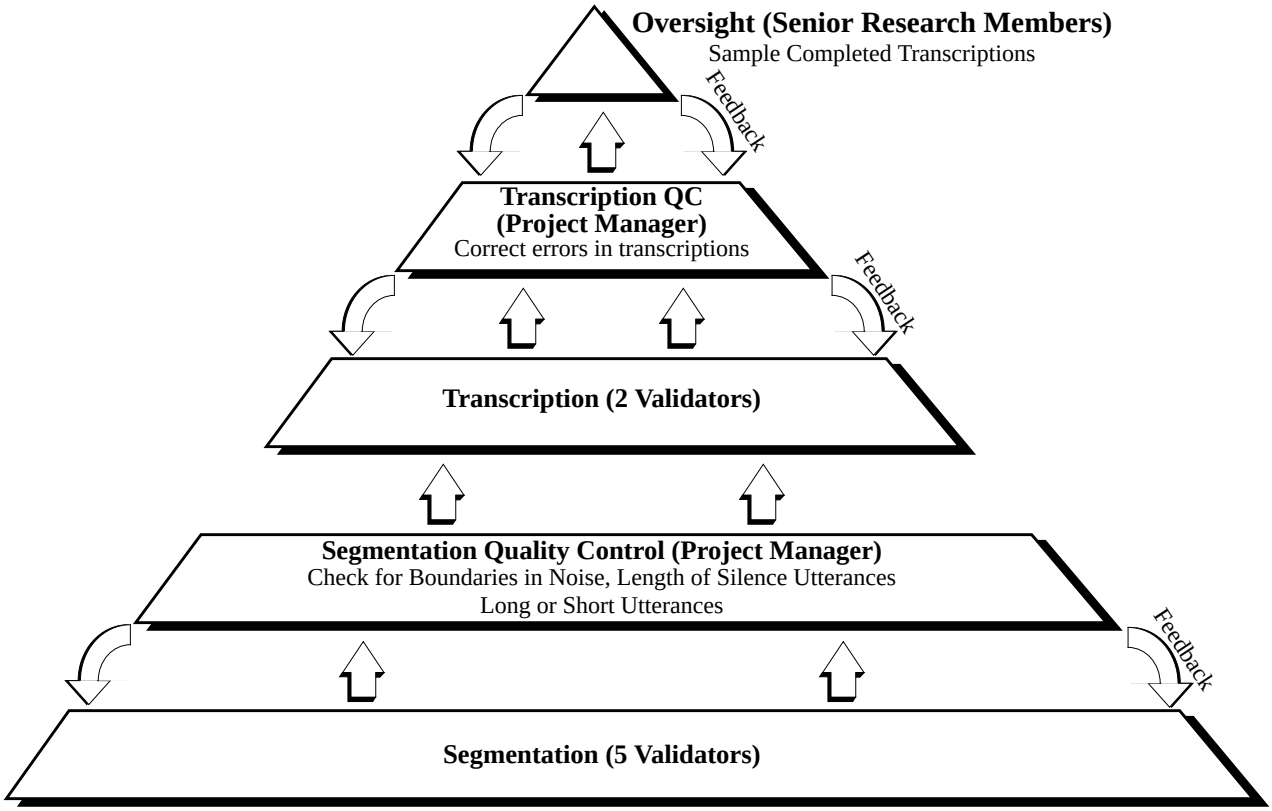

Figure 3. New workflow for segmentation and transcription. Notice the multiple quality control steps and the multiple feedback paths. This environment has resulted in a much more stable set of conventions which are understood and followed by all. In turn, our real-time rates have decreased and our accuracy has increased.

One might think that the total time required to validate an entire conversation would now be greater than if one person did the segmentation and transcription, but we have found this not to be true. Previously, our average validators could segment and transcribe the data at 17 times real-time (xRT). Now, our average segmenter expert can segment data at 7 xRT and our average transcriber can transcribe the data at 8 xRT. So, the real-time rate for an entire conversation has actually lessened, and, as we will show below, we are producing cleaner and more accurate utterances using this method.

In addition to the two validators, this procedure provides a quality control review after both segmentation and transcription. This review is carried out by the project manager and typically will result in the manager reviewing all of the data via the quality control utilities and 10%-20% of the data in detail. In this stage, problem utterances are marked by the quality control scripts and each of these is reviewed and corrected if necessary. This is also when the questions logged by the validators regarding segmentation and transcription are reviewed and decided upon. This quality control step is performed at least once per week so that the project manager is never out of synch with the validators and can address any persistent problem before it permeates a large portion of the database.

The final step in this internal incremental review process is carried out by our best Ph.D. students who have significant experience in building LVCSR systems. Since our goal is to produce data

which can be used by the speech research community to help build robust recognition systems, it is important to view the output of the validation process with a system-building mindset. We ask these researchers to randomly review 2-3 completed conversations out of every 20. At this point, the conversations they review have been through our entire battery of quality control tests and should have zero errors. In this review they are not only searching for any obvious errors but are also looking for persistent problems with the data. It was a review of this sort that turned up our problems with boundaries being marked in echo and noise.

Beyond this set of internal reviews, we have also continued to issue weekly incremental releases of data to the public-domain and to maintain the SWB FAQ [4] and SWB mailing list (*swb@isip.msstate.edu)*. The releases allow the community to use the data as soon as it is available. They also provide interested parties the opportunity to make sure that the data we generate is useful to their research by providing feedback to us. There has not yet been much response to our incremental releases, but our FAQ and mailing list have drawn some very insightful and useful suggestions that we have attempted to incorporate into our work. Marking of asides in conversations is one example of feedback which has been adopted into our working transcription guidelines.

4.2. Details of the quality control process

As mentioned above, we have developed a more strenuous quality control process. At the core of this regimen are a set of utilities that automatically tag utterances which have common errors such as misspellings and boundaries in noise. The sequence of scripts used is shown in Figure 4. Notice that the process is iterative as each marked problem must be adjudicated before the conversation is released. Below, we describe each of the quality control utilities in detail.

> ***check_bounds***: In the early stages of the project the validators were not protected by the segmentation tool from mistakes such as putting the right boundary before the left boundary. The *check_bounds* utility will find all such gross errors in the boundary alignment. The utility verifies that every sample of data in the speech file is accounted for by the transcription start and end times. It does so by making sure that the start time of every utterance (or word in the case of word alignment files) is equal to the end time of the previous utterance or word. It also checks that the end time of the last utterance or word is equal to the last sample in the file and that the start time of the first utterance is zero.
>
> ***check_silence***: One of our transcription conventions is that every utterance marked as containing only silence should be at least 1.0 second long. At times the validators intend to merge a pair of utterances but unintentionally leave a dangling silence-only utterance which is extremely small. This utility finds these problems by tagging all utterances that are transcribed as “[silence]” but are shorter than a specified minimum duration. For our quality control process, the minimum duration is set to 1.0 seconds.
>
> ***utterance_hist***: It is our belief that the average SWB utterance should be between 6 and 8 seconds long and should rarely be greater than 15 seconds or less than 2 seconds. In our review of the data we found that the validators were not paying close attention to these parameters. This *utterance_hist* utility accepts a list of transcription files and for those files flags those utterances whose duration falls outside of the accepted range (2 secs - 15 secs).

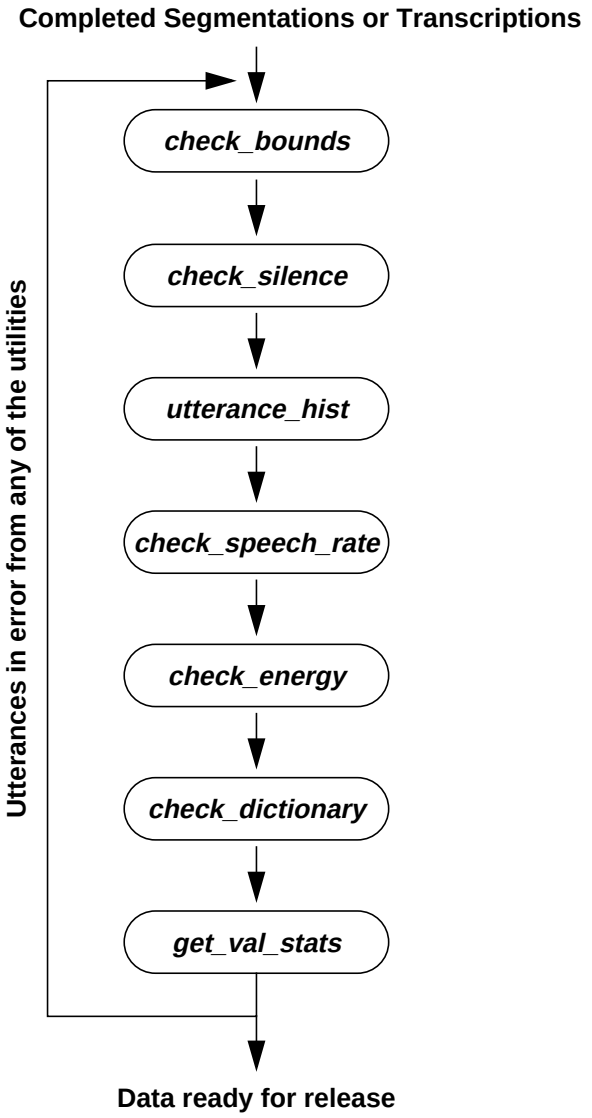

Figure 4. Sequence of quality control utilities used to check for segmentation and transcription errors. This battery of tests is run after both segmentation and transcription of each conversation.

It also produces comprehensive statistics for that list of files including:

- Number of conversations processed
- Number of non-silence and silence-only utterances
- Number of words
- Hours of non-silence and silence-only data in the conversations
- Mean duration of non-silence utterances
- Standard deviation of duration among non-silence utterances
- Maximum and minimum utterance lengths

We use these statistics to characterize the data we are producing and to search for any trends in the data which would lead us towards problems in our transcriptions.

check_speech_rate: We have found that most gross errors in transcriptions such as accidentally replicating part of the transcription twice in one utterance can be easily found by examining the speech rate of each utterance. This is a measure of the number of words transcribed per second of speech in the utterance. We have found that a vast majority of correct utterances have rates between 0.5 and 5.0 words per second. Thus, our quality control script flags any utterances which have speech rates outside of this range. There are, of course, utterances which are in error yet still fall within the range of accepted rates. The number of these is minimal in our released data and will be corrected by the word alignment process.

check_energy: As described earlier, one of the primary reasons for changing our quality control process is to adjust boundaries that are in large bursts of noise or echo. This utility is our primary means for verifying that our validators are following the rules for placing boundaries in low-energy area. The utility uses a standard algorithm [8] to determine the nominal channel energy level. For each utterance in a conversation, *check_energy* finds the average energy of a window around the boundary. If that average energy is larger than the

noise floor of the conversation by a certain amount (typically 25 dB) then the boundary is flagged as occurring in an impulsive noise. This method has been extremely successful in finding boundaries placed in noise or echo and has helped us demonstrate to the validators examples of correct and incorrect boundary placement.

check_dictionary: In our last report we described the revised dictionary built from our improved transcriptions. This dictionary provides a pronunciation for each word in the conversations. With each corrected transcription comes words that are currently not in the dictionary — these are usually partial words, proper names, or laughter words. *check_dictionary* is used to find those words that are not in the dictionary. Each of these are individually reviewed and, if the word is correct in the transcription, are added to the dictionary. This allows us to find any misspelled words or misused words. Using this utility is not foolproof since words can be mistranscribed in the transcription though they do appear in the dictionary. An example of this is a transcription of "World War I" which should be transcribed as "World War One", but since "I" is in the dictionary, *check_dictionary* will allow this phrase to pass. Errors like this will either be caught as we work through the other quality control scripts or when we perform manual word alignments on the data.

get_val_stats: One of the best indicators of our progress in reframing the transcription and segmentation procedures has been the increased performance of our validators accompanied by an increase in accuracy. *get_val_stats* is used to generate statistics on a per-validator basis. With this utility, we can determine the hours of data transcribed, the number of conversations completed and the real-time rates of the validators over a given period of time. We have found that daily feedback to the validators on their real-time rates and data production has been a great motivator for them to continue to work hard.

4.3. Cross-Validation

Cross-validation has become central to evaluating the performance of our validators as well as the quality of our transcriptions. In these tests, each worker validates the same conversation and their transcriptions are compared for accuracy and consistency. Each validator's transcriptions are checked against a reference determined upon careful review by the project manager and a panel of Ph.D. students. This is a blind test, so the validators are unaware that they will be scored on this particular conversation. The transcriptions of each validator and the original LDC transcriptions are compared to the reference to provide an estimate of the improvement in SWB transcriptions after resegmentations.

In the August 1998 report [9], we showed an average validator performance of around 3% on conversation sw3909. To verify that our new procedures were reducing the errors in the conversations, we had a new validator retranscribe the same conversation. The results of this experiment are shown in Table 1. We are encouraged that the error rate for the revised transcriptions have reduced greatly from the previous validation of this conversation. A review of the conversation also shows that the segments now contain more meaningful phrases.

Since making changes to our transcription and segmentation process we have also performed two

cross-validation experiments which are detailed in Tables 2 and 3. The first of these is a transcription cross-validation where the validators transcribed data from the same segmentation of conversation sw2137 and were scored against a reference that was also transcribed from that segmentation. The errors shown in the table are *significant* errors which only include deletion, insertion, or substitution of a word. These specifically do not include minor differences in partial words, differences in transcription conventions (when scoring the LDC data), and marking of noises. The ISIP error rates quoted are averages across all validators. Our worst validator performance was 1.6% and our best was 1.4%. We can see from the table that our revised transcriptions continues to better the LDC transcriptions by a significant margin and we have halved the error rate of our previous work. We fall just short of achieving under 1% on this data. We believe that our manual word alignment efforts will be able to bring this result to less than 1% as predicted.

Table 3 shows results of our segmentation cross-validation. In this test, we compare the validators boundary locations with the reference segmentation. Any boundary that is in the same area — within .2 secs — of the reference location and is not in noise or echo (if possible) is considered to be a correct segmentation. The table shows that our new process has brought about more consistent markings of the boundaries as well as boundaries that more closely conform to the preferred segmentation. Two of the validators who participated in this cross-validation had only been training for one week. Each performed as well as veteran validators who had been segmenting data for months. Our ability to quickly train new validators to perform at the level of veteran validators in one week's time has helped our production rate tremendously.

Transcriber	WER
LDC	7.9%
ISIP before revisions	2.7%
ISIP after revisions	0.7%

Table 1. Transcription error rates on sw3909 for the LDC transcriptions and for the ISIP transcriptions before and after our revised guidelines. These errors do not include those due to convention differences, marking of noise or partial-word marking.

Transcriber	WER
LDC	5.4%
ISIP before revisions	3.7%
ISIP after revisions	1.5%

Table 2. Transcription error rates on sw2137-A for the LDC transcriptions and for the ISIP transcriptions before and after our revised guidelines.

	Error Rate
before revisions	7.0%
after revisions (senior validator)	0.0%
after revisions (new validator)	1.41%

Table 3. Segmentation error rates on sw4045-A for the ISIP data before and after implementation of our new guidelines. We have now achieved consistency amongst the validators.

5. ANALYSIS OF RELEASED DATA

Thus far, we have corrected and released over 150 conversations which conform to what we believe are our final transcription conventions. This set covers the WS'97 devtest and eval set as well as a portion of the WS'97 training set. Table 4 illustrates the details of the released data. The following sections give an interesting analysis of the data available at present.

Conversations	147
# of non-silence utterances	11178
# of silence-only utterances	6306
# of words	141167
hours of data	23.3
hours of speech data	14.8
Mean utterance duration (seconds)	4.8

Table 4. Statistical analysis of the released data to date

5.1. Words per utterance

A primary goal of this work is to produce utterances which contain meaningful phrases such as sentences or complete thoughts. From this, one would think that, on average, the number of words per utterance would be large. The histogram of Figure 5 bears this out but also reveals an interesting trend. From the figure, one sees

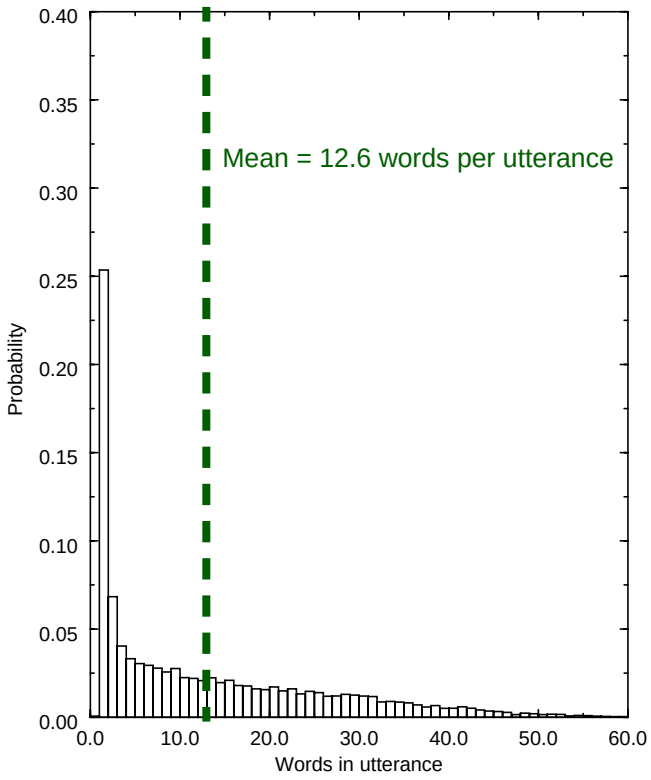

Figure 5. Histogram of words per utterance. There is a large bin due to one-word utterances and the distribution tapers off thereafter.

Word	Count	Cumulative coverage
yeah	738	26.3%
um-hum	674	50.3%
uh-huh	391	64.2%
right	203	71.5%
hm	97	74.9%
oh	95	78.3%
okay	71	80.8%
um	61	83.0%
yes	42	84.5%
huh	41	85.9%
so	33	87.1%
sure	31	88.2%
no	28	89.2%
uh	26	90.2%

Table 5. Distribution of words in the one-word utterances. Not surprisingly, affirmative statements and pause fillers make up the majority of these utterances.

that over 25% of the utterances are one-word utterances explaining the relatively short mean utterance duration from Table 4. There is a long tail after the one-word utterances which gives a mean value of over 12 words per utterance.

The more significant result of this plot is related in Table 5. Four words (all affirmations) account for over 70% of the one-word utterances. With only 14 words we can cover 90% of the one-word utterances. It is likely that this information could be used to tune a language model to short utterances such as affirmations or to constrain advanced systems which are able to determine the number of words in the utterance before hand.

5.2. Utterance lengths

We believed that the majority of SWB utterances containing a single phrase would be between seven and eight seconds in length with sufficient silence buffers. The histogram of Figure 6 tells a different story. A large portion of the utterances (close to one-third) are less than two seconds long. This directly correlates with the distribution shown in Figure 5 where the one and two-word utterances are dominant and is a fall-out of conversational speech — one-word replies abound. If we remove the one-word utterances from the data then we do find that the distribution of utterance lengths has a mean of close to 6.5 seconds which is more reasonable for the desired long phrases.

5.3. Speech rate

It is well known that speech rate is directly related to one's ability to accurately transcribe speech data. The speech rate is also proportional to speech recognition system performance. In Figure 7 we see that the SWB speech rates actual take on a bimodal distribution. A large percentage of the utterances have speech rates less than one word per second. For the most part, these are one-word utterances where the amount of speech used to calculate the speech rate is masked by the silence buffers on either side. Our quality control scripts flag all utterances with speech rates less than 0.5 words per second and greater than 4.5 words per second.

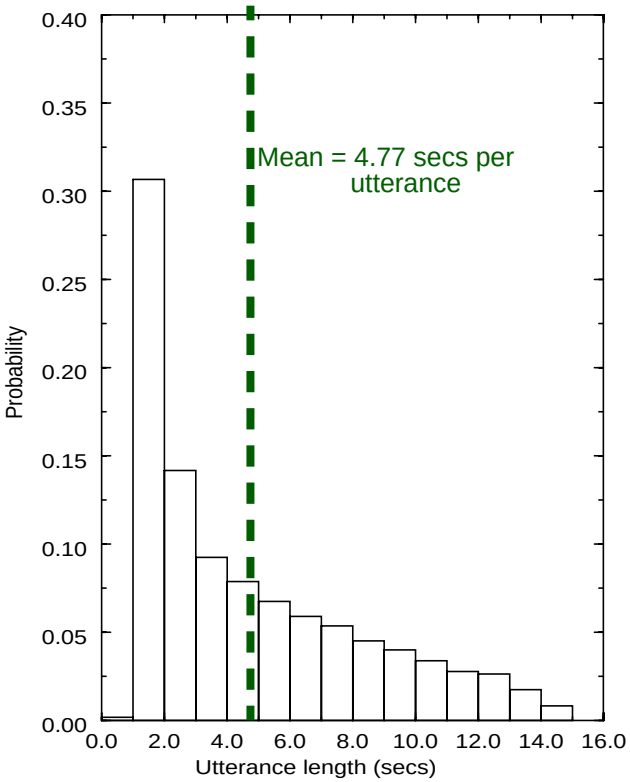

Figure 6. Histogram of utterance durations. The peak close to one second directly corresponds to the large number of one-word utterances.

6. PLANS AND ISSUES

As of November 15, we have released over 150 revised conversations with new segmentations and transcriptions. We have also resegmented close to 400 more conversations. Though this seems like a

step backwards from our previous report, the transcriptions and segmentations being produced now correct the problems present in our last releases which would have rendered them problematic for use in research. We have greatly improved our quality control procedures by using a multi-pass review process to insure accuracy and conformity to the guidelines laid out in Appendix A. Most importantly, we have managed to train our validators to generate more accurate transcriptions and segmentations in a lesser amount of time.

Though we have experienced a setback from the goals stated in the last project report, we believe that our new process will allow us to make up that time quickly. A revised timeline of our expected progress is shown in Figure 8. Our goal is to re-release the 1000 previously released conversations with corrected segmentations and transcriptions by January 1, 1999. We have added three new validator positions to offset the time spent in revision. This increased work output should allow us to catch up with our original timeline by March 1999. We had projected a release of 1800 conversations by that point and we believe that this is still a realistic expectation. By July, we plan to have completed all of the conversations and be working solely on word alignments and extended quality control work for the entire database of transcriptions.

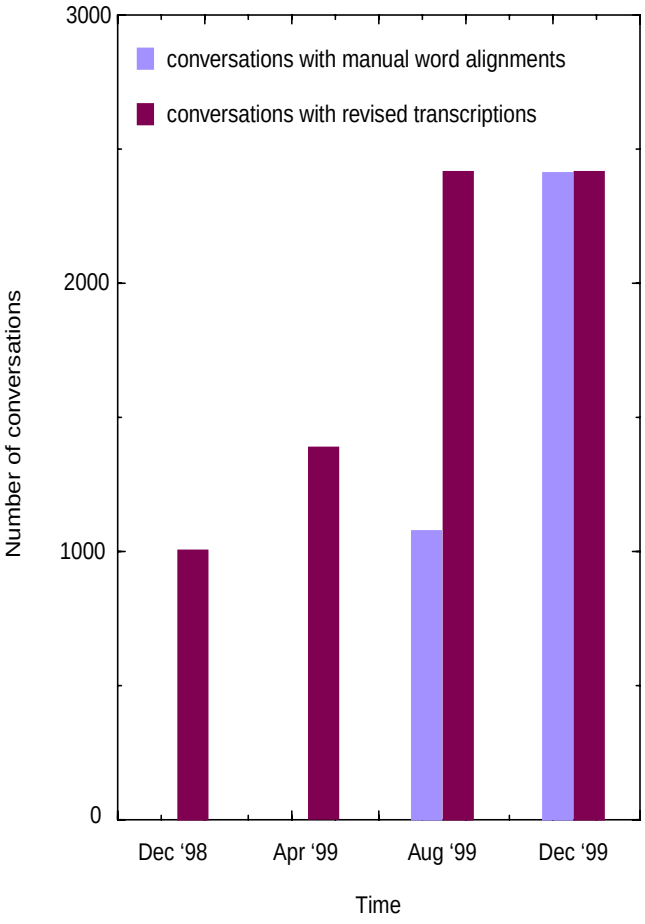

Figure 8. Timeline for the remainder of the SWB resegmentation project. Our anticipated completion date remains at December 1999.

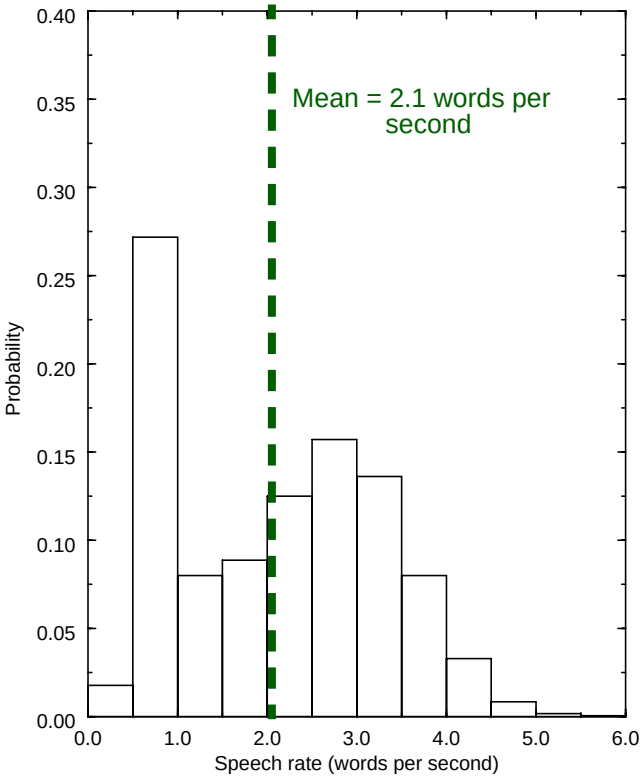

Figure 7. Histogram of utterance speech rates for the released data. This takes a multi-modal distribution for the short utterances and the longer ones

As part of our ongoing work with the SWB project and our public-domain speech recognition system [10], we also plan to begin running experiments with the new data. With training on the previously released data, we found that our

WER decreased by almost 2%. We expect these results will hold with the next release of data since the revised segmentations and transcriptions are better suited as training data than the previous release. In January of 1999, we plan to repeat these experiments using the 1000 revised conversations as training data and will report these results in our Spring progress report.

7. ACKNOWLEDGEMENTS

We greatly appreciate the increased support of our efforts from the speech research community over the past three months. This project is producing data that we hope will aid our colleagues in building robust LVCSR systems so we are encouraged by the interest many have taken in the SWB work. To all who have given comments, suggestions and encouragement during the recent months of this project, we extend our deepest appreciation. In particular, the assistance of Dr. William Fisher of NIST has been invaluable in helping us to conform to industry-standard practices in transcribing data. Our continued gratitude is extended to the LDC for supplying us with copies of the SWB CDs and for helping us with many transcription and database-related issues. We would also like to express our continued appreciation to Dr. Jack Godfrey for his continued support in all things related to the SWB Corpus, linguistics, and data collection.

8. REFERENCES

[1] J. Godfrey, E. Holliman and J. McDaniel, “Telephone Speech Corpus for Research and Development,” *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 517-520, San Francisco, California, USA, March 1992.

[2] B. Wheatley, G. Doddington, C. Hemphill, J. Godfrey, E.C. Holliman, J. McDaniel, and D. Fisher, “SWITCHBOARD: A User’s Manual,” *http://www.cis.upenn.edu/~ldc/readme_files/switchbrd.readme.html*, Linguistic Data Consortium, University of Pennsylvania, December 1995.

[3] N. Deshmukh, A. Ganapathiraju, R. Duncan, and J. Picone, “An Efficient Tool For Resegmentation and Transcription of Two-Channel Conversational Speech,” *http://www.isip.msstate.edu/resources/technology/software/1998/swb_segmenter,* Institute for Signal and Information Processing, Mississippi State University, August 1998.

[4] J. Hamaker and J. Picone, “The SWITCHBOARD Frequently Asked Questions (FAQ),” *http://www.isip.msstate.edu/resources/technology/projects/current/switchboard/faq*, Institute for Signal and Information Processing, Mississippi State University, August 1998.

[5] J. Hamaker and J. Picone, “A Statistical Guide to SWITCHBOARD: Topic Statistics,” *http://www.isip.msstate.edu/resources/technology/projects/current/switchboard/doc/ statistics*, Institute for Signal and Information Processing, Mississippi State University, August 1998.

[6] J. Hamaker, A. Ganapathiraju, and J. Picone, “SWITCHBOARD Educational Resources,” *http://www.isip.msstate.edu/resources/technology/projects/current/switchboard/doc/education*, Institute for Signal and Information Processing, Mississippi State University, August 1998.

[7] J. Hamaker, Y. Zeng, and J. Picone, “Rules and Guidelines for Transcription and Segmentation of the SWITCHBOARD Large Vocabulary Conversational Speech Recognition Corpus,” *http://www.isip.msstate.edu/resources/technology/projects/current/switchboard/doc/transcription_guidelines,* Institute for Signal and Information Processing, Mississippi State University, July 1998.

[8] J. Picone, “Fundamentals of Speech Recognition: A Short Course,” *http://www.isip.msstate.edu/resources/courses/isip_0000/lecture_notes.pdf*, Institute for Signal and Information Processing, Mississippi State University, May 1996.

[9] J. Hamaker, N. Deshmukh, A. Ganapathiraju, and J. Picone, “Improved Monosyllabic Word Modeling,” *Department of Defense*, August 15, 1998.

[10] N. Deshmukh, A. Ganapathiraju, J. Hamaker and J. Picone, “Large Vocabulary Conversational Speech Recognition”, *http://www.isip.msstate.edu/resources/technology/projects/1998/speech_ recognition/*, Institute for Signal and Information Processing, Mississippi State University, 1998.

APPENDIX A. GENERAL INSTRUCTIONS FOR SWITCHBOARD TRANSCRIPTIONS

This document is structured into two sections: the original SWITCHBOARD (SWB) transcription guidelines and the ISIP modifications to this standard. Historically, the problem with any SWB convention document has been that the data delivered does not conform to the guidelines. Hence, the ISIP modifications are somewhat a documentation of what conventions are embedded in the current corpus, along with some new conventions based on the goals of our project. The ISIP modifications appear first followed by the SWB standard. If a particular issue is not covered in the ISIP amendments section, then assume we are following the original SWB convention.

A.1 Appended Instructions

The following guidelines for segmentation and transcription of SWB take precedence over the original SWB transcription conventions supplied by LDC (and described in Section A.2).

A.1.1. Segmentation

The original goal of this project was to provide a new segmentation of the database to support improved acoustic training for speech recognition. It is important to remember this goal when discussing the challenging problem of SWB segmentation. Note that we do not pay attention to turns and such linguistic phenomena in performing the segmentation. Our segmentation will be largely based on the acoustic data.

Conversations will be broken into a sequence of segments which we refer to as utterances. Utterances will consist of either speech padded by 0.5 secs of silence on each side, or consist of only silence (background noise). Further, a design goal of the project is that an utterance be no more than 15 seconds in length. Ideally, breakpoints will be inserted at natural linguistic points in the utterance such as sentence or phrase boundaries. When no suitable boundary can be found, we progressively relax the requirement that the silence padding be 0.5 seconds in duration. Below are some general rules about segmentation.

1. Each utterance should be padded by a nominal 0.5 second buffer of silence on both sides. In general, these silence buffers can range from 0.35 to 0.75 seconds.
2. The boundary can **only** be placed in a “silence” consisting solely of channel noise and background noise. Whenever possible place the boundary in a section with very low energy (visually speaking, this is a flat part of the signal in the segmentation tool)
3. The 0.5 second buffers can contain breath noises, lip smacks, channel pops, and any other non-speech phenomena. However the boundary **can not** be placed in a noise of this sort.
4. No utterance can be longer than 15 seconds. As an utterance approaches 15 seconds in length, the validator is allowed to find a point of segmentation that will generate silence buffers less than 0.5 seconds but not less than 0.1 seconds. If this segmentation can not be found then that utterance should be marked as “NEEDS_REVIEW” in the log file and the validator should send an e-mail to the adjudication team explaining the problem.
5. Every utterance containing only silence must be greater than 1.0 seconds in duration.

6. Whenever possible choose a segmentation that maintains the phrase structure of the conversation. This means that, ideally, we would like every utterance to contain a single phrase. However, due to the nature of the SWB data, we realize that this is not always possible. **Note**: The previous instructions take precedence over this one.
7. The end of the preceding utterance coincides with the start of the next utterance. Hence all data is accounted for. Segmentation essentially involves placing a boundary between two utterances.
8. Consider a stretch of silence which has small amplitude noises embedded in it as a silence only utterance - do not mark the noise and do not segment the noises into separate utterances. However, if a noise has a particularly high amplitude, then segment it into its own utterance.

A.1.2. Transcription

1. Transcribe “verbatim,” without correcting grammatical errors: “i seen him,” “me and him gone to the movies,” etc.
2. Standard reductions and alternate pronunciations: Unless otherwise noted below, if “no” is meant but said as “naw” or “nah”, transcribe it how it is spoken. e.g. “y’all” instead of “you all”; “gonna” instead of “going to”; “wanna” instead of “want to”. However, in cases where there is severe reduction of a preposition such as in “kinda”, “sorta”, “gotta”, etc., transcribe the phrase as it was intended to be spoken. e.g. “kind of”, “sort of”, “got to”.
3. Follow the dictionary on hyphenating compounds in clear-cut cases. But “when in doubt, leave them out.”
4. Compound words: All compound words should be transcribed as one word when such a word exists in the dictionary unless there is an acoustical pause between the two words. e.g. “someone”, “everyday”, “cannot”, etc.
5. Try to avoid word abbreviations: Fort Worth, not Ft. Worth; percent, not %; dollars, cents, and so forth.
6. Contractions are allowed. e.g. “there’ll”, “it’s”, “can’t”, etc.
7. Capitalization: Use normal capitalization on proper nouns. Do not capitalize the beginning of the sentence. Titles should be capitalized using the standard grammar rule: the first word of a title is always capitalized, prepositions within a title that are under five letters are always lowercase, and the last word of a title is always capitalized.

   Example: “Dances with Wolves”, “Gone with the Wind”
8. The pronoun “I” should not be capitalized, instead it should be typed as “i”. Titles containing the word “I” are exceptions to this rule.

   Examples: i am tired of talking to you

   are you as tired as i am of listening to this
9. No punctuation should be used in the transcriptions.
10. Remember to watch for common spelling confusions like: its and it’s, they’re, there and

their, by and bye, to and too, etc.

11. Numbers: Spell out all number sequences except in cases such as “123” or “101” where the numbers have a specific meaning. Transcribe years like 1983 as spoken — “nineteen eighty three.” Do not use hyphens (“twenty eight”, not “twenty-eight”).

12. Letter sequences: Spell out letter sequences: DFW, USA, FBI, NASA, ROM. When a letter sequence is used as part of an inflected word, add the inflection to the end of the letter sequence: e.g. TIer, BSing, the Oakland As, a witness IDed him. Transcribe a spoken spelling in all capital letters, each separated by a space: e.g. “dog is spelled D O G”; “my name is Tirelly, that’s T I R E L L Y”. If letter sequences contain special symbols then transcribe them as they would be written not as they are spoken: e.g. “AT&T” not “AT and T”; “Texas A&M” not “Texas A and M”.

13. Classifications of music are not titles, should not be transcribed in uppercase: e.g. “country western”, not “Country Western”; “rock ‘n’ roll”, not “Rock ‘n’ Roll”.

14. Possessives: Use standard grammar rules to denote possession: the US’s policy, Sally’s book, the drivers’ cars, the CEO’s decision, the dancers’ shoes.

15. Partial words: If a speaker does not completely pronounce a word and the word is not a standard reduction then spell out as much of the word as is pronounced, and inside brackets spell out the part of the word that was not pronounced. Use a single dash after the brackets if the last part of the word was not pronounced and a single dash before the brackets if the first part of the word was not pronounced to flag that a partial word was spoken. Context should be used to determine what word was intended to be spoken. If, from context, a reasonable intended word can not be determined, mark it as [vocalized-noise]

    Example: If a person begins to say the word “went” but only pronounces the “w”, transcribe it as “w[ent]-”.

    If a person says only the “at” portion of “that”, transcribe it as “-[th]at”.

16. Restarts of “i”: If a speaker restarts when saying the word “i”, it should be transcribed as “i-”. This should only be used when the first “i”s are not completely pronounced.

    Example: i- i really felt like i’ve been working now for about four years

17. Mispronunciations: If a speaker mispronounces a word and the mispronunciation is not an actual word, transcribe the word as it is spoken followed by the word that was intended. Divide these two words by a forward slash and enclose both words in brackets.

    Example: i wasn’t sure that they were blaming that [splace/space] space disaster on one company

18. Coinages: If a speaker uses and gives meaning to a word that is not an actual word, spell the word out as it sounds and enclose it in braces.

    Example: How are things for you {weatherwise}

19. Asides: If one of the speakers involved in the conversation talks to someone in the background and the words can be understood, then transcribe it as an aside enclosed in the markers, <b_aside> and <e_aside>. This only applies if one of the conversation

speakers is involved in the background conversation. If just background speakers can be heard then this can be thought of either as noise or background noise depending energy level of the background speakers. compared to the foreground speakers.

Example: “yeah i know what you <b_aside> honey i can’t play with you right now i’m on the phone <e_aside> sorry you know kids always want mommy all to themselves”

20. Hesitation sounds: Use “uh” or “ah” for hesitations consisting of a vowel sound, and “um” or “hm” for hesitations with a nasal sound, depending upon which transcription the actual sound is closest to. Use “huh” for the aspirated version of the hesitation as in: "huh? <other speaker responds> um ok, i see your point."

21. Yes/no sounds: Use “uh-huh” or “um-hum” (yes) and “huh-uh” or “hum-um” (no) for anything remotely resembling these sounds of assent or denial; you may use “yeah,” “yep,” and “nope” if that is what the words sound like.

22. Non-speech sounds during conversations: transcribe these using only the following list of expressions in brackets:

    [laughter] [noise] [vocalized-noise]

    Pick the closest description ([noise] will be adequate in most cases).

23. Laughter during speech: If laughter occurs directly before a word, place the [laughter] tag before the spoken word. If laughter occurs after a spoken word, place the [laughter] tag after the word. If the speaker laughs while saying the word, but the word is still understood, transcribe this as [laughter-word], where "word" is the word spoken during the laughter. If the speech is obliterated by the laughter, transcribe it strictly as [laughter]. If a speaker laughs while saying several words and the words are understood, transcribe each word in the phrase as [laughter-word]. Laughter throughout the phrase, “you don’t say,” would be transcribed as: [laughter-you] [laughter-don’t] [laughter-say].

24. Pronunciation variants: The convention of "word_1" is used to denote a common variation in the pronunciation of a word. A list of these words will be kept in the transcription conventions documentation. Examples of pronunciation variants currently in use are:

about_1	b aw t	because_1	k ah z
depends_1	p eh n d z	them_1	eh m
okay_1	m k ey	especially_1	s p eh sh ax l iy

    These are to be used judiciously, and only to capture frequently occurring reductions which are easy to distinguish from the baseform.

25. Continuous background noise: Consider it as part of channel. For example, if a baby cries at a consistent energy level throughout the conversation then treat it as background noise. Only consider it as noise if the noise grows much louder than the normal level — in our example above the baby screaming would warrant considering it as noise. In this case mark it as [noise].

26. Special lexicon issues:

- Use "all right" instead of “alright” in all cases.
- Use "Walkman" when the speaker is referring specifically to the Sony Walkman, and use “walkman” when there is no reference to Sony.

  Example: i like to listen to my walkman when exercising
  i wonder how many transistors a Sony Walkman has?

- Use “doggy” instead of “doggie” in all cases.
- Use "God" instead of "god" in all cases.

  Example: it’s like you know God what are they doing

A.2. Original Instructions

Following is the original set of guidelines and instructions for transcription of SWITCHBOARD. We propose to deviate from these in a manner explained previously in Section 1.

A.2.1. General Instructions

1. Transcribe “verbatim”, without correcting grammatical errors: "i seen him,” “me and him gone to the movies,” etc.
2. Do not try to imitate pronunciation; use a dictionary form: “no” will do for “naw,” “nah,” etc., “oh” for “aw,”; “going to” (not gonna or goin to); “you all” rather than “y’all”; “kind of” instead of “kinda”; etc. Nonstandard words which are not in the dictionary (e.g., kiddo) should be typed normally, i.e. without quotes or other special notation.
3. Follow the dictionary on hyphenating compounds in clear-cut cases. But “when in doubt, leave them out.”
4. Try to avoid word abbreviations: Fort Worth, not Ft. Worth; percent, not %; dollars, cents, and so forth.
5. Contractions are allowed, but be conservative. For example, contraction of “is” (it’s a boy, running’s fun) is common and standard, but there’ll (there will) be forms that’re (that are) better left uncontracted. It is always permitted to spell out forms in full, even if the pronunciation suggests the contracted form. Thus it is O K to type he is and they are and we would even if it’s he’s and they’re and we’d you heard.
6. Use normal capitalization on proper names of persons, streets, restaurants, cities, states, etc., but put titles (of books, journals, movies, songs, plays, TV shows, etc.--what would properly be in italics.) in ALL CAPS, i.e., uppercase letters.
7. If it is necessary to use accent marks, insert the number 3 before the letter which would receive the accent, e.g., fianc3e.
8. Punctuation: although normal punctuation rules apply, spontaneous conversational speech is full of difficult situations. Strive for simplicity and consistency, with the following specific guidelines:
   - terminate each sentence with a period unless a question mark or exclamation

point is clearly justified;

- use a comma instead of ... or -- or fancier punctuation when speakers change thoughts or grammatical structures in the middle of a sentence;
- for more detail, and for special rules involving interruptions, etc., see below under SPECIAL CONVENTIONS.

9. Be sure to run a spell check upon completion of the transcript. Remember to watch for common spelling confusions like: its and it's, they're and there and their, by and bye, etc.

A.2.2. Special Conventions for SWITCHBOARD Conversations

1. Speakers should be indicated by "A: " and "B: " at the left margin, with two spaces after the colon, and with a blank line between speakers (i.e., an extra carriage return before each A: or B:). On the audio tape, A will be THE SPEAKER ON THE FIRST OF THE TWO SEPARATELY RECORDED SIDES. IT IS IMPERATIVE TO KEEP THIS DESIGNATION CORRECT AND CONSISTENT, even when the crosstalk or echo is so strong that both speakers are equally loud. The log sheet for each conversation will show the first few words by each speaker, to help you confirm the assignment.

   EXAMPLE:
   A: Blah blah blah blah.
   B: Blah blah blah.
   A: Etcetera.

2. Spell out letter and number sequences: D F W, seven forty-seven, US A, one oh one, F B I, etc., unless the letter sequence is pronounced as a word, as in NASA, ROM, DOS.
3. Transcribe years like 1983 as "nineteen eighty-three," with hyphens only between the tens and ones digits.
4. When a letter sequence is used as part of an inflected word, add the inflection with a dash: T I -er, B S -ing, the Oakland A -s, a witness I D -ed him. This leads to clumsy-looking possessive forms, as in: the U S -'s policy, the T I -er's last name, all the C E O -s' votes, but it saves lots of time later on.
5. Partial words: if a speaker does not finish a word, and you think you know what the word was, you may spell out as much of the word as is pronounced, and then use a single dash followed by a comma, -,. If you cannot tell what word the speaker is trying to say, leave it out.

   EXAMPLE:
   A: Well, th-, that's what they kept tell-, wanted me to believe.
   B: I, I, I just am not to-, totally sure, uh, about that.

6. Hesitation sounds: use "uh" for all hesitations consisting of a vowel sound (rather than trying to distinguish uh, ah, er, etc.), and "um" for all hesitations with a nasal sound

(rather than uhm, hm, mm, etc.)

7. Yes/no sounds: use “uh-huh” (yes) and “huh-uh” (no) for anything remotely resembling these sounds of assent or denial; you may use “yeah,” “yep,” and “nope” if that is what the words sound like.

8. Punctuation: use commas instead of ... or -- or other “fancy” punctuation when speakers change thoughts or grammatical structures in the middle of a “sentence.” Terminate each sentence with a period unless a question mark or exclamation point is clearly justified. Only use suspension dots ... if a speaker leaves a sentence unfinished at the end of his/her turn, and a period cannot be used, or at the end of a conversation where the speaker’s turn was cut off by the computer timing out:

   EXAMPLE:
   A: I was going to do that, but then I ...
   B: Right, me too.

9. Use a double dash if a speaker breaks a sentence off and picks it up at the beginning of the next turn, with another double dash where the pickup begins:

   EXAMPLE:
   A: I was going to do that, but then I --
   B: Right, me too.
   A: -- thought I better not after all.

10. Non-speech sounds during conversations: indicate these using only the following list of expressions in brackets. When making judgments, pick the closest description; [noise] will be adequate to describe most sounds that are not represented below. Note underscores (not spaces or hyphens) to connect the double word descriptions.

11. If the event being described lasts longer than a few words, then indicate the beginning in brackets [ ], and the end in brackets with a “/”, [/].

    EXAMPLE:
    1. Separate multiple sounds by a space, each one in brackets:
       A: Oh, that’s funny. [laughter] [cough] Excuse me, I have a cold.
       B: That’s all right, [sneezing] so do I. [barking] [child_talking]

    2. Use “/” to show end of a continuous sound:
       A: Well, it all depends, uh, on, you know, [baby_crying] how the family reacts. I mean, it can be a positive or a negative thing, you know?
       B: Yeah, well, I guess so. It just seems [/baby_crying] to me that it’s a very difficult, uh, difficult issue.

12. When a comment is needed to describe an event, put the comment in curly braces { }: {very faint}, {sounds like speaker is talking to someone else in the room}, {speaker imitates a woman’s voice here}.

EXAMPLE:

1. Curly braces to describe the speech:
   B: Yeah, yeah, I agree {very faint} right.

2. Combine curly braces and brackets if more explanation is needed to describe the word in the brackets:
   A: Did it sound like this? [clicking] {sounds made with mouth}
   B: No, more like [clicking] {sounds like a pencil tapping on a table} this.

13. When a word or phrase is not clear, type DOUBLE PARENTHESES (( )) around what you think you hear. If there is no way to tell what the speaker said, leave 1 blank space between the double parentheses, indicating speech has been left out because it was unintelligible.

EXAMPLE:

A: So when I finally did ((take up)) the violin, progressed pretty quickly in the beginning.
B: Of course, that was in college which was a long time ago, so (( )) I remember.

14. Marking untopical speech for possible trimming: Use an “at sign”, @, and a double “at sign”, @@, to designate potential “trim points” at the beginning or end of conversations. These would exclude speech that either is not part of the conversation itself, or refers directly to the protocol. For example, it sometimes happens that callers accidentally press the touchtone button that begins recording, and are being recorded during the “warmup period” and don’t know it. All such speech should be marked for trimming. Other examples would be speech that:

    a) explicitly refers to the SWITCHBOARD protocols;
    b) refers to the process of making the call;
    c) uses the TITLE of the prompt (e.g., “music”); or
    d) repeats or paraphrases the PROMPT itself.

15. [The TITLE and the PROMPT for each topic will be found on your information sheet; they are keyed to the topic number, which is on the log sheet for each conversation.]

16. Marking these trim points means that EVERYTHING BEFORE ‘@’ AND/ OREVERYTHING AFTER ‘@@’ may be discarded without losing the main body of the conversation on the topic. These symbols may therefore only be used ONCE AT THE BEGINNING (@) AND/OR ONCE AT THE END (@@) of the conversation. They must also be used ONLY AT TURN-TAKING POINTS, i.e., at the left hand margin, before an “A:” or “B:”, NOT part of the way through someone’s turn. One or both may be used in a single conversation, i.e., trimming of material at the beginning is independent of trimming at the end.

17. Social niceties and transitional talk are neutral. That is, they may be left alone, but should be trimmed if they occur next to material that definitely deserves trimming.

EXAMPLE:

A: Okay, so what am I supposed to do now? Wait, let me read,
B: I think you're supposed to push one.
A: let's see, it says here to push, okay, but I think I already, okay are you ready?
B: Yep. [Talking about protocol up to here.]
A: Here we go. Alright, now, tell me, what is your favorite kind of music? [Using topic TITLE explicitly.]
@B: I enjoy Mozart and reggae, but I really love rap. [OK]
.
. <body of conversation is here>
.
A: I've certainly enjoyed hearing what you have to say. [Trim optional here.]
@@B: Well, if we've talked enough, do I need to push a button or anything? I guess not, we can just hang up. So long. [Talk of protocol should be trimmed.]
A: Bye. Nice talking to you.

ANOTHER EXAMPLE:

A: Hi, there, how are you doing?
B: Fine, how about you?
A: Just great, except for all this heat. [Chitchat up to here could be left alone if no reason to trim occurred.]
B: Well. Care of the elderly, huh? That's our topic? [Need to trim because it mentions the topic TITLE.]
@A: Yes. Do you have any relatives that need special care? [This is OK as part of the conversation, since only the word "care" is repeated from the prompt. It is not trimmed--initial trimming ends with the '@'.]
.
.
.
@@B: Well, I guess we have solved the problem of care of the elderly, and how to choose nursing homes, haven't we? [Trimmed because it contains both TITLE and a paraphrase of prompt.]
A: Sure did. I hope your grandmother gets better. So long now, it's been fun talking to you. [Social pleasantries would not be trimmed themselves, but no harm in trimming them in order to get rid of the previous turn.]

18. Simultaneous talking: Wherever possible, mark where both speakers talked simultaneously with TWO PAIRS of pound signs (#), ONE BEFORE AND ONE AFTER each of the segments spoken at the same time. One of these segments MUST BEGIN A TURN; in other words, if one person is an "interruptor", his interruption starts a new turn. Remember, BOTH speakers' turns must contain TWO pound signs each.

A SIMPLE EXAMPLE:

A: Okay, well, I guess that's about it.

B: Yeah.
A: Nice talking to you.
B: # Right, bye. #
A: # Bye bye. #

ANOTHER EXAMPLE:

A: I never heard such nonsense, you know,
B: # Yeah, I know. # [B interrupts while A continues.]
A: # as I heard that # day when I blah blah blah. [A continues beyond the simultaneously spoken words.]

WHICH COULD ALSO BE WRITTEN:

A: I never heard such nonsense, you know, # as I heard that #
B: # Yeah, I know. #
A: day when I blah blah blah

ANOTHER EXAMPLE:

A: I never heard such nonsense, # you know, # [A starts.]
B: # Yeah, # [B starts to step on A.]
A: as I heard that day when # I was at that meeting. # [A continues without stopping.]
B: # I agree with you all the way # [B comes in over A again.]