Distributed Listening in Automatic Speech Recognition

by

Yolanda McMillian

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 9, 2010

Keywords: Automatic Speech Recognition, Spoken Language Systems, Distributed Listening

Approved by

Juan E. Gilbert, Chair, Professor of Computer Science and Software Engineering
Cheryl Seals, Associate Professor of Computer Science and Software Engineering
Gerry Dozier, Professor of Computer Science and Software Engineering

Abstract

While speech recognition systems have come a long way in the last forty years, there is still room for improvement. Although readily available, these systems are sometimes inaccurate and insufficient. The research presented here outlines a technique called Distributed Listening which demonstrates noticeable improvements to existing speech recognition methods. The Distributed Listening architecture introduces the idea of multiple, parallel, yet physically separate automatic speech recognizers called listeners. Distributed Listening also uses a piece of middleware, called an interpreter, which resolves multiple interpretations using a phrase resolution algorithm. The subsequent experiments of the research show that these efforts work together to increase the accuracy of the transcription of spoken utterances and Distributed Listening at worst, is as good as the best individual listener.

Acknowledgments

I want to thank my family and friends who provided unending support and encouragement through this entire process. Specifically, my mother, Eloise McMillian, my sisters, Shauna and Tonya McMillian, my father Joseph McMillian, who lives on in the loving memory of his family, and my support circle, Hilary Boyd, Regina Bolden, Nicole Harris, Cynithia Landry, Tamika Austin, Kenitra Fewell, and Shalonna Banks.

I especially want to thank my advisor, Dr. Juan Gilbert, for his patience and mentoring. I would also like to thank my committee members, Dr. Gerry Dozier and Dr. Cheryl Seals, along with my outside reader, Dr. Jared Russell, for their assistance in making this possible.

I also must thank the members of the Human Centered Computing Lab, both past and present, for their unconditional help and advice, especially Jerome McClendon, Philicity Williams, Dr. Ken Rouse, Dr. E. Vincent Cross, Kamilah Walker, Michele Williams, Kenishia Sapp, and Dr. Dale-Marie Wilson.

Most of all, I give thanks to my Lord and Savior Jesus Christ, because through Him, I can do anything but fail.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation

Research in the area of speech and natural language processing has been on-going for over forty years (Natural Language Software Registry 2004 and Jurafsky 2000) with foundations in a number of overlapping disciplines (Jurafsky 2000); however, there is still room for improvement with mainstream speech recognition systems (Deng and Huang 2004). Spoken language is quite pervasive which leads to frustrations when spoken language systems do not meet a user's expectations. In theory, these systems have the ability to save both time and money, all while executing on a consistent basis, something humans are not easily able to do. For example, an automated system can respond, "…tirelessly, patiently, perkily, consistently, and to the best of her abilities" time after time (Price 2010). Yet, the annoyance of such systems to a user far outweighs these advantages. Until these systems outperform a typical person in most areas, there will always be room for improvement. Additionally, until systems achieve a conversational and casual style of speech interaction, the challenges for current speech recognition technology will persist (Deng and Huang 2004). A final fundamental obstacle of mainstream spoken language systems is overcoming accuracy rates in noisy environments. Although strides have been made, there are still practical limitations that need to be addressed (Deng and Huang 2004). Since ASR systems have the ability to be superior to people with regard to consistency and data management, research of these systems is unending.

Distributed Listening will further research in this area. The concept is based around the idea of multiple speech input sources. Previous research activities involved a single microphone with multiple, separate recognizers that all yielded improvements in accuracy. Distributed Listening uses multiple, parallel speech recognizers, with each recognizer having its own input source. Each recognizer is known as a listener and works in parallel with the other listeners. Each listener also serves as an interpreter. Once input is collected from the listeners, one machine, the master interpreter, processes all of the input (see figure 1).

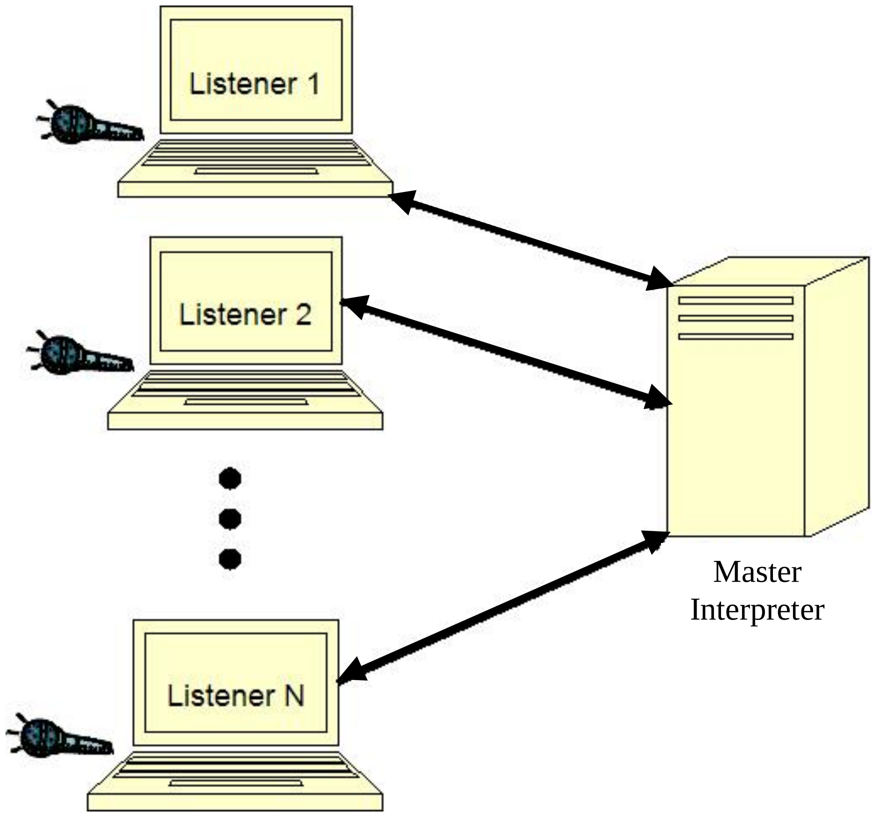


Figure 1 Distributed Listening Architecture

To process the spoken input, a phrase resolution algorithm is used. This approach is analogous to a crime scene with multiple witnesses (the listeners) and a detective (the interpreter) who pieces together the stories of the witnesses using his/her knowledge of crime scenes to form a hypothesis of the actual event. Each witness will have a portion of the story that is the same as the other witnesses. It is up to the detective to fill in the blanks. With Distributed Listening, the process is very similar. Each listener will have common recognition results and the individual interpreters will use a phrase resolution algorithm to propose phrases, with the master interpreter resolving conflicts.

All domains that utilize spoken language systems can benefit from Distributed Listening. The increase in speech recognition accuracy will result in more effective communication by improving the automatic transcription of spoken words.

1.2 Problem Description

Distributed Listening uses multiple perspectives collected from distributed speech recognizers working in parallel. Distributed Listening also uses a phrase resolution algorithm to reconcile the results of each recognizer. This approach addresses the followings issues found with current speech recognition systems:

1. Less than favorable recognition results in sub-optimal environments. This includes environments with considerable background noise and environments where the system has not been trained for a specific individual.
2. Bad recognition accuracy due to distorted input. Current systems use the same input source for each recognizer, so poor input will result in undesirable recognition rates.

1.3 Overview of Research Goals, Approaches and Contributions

Distributed Listening answers the question, “Is there a way to improve the accuracy of current speech recognition systems?” Researchers found that speech recognition accuracy rates fall to 0% when grammars reach a certain size (Gilbert 2003). While current speech recognition systems are robust and usable, in certain domains that is not enough. For example, students with hearing impairments have enough trouble keeping the pace of the other students in the classroom. The greater the accuracy of the speech recognition system, the better it is for these students.

To enhance speech recognition systems, Distributed Listening aims to simulate the way people hear. Humans use a psychological method called Dichotic Listening, where people listen to different voices in each ear, at the same time (Bruder 2004). It’s a natural extension to enable systems to hear in a manner similar to people.

The success of Distributed Listening will benefit not only the field of computer science, but society in general. A more accurate speech recognition system can be applied to all domains that utilize spoken input resulting in a more precise form of communication.

1.4 Organization

The following chapters will discuss this research task in detail. Chapter 2 examines the area of research that supports the development of Distributed Listening; Automatic Speech Recognition. Chapter 3 will present the research question and the approach that was used to support the hypothesis, including the system design and implementation details. Chapter 4 will explain the experiment that was performed, with a focus on the experiment design and settings. Chapter 5 will detail results, including a comprehensive data analysis and discussion. Chapter 6 provides the summary, along with contributions and directions for future research. The document ends with a reference list and appendices.

2 Literature Review

This chapter provides a description of Automatic Speech Recognition (ASR) systems, which is the core and concentration of this body of work. The scope of ASR systems is far-reaching, yet the research presented here is focused on ASR systems that use multiple speech recognizers. Thus, this chapter will also provide a review of such systems, which is the fundamental focus of this project.

2.1 Automatic Speech Recognition (ASR) Systems

Automatic speech recognition systems convert a speech signal into a sequence of words, usually based on the Hidden Markov Model (HMM) (Young 1990), in which words are constructed from a sequence of states (Baum 1972, Furui 2002), based on an extensive vocabulary of training data (Price 2010). The speech signal itself is based on phonemes (Young 1989). The sequence of words produced from the speech signal is returned as transcribed text of the speech input.

ASR systems do not include a component that determines the identity of the speaker, nor do ASR systems determine the meaning of the words that are spoken. Such systems are known as speaker recognition/verification and natural language processing, respectively, and are separate entities from ASR systems.

Among other things, these systems must overcome two obstacles (Price 2010):

1. Noise: The environment that surrounds the user directly impacts the accuracy of spoken language systems. In a controlled environment with minimal noise and

competing speech, it is common to have accurate recognition results. This accuracy degrades as background noise is introduced.

2. Dialect: The dialect of the user also directly impacts the accuracy of spoken language systems. Since such an large training set is needed for ASR systems, speakers who speech match the training set will notice better accuracy than those speakers who have a noticeably different dialect.

Even with the obstacles that effect ASR systems, there are four definitive advantages of such systems (Furui 1989):

1. Users do not need a specialized skill, like typing, to use speech recognition systems. For most people, speech is an inherent skill that comes natural and is cultivated from an early age.
2. Using speech is significantly faster than other forms of communication like typing or writing. A user can communicate with speech up to 10 times faster than writing on paper.
3. ASR systems allow the use of multiple modalities. Meaning, users can speak while doing other activities with their hands, legs, eyes, or ears.
4. The input methods of automatic speech recognition systems are economical. Specifically, microphones and telephones are very affordable.

Given the inherent nature of speech and the pervasive and ubiquitous qualities of computers, it is not surprising that ASR systems are heavily researched as practical applications to supplement everyday life. Within this domain, a focus has been maintained on multiple ASR systems that work together to improve accuracy. A review of such systems will be presented in the remainder of this chapter.

2.2 Enhanced Majority Rules

Barry (et. al. 1994) took three different Automatic Speech Recognition systems, along with an Enhanced Majority Rules (EMR) software algorithm and a fourth master system, to increase accuracy within the domain of aircraft cockpits, as shown in Figure 2.

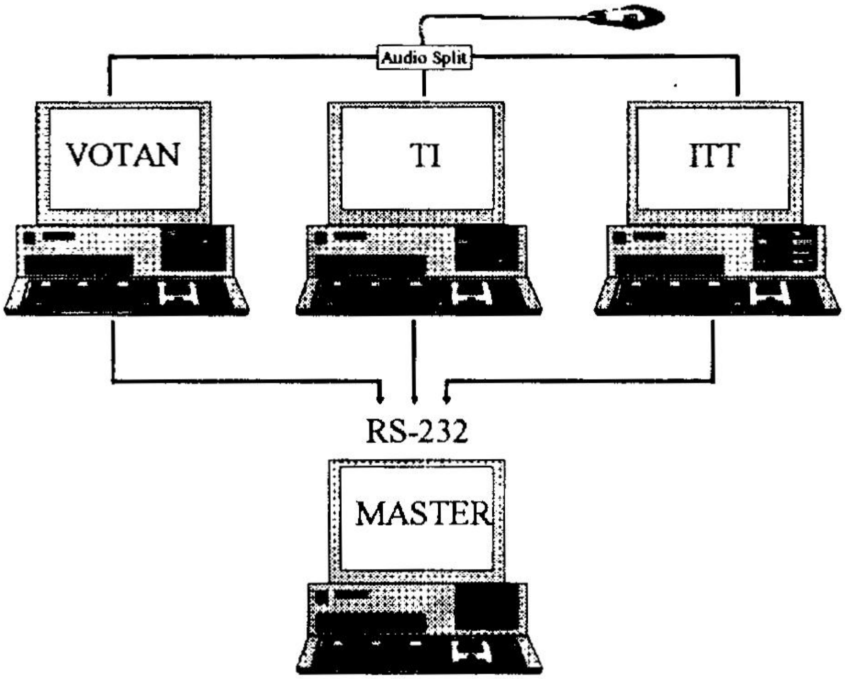
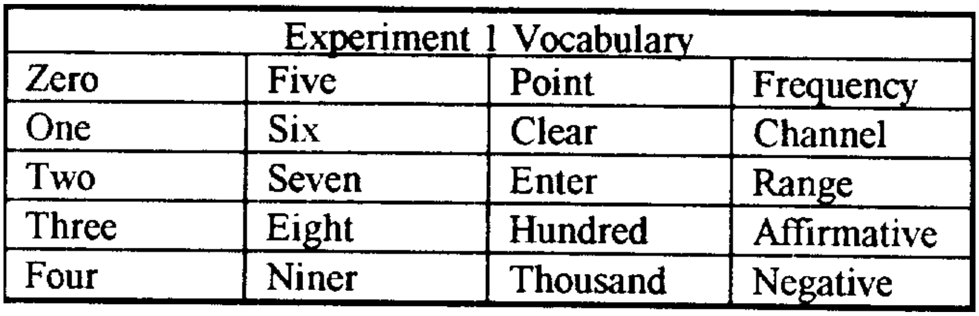


Figure 2 Barry (et. al. 1994) Hardware Configuration

Each of the three individual systems received the same input, performed speech recognition, and sent the result to the master system. The result included the recognized word along with a distance score, as well as a second choice word and its distance score. The distance score was the confidence of the system in choosing a particular word. The inconsistencies from the three individual systems were resolved using the EMR algorithm. The EMR algorithm resolved these inconsistencies by first looking for agreement from the individual systems for the recognized word. If there was no majority agreement, the EMR algorithm added the second word, with equal weight, to the collection of words and looked for a majority agreement. If there was still no agreement, the algorithm relied on the distance scores. At times, the individual systems would produce extra recognized words, or insertions. In those cases, the insertions were also added to the collection of words, and the algorithm proceeded in the same manner as described previously.

This architecture was used to complete two experiments. There was a twofold objective to the experiments; 1) to determine the recognition accuracies of the individual systems using an “easy” and a “hard” vocabulary and 2) to determine if the addition of the EMR algorithm would produce accuracy rates greater than the rates produced by the individual systems.

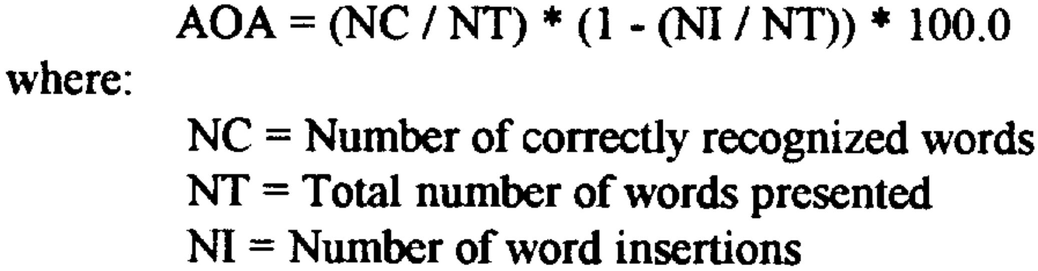
The first experiment relied on a simple, or easy, vocabulary (table 1) that consisted of 20 words common to the commands used by pilots in a cockpit, spoken by six male and six female pilots who were not experienced with ASR systems. Each pilot was randomly presented all 20 vocabulary words in 5 separate trials, resulting in 100 words spoken by each pilot. Before the start of the set of trials, which were run consecutively, each pilot trained the three individual systems to recognize his/her voice.



Experiment 1 Vocabulary			
Zero	Five	Point	Frequency
One	Six	Clear	Channel
Two	Seven	Enter	Range
Three	Eight	Hundred	Affirmative
Four	Niner	Thousand	Negative

Table 1 Barry (et. al. 1994) Experiment 1 Vocabulary

To determine if the resulting system produced better results than the individual systems, a measure of the mean Adjusted Overall Accuracy (AOA) was used along with a statistical gender by recognition system within subjects factorial design. The AOA consisted of a count of the number of correctly recognized words, the number of words presented, and the number of word insertions, as shown in Figure 3.



$$AOA = (NC / NT) * (1 - (NI / NT)) * 100.0$$

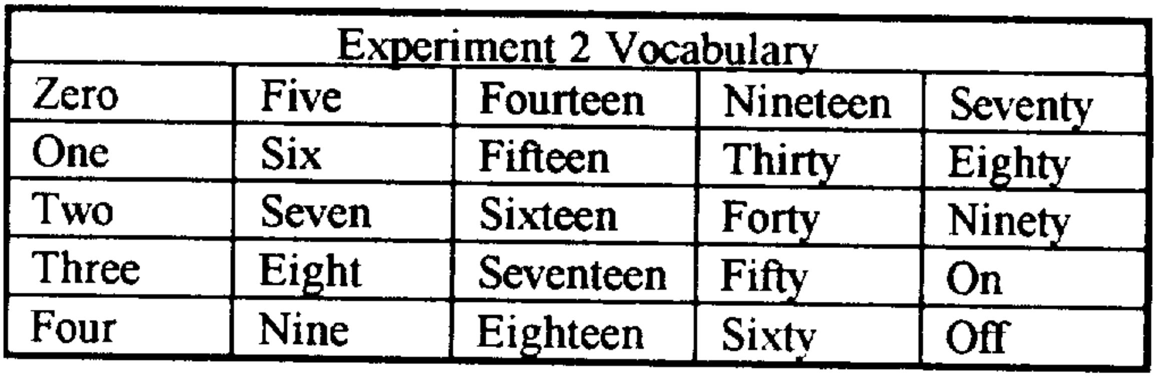
where:

NC = Number of correctly recognized words
NT = Total number of words presented
NI = Number of word insertions

Figure 3 Barry (et. al. 1994) Adjusted Overall Accuracy Measure

The resulting data analysis showed that the EMR algorithm produced statistically significant better recognition accuracy than two of the three individual systems, but failed to do the same when compared to the third system. Additional analysis of the data showed that there was no effect of gender differences.

The second experiment used the same architecture, algorithm and procedure, but differed in the robustness of the vocabulary that was used. The new vocabulary still reflected common words used in an aircraft cockpit environment, but was chosen based on the potential “confusability” of the words in the vocabulary. The 25 words, as shown in table 2, were chosen based on how closely a word sounded like another word in the vocabulary. Ten of the 12 pilots from the first experiment participated in the second experiment, along with 2 new participants. The resulting recognition accuracy and data analysis was the same as in the first experiment.



Experiment 2 Vocabulary				
Zero	Five	Fourteen	Nineteen	Seventy
One	Six	Fifteen	Thirty	Eighty
Two	Seven	Sixteen	Forty	Ninety
Three	Eight	Seventeen	Fifty	On
Four	Nine	Eighteen	Sixty	Off

Table 2 Barry (et. al. 1994) Experiment 2 Vocabulary

Overall, the use of the EMR algorithm with three individual recognition systems produced better recognition accuracy than the individual systems. While an improvement was made, the architecture can suffer from distorted input. Since each system receives the same input, if the input signal is not good, then all of the individual systems will receive that bad input signal.

2.3 Virtual Intelligent Codriver

The Virtual Intelligent Codriver (VICO) project also used multiple ASR systems in parallel within automobiles to increase the accuracy of hands-free communication inside a car (Brutti et. al. 2004, Cristoforetti et. al. 2003). The specific aim of the VICO project was to assist drivers in accessing tourist information and driving assistance while inside an automobile.

The VICO structure had multiple ASR systems with each receiving the same input and having its own specialized language model (see figure 4). There were several distinct language models as an alternative approach to one comprehensive language model and large vocabulary. The input that each system received was first processed and included an optional background noise reduction procedure. The multiple systems then performed independent speech recognition on the input. The resulting interpretations from each ASR system were then passed to a module that selected the final output. This output was chosen by comparing the interpretations to each other using confidence scores. The interpretation with the maximum likelihood was selected.

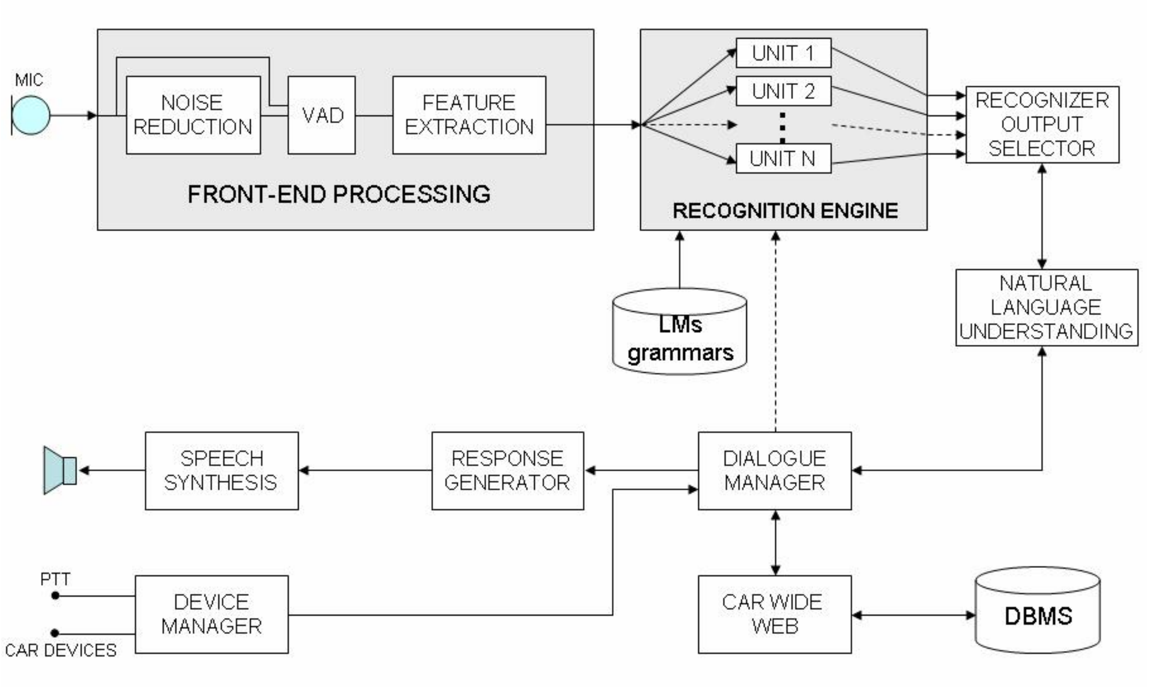
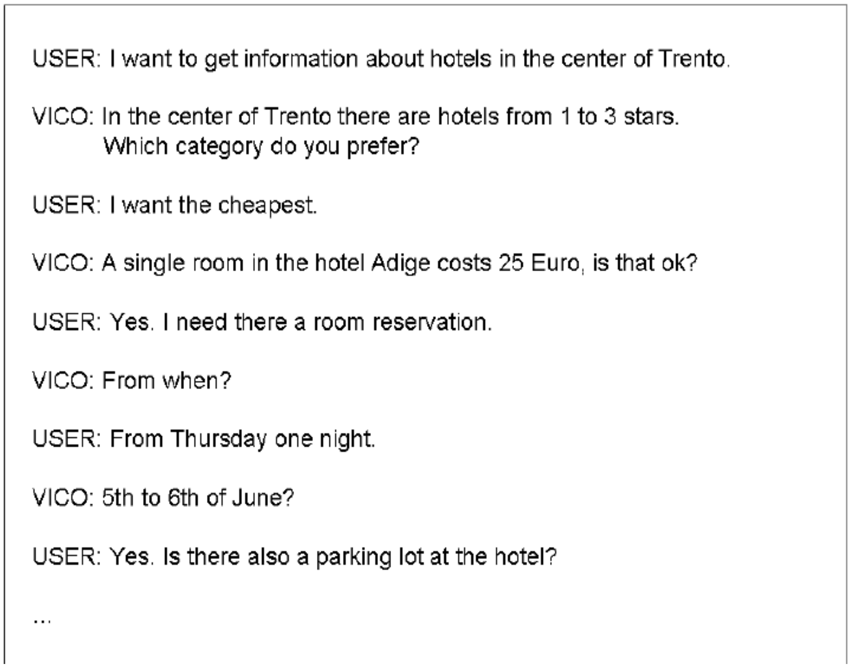


Figure 4 Brutti (et. al. 2004) Distributed Listening Architecture

The final output was then passed to a module that performed natural language understanding in that the module took the recognized interpretation and produced a semantically correct representation that could be passed to a dialogue manager module and ultimately a response generator module. To produce the semantically correct representation, the natural language module parsed the recognized interpretation and identified select elements within the phrase. This module did not check for incomplete phrases, grammatically incorrect phrases, or other inconsistencies in the recognition. Rather, meaningful words of the phrase were identified and subsequently passed to the dialogue manager module.

The dialogue manager module was then responsible for selecting responses to the spontaneous requests of the driver, based on the semantic information from the natural language understanding module. The selected response was then passed to the response generator with instructions on how to generate the response that would in turn undergo speech synthesis before being presented to the driver. An example dialogue of the VICO system is shown in figure 5.



USER: I want to get information about hotels in the center of Trento.

VICO: In the center of Trento there are hotels from 1 to 3 stars.
Which category do you prefer?

USER: I want the cheapest.

VICO: A single room in the hotel Adige costs 25 Euro, is that ok?

USER: Yes. I need there a room reservation.

VICO: From when?

USER: From Thursday one night.

VICO: 5th to 6th of June?

USER: Yes. Is there also a parking lot at the hotel?

...

Figure 5 Brutti (et. al. 2004) Sample Dialogue

A number of experiments were performed within an automobile using an architecture that consisted of 5 ASR systems. The spoken data came from 8 female and 8 male speakers, with a combined total of 1612 utterances that translated to 9150 word occurrences and 918 vocabulary words. The research team was able to perform synchronous experiments by using a “close-talk head-mounted” microphone, as well as a “far-microphone” that was located on the ceiling of the automobile. The experiments showed a noticeable improvement over a structure with individual ASR systems. Specifically, the close-talk microphone experiment resulted in a 3.7% increase in recognition rates compared to using individual ASR systems. Likewise, the far-microphone experiment resulted in a 2.4% increase in recognition rates.

The researchers found that while using the maximum likelihood to select the final output from the multiple ASR system did show a decrease in error rates, that method represents the simplest choice. The project team recognized this fact and noted other ASR systems reconciliation methods for future research that included confidence measures and word graph hypotheses.

Although the VICO project did indeed show an improvement in speech recognition accuracy when using multiple ASR systems with specialized recognition units, there are two shortcomings. First, if the input signal is distorted, then each recognizer will receive bad input. Second, if each recognizer contains a piece of the optimal interpretation, then this architecture falls short. In order to address this problem, a post-processing combination algorithm is required.

2.4 Recognized Output Voting Error Reduction

The Recognized Output Voting Error Reduction (ROVER) system is a composite of multiple ASR systems that uses a voting process to reconcile differences in the individual ASR system outputs (Fiscus 1997). The ROVER architecture, as shown in figure 6, consists of multiple recognition engines, an alignment module and a voting module. The multiple interpretations from the recognition engines are passed to the alignment module. The alignment module iteratively builds a composite linear topology Word Transition Network (WTN). The ROVER system subjectively selects one of the ASR system outputs to act as the base of the WTN. The WTN depicts the order and transition from one word to another in a given ASR system output. Each ASR system output is added to the WTN until a composite network is built that shows the word similarity, by position, between the individual ASR systems outputs. Once aligned, the voting module is called. The ROVER project investigated three different voting schemes. First, votes were tallied based on frequency of occurrence. Second, votes were tallied by frequency of occurrence along with the average word confidence score. Lastly, votes were tallied by frequency of occurrence along with the maximum word confidence score. No matter the voting scheme used, the voting module scores each word within the composite WTN vertically and the words with the highest scores are chosen, at any given position within the network. If there were ties between words, the ROVER system arbitrarily selects a winner.

Tests were performed for each of the voting schemes, using 3 ASR systems for each experiment.

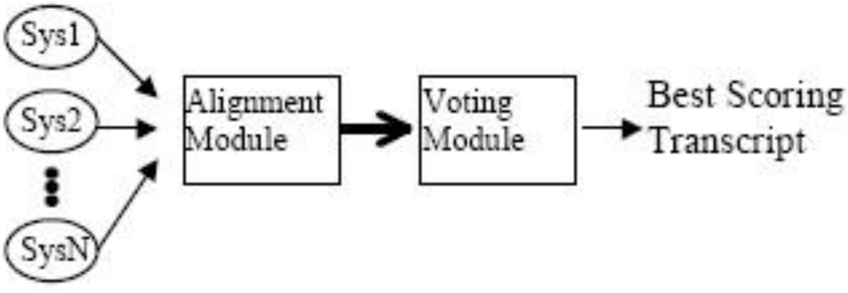


Figure 6 ROVER Architecture

The first voting scheme performed the poorest, relative to the other voting schemes, but still achieved a decrease in the error rate compared to an individual system. This was the only voting scheme that resulted in ties within the WTN. Out of approximately 30,000 words, 5,320 resulted in a tie.

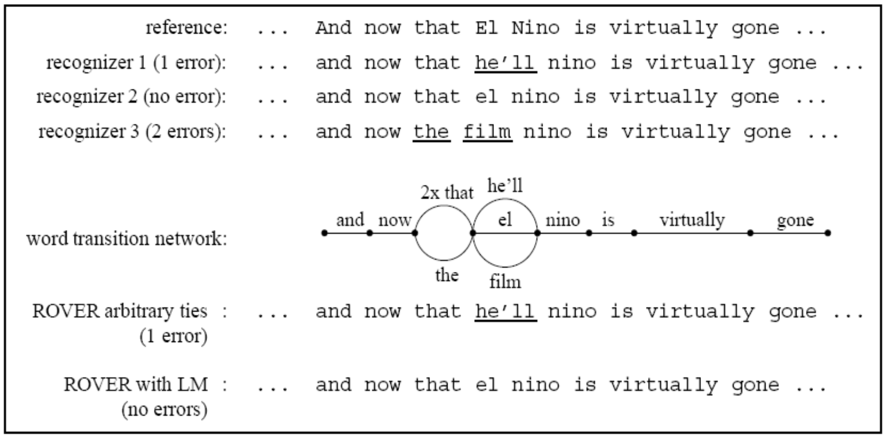
The second voting scheme, which used average confidence scores, showed even more of an improvement over an individual system. The fact that this voting scheme did not have to resolve any word score ties was considered a significant improvement alone.

The third voting scheme showed the best improvement over an individual system in addition to not having to resolve any word score ties.

On average, this composite ASR system produced a lower error rate than any of the individual systems, but suffers from order of combination into the WTN and ties within the voting module. To overcome these shortcomings, a future research direction of this project was to investigate other voting methods that take advantage of different knowledge sources, including decision trees and artificial neural networks. Additionally, an alternative method for aligning the results of the individual ASR systems into the WTN by way of phonologic mediation is of interest to this research team.

2.5 Modified ROVER

To solve the problem that resulted from the order of combination and ties of the original ROVER system, Schwenk proposed a modified ROVER system that used a dynamic programming algorithm built on language models (Schwenk and Gauvain 2000). To accomplish the task, the modified ROVER system took advantage of the composite WTN of the original ROVER system, where the most likely word is chosen from each branch of the WTN, but changed the order of the systems when producing the composite WTN. Moreover, the modified system also added a normalization procedure before combining the systems. The normalization mapped alternate spellings of words and abbreviated forms of words to a common form. In the case where the WTN produced more than one result, the Modified ROVER system used a language model to select the most likely word sequence based on contextual information. Meaning, when a position in the WTN resulted in a tie, all of the variations were kept, and the resulting word sequences were analyzed according to the perplexity of the sequence according to the language model. The word sequence that minimized perplexity was the sequence that was chosen. To ensure that the dynamic programming algorithm didn’t automatically prefer short word sequences, a penalty was applied for null arcs of the WTN. An example WTN is shown in figure 7, where incorrect words are underlined.



reference: ... And now that El Nino is virtually gone ...

recognizer 1 (1 error): ... and now that he'll nino is virtually gone ...

recognizer 2 (no error): ... and now that el nino is virtually gone ...

recognizer 3 (2 errors): ... and now the film nino is virtually gone ...

word transition network:

ROVER arbitrary ties (1 error) : ... and now that he'll nino is virtually gone ...

ROVER with LM (no errors) : ... and now that el nino is virtually gone ...

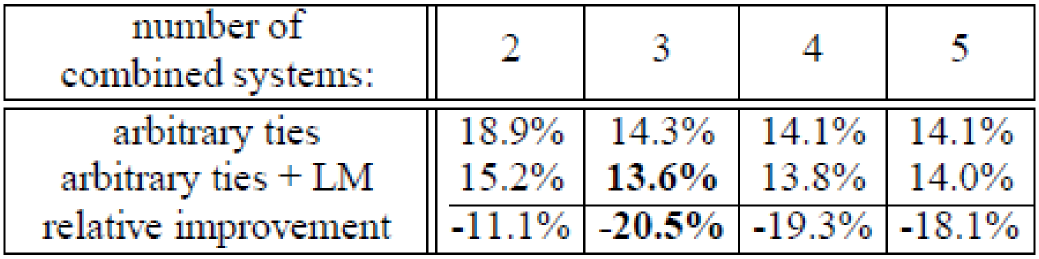
Figure 7 Modified and Standard ROVER Example WTN

The modified ROVER system was tested using two different speech recognition corpora using a varying number of ASR systems, from 2 up to 9 combined systems. Table 3 displays the word error rate, sentence error rate, and perplexity of the improved ROVER system and compares the same rates with the original ROVER for the first corpus. When combining between 2 and 7 systems, an increase in accuracy is seen for the applicable metrics. Notice, however, that when combining 8 or 9 systems, the word error and sentence error rates are worse for the improved ROVER system, yet the perplexity is better.

number of combined systems:	2	3	4	5	6	7	8	9
arbitrary ties:								
word error:	13.8%	11.6%	10.7%	10.1%	10.1%	10.0%	10.2%	10.4%
sentence error:	81.0%	76.3%	74.3%	73.0%	73.8%	73.4%	73.4%	74.6%
perplexity:	183.8	171.6	166.1	164.2	161.7	160.2	159.3	159.6
using LM to break ties:								
word error:	12.5%	11.1%	10.3%	10.1%	10.1%	10.0%	10.3%	10.5%
sentence error:	79.9%	75.4%	73.3%	72.6%	73.0%	72.9%	74.2%	74.7%
perplexity:	137.2	145.8	146.5	151.2	149.6	150.8	150.0	151.1

Table 3 Modified ROVER Word Error Rates, Sentence Error Rates, and Perplexity

The second experiment showed similar results. Between 2 and 5 systems were combined on a second corpus to produce a decrease in word error rates over the original ROVER system, as shown in table 4. The relative improvement as displayed in the table is relative to the best single recognizer that achieved a 17.1% word error recognition rate.



number of combined systems:	2	3	4	5
arbitrary ties	18.9%	14.3%	14.1%	14.1%
arbitrary ties + LM	15.2%	**13.6%**	13.8%	14.0%
relative improvement	-11.1%	**-20.5%**	-19.3%	-18.1%

Table 4 Modified ROVER Word Error Rates

The experimental data analysis showed that using language model information is advantageous, but performance can degrade when too many systems are combined, especially those systems with the highest word error rates. Overall, the improved ROVER system resulted in a 5% relative word error reduction over the original ROVER system.

2.6 Post-Labeling Integration

Paul Duchnowski researched a method termed Post-Labeling Integration (figure 8) that used multiple recognizers, called sub-recognizers, that worked independently before being integrated together to produce the final output decision (Duchnowski 1993).

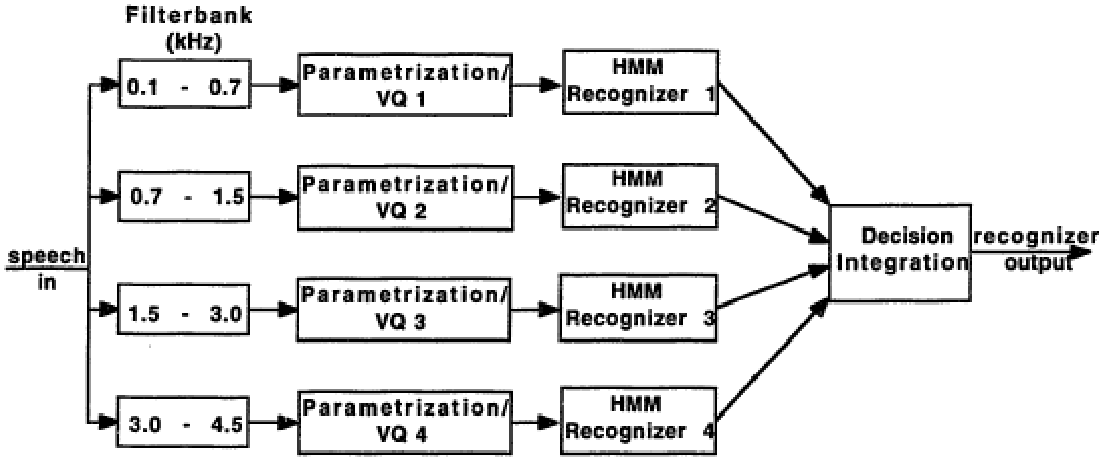
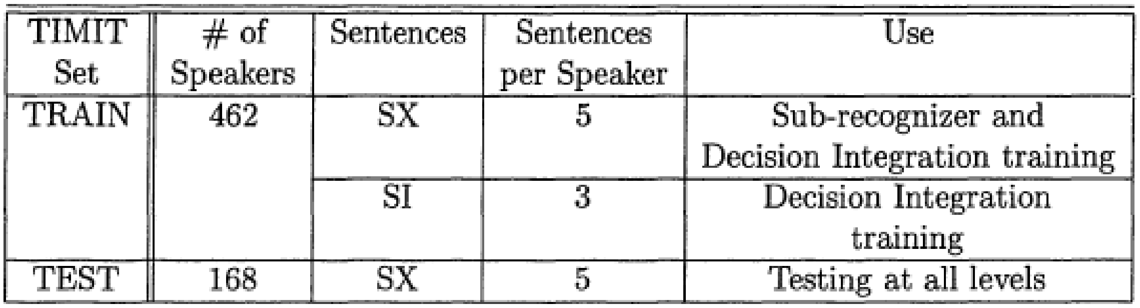


Figure 8 Duchnowski Block Diagram of the Proposed Recognizer

To accomplish the task, Duchnowski used a single speech signal that was filtered into frequency bands and parameterized before being fed into 4 sub-recognizers. The parameterization was consistent across each independent channel and was used to extract "the most salient, information bearing features" of the speech signal. The individual outputs of the sub-recognizers were processed using a combination rule that merged the outputs to produce the recognized phones, or the smallest identifiable elements of speech. To integrate the outputs of the sub-recognizers, the speech signals were aligned by identifying the timing between successive phones and using probabilistic functions and a bigram language model to select the final sequence of phones.

The subsequent experiments of this architecture were performed using the National Institute of Standards and Technology (NIST) version of the TIMIT database, which is a "readily available, large, multi-speaker database that has been phonetically transcribed" and contains 630 speakers, with twice as many males as females. Each of the 630 speakers supplied 10 sentences and they collectively represented 8 regional dialects of

American English. Of the 6300 sentences available, a portion of those sentences was not used because they were the same for all of the speakers and would have put a bias towards certain phones. Table 5 shows the portions of the TIMIT database that were used for the experiments.



TIMIT Set	# of Speakers	Sentences	Sentences per Speaker	Use
TRAIN	462	SX	5	Sub-recognizer and Decision Integration training
		SI	3	Decision Integration training
TEST	168	SX	5	Testing at all levels

Table 5 Duchnowski (1993) Breakdown of TIMIT Database

The resulting data analysis showed that 54% of the phones were correctly recognized. Duchnowski stated that the resulting recognition rate was similar to comparable systems, but not a major improvement.

It is worthwhile to mention that this research project is distinct from some in this chapter in that the focus was on the phone-level, as opposed to the word-level.

2.7 Multiple Japanese LVCSR Models

This research project combined multiple Japanese LVCSR models and was also motivated by the ROVER project (Kodama et. al. 2001). The research team hypothesized that if a simple voting scheme such as the one used with ROVER could produce word error reduction then it is possible to further improve results by “simply exploiting more than one speech recognizers’ output”. To accomplish the task, the confidence of agreement between the outputs of the systems was evaluated, with the systems having different acoustic models.

There were two different acoustic models that were evaluated, the first being a phoneme-based HMM and the second being a syllable-based HMM, both based on Gaussian mixture HMM. The phoneme-based acoustic model was gender-dependent (male) and included 43 Japanese phonemes. The syllable-based HMM was also gender-dependent (male) and included 114 Japanese syllables.

To evaluate the confidence of the combined Japanese LVCSR models, a metric that determined the recall/precision rate of estimating correctly recognized words was used, according to the formulas shown in figure 9. The formulas rely on an agreed word list that is a collection of those words that were aligned through dynamic programming and had identical lexical form.

$$\begin{aligned} Recall &= \frac{\text{\# of correct words in the agreed word list}}{\text{\# of words in the reference sentence}} \\ Precision &= \frac{\text{\# of correct words in the agreed word list}}{\text{\# of words in the agreed word list}} \end{aligned}$$

Figure 9 Multiple Japanese LVCSR Models Recall and Precision Formulas

To train and test the system, two datasets were used. One dataset was composed of 100 newspaper sentence utterances spoken by 10 males and consisting of 1,565 words. This dataset was considered relatively easy for speech recognizers. The second harder dataset was comprised of 175 broadcast news speech utterances and consisted of 6,813 words spoken by 10 male speakers, of which 8 were announcers and the other 2 were reporters. Half of the data were used for training and the other half were used for testing.

The resulting data analysis showed that relying on the agreement between outputs of multiple LVCSR models, along with different acoustic models, performs well. Most notable is that the composite system was found to have quite high precision with less than 10% loss of recall from a single LVCSR model. The precision is over 99% accurate when the word recognition rate of a single LVCSR model is approximately 90% and over 93% accurate when the baseline word recognition rate is below 60%.

2.8 Segmental Minimum Bayes-Risk Recognition

Goel et. al. (2000) were also motivated by the ROVER project and attributed the success of ROVER and subsequent voting schemes on the Minimum Bayes-Risk (MBR) framework. To further extend the research in this area, a segmental MBR Recognition procedure was developed with a derivative of the voting procedure to create N-best ROVER. Another extension was made that produced the Extended (e)-ROVER.

N-best ROVER first constructed a WTN based on the N-best outputs of N systems. A posterior probability equation based on the WTN and a distribution was then used for each correspondence set of words. The word with the highest posterior probability from each set was selected and concatenated to produce the final output.

The research team then improved segmentation that resulted in e-ROVER. This allowed two or more consecutive words in each correspondence set and was achieved by joining the two consecutive sets. The joined, expanded set replaced the two sets while maintaining the paths from the original sets (figure 10) within the WTN. The addition of the segmentation was further derived from a “pinching” procedure that ignored those word sets with a posterior probability above the pinching threshold. An example e-ROVER WTN is shown in figure 11.

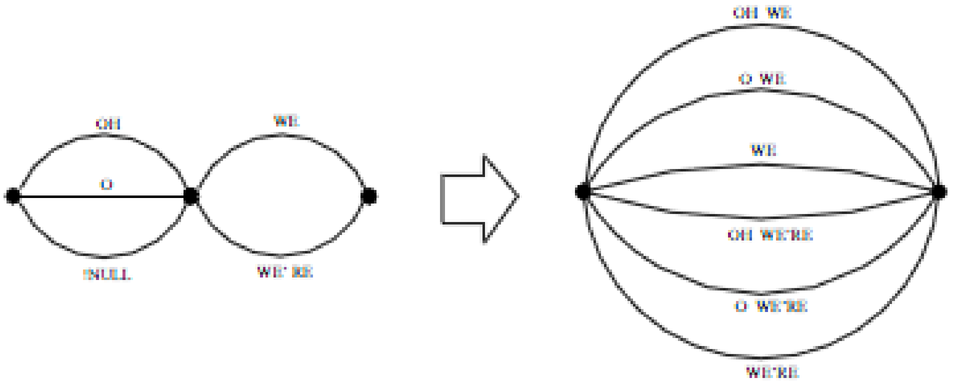


Figure 10 e-ROVER Joining Two Correspondence Sets

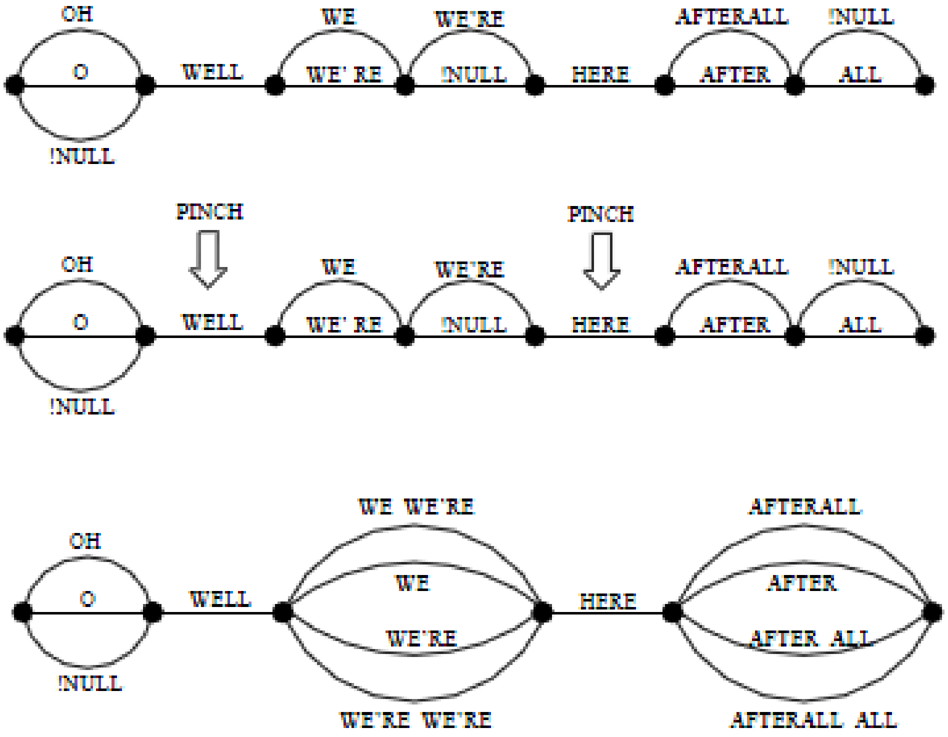


Figure 11 e-ROVER WTN Construction

The experiments that were executed were based on a multi-lingual acoustic modeling task that included combined Czech recognition outputs from three systems. The three individual systems had error rates of 29.42%, 35.24%, and 29.22% and are used as baseline metrics for comparison of N-best ROVER and e-ROVER.

The best individual system achieved an error rate of 29.22% and N-best ROVER was able to improve that baseline by 3.28%. An additional improvement of .56% was achieved from e-ROVER to produce a 3.84% improvement over the baseline. A comparative analysis of N-best ROVER and e-ROVER is shown in figure 12.

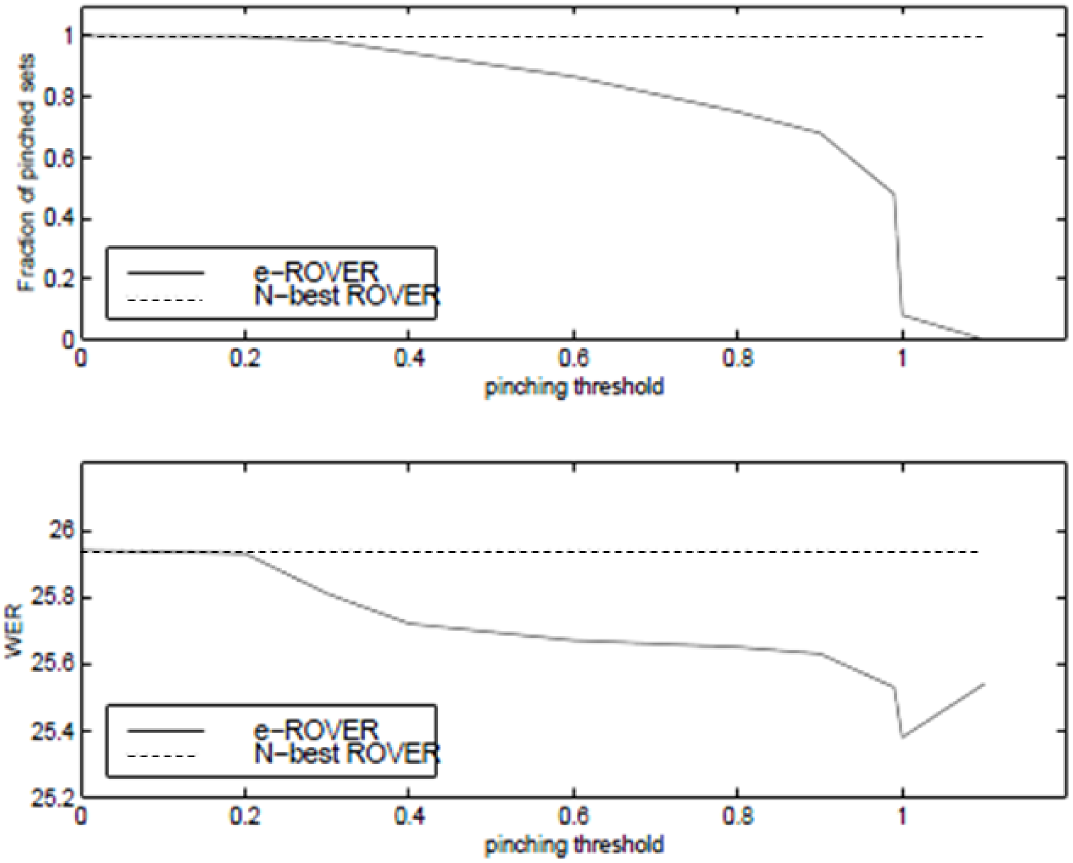


Figure 12 N-Best ROVER and e-ROVER Comparative Analysis

Overall, the resulting tests provided a small, yet significant, improvement. The primary shortcoming of e-ROVER is the method used for pinching. The researchers believe pinching by overall Bayes-Risk and not just posterior probabilities will result in further improvements.

2.9 Posterior Probability Decoding with Confidence Estimation

The ROVER project motivated another research team that evaluated posterior probabilities with confidence score estimation using multiple recognizers (Evermann and Woodland 2000). This project relied on confusion networks that provided a representation of the most likely word along with the associated word posterior probability. The posterior probabilities were used to create the confidence scores. The combination of confidence scores based on posterior probabilities, along with the confusion networks, were the features of the system.

The word-level posterior probabilities were “derived from the acoustic and language model (LM) likelihoods of the word sequences hypothesised by a Viterbi decoder” and represent the competing words and scores. The confusion network is a linear graph composed of a word lattice that has been clustered and transformed using a Viterbi decoder. The resulting confusion network was used with the dynamic programming alignment procedure of the original ROVER project and contains word posteriors (figure 13).

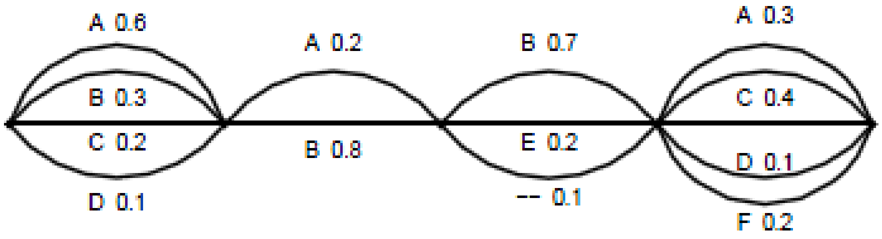


Figure 13 Posterior Probability Example Confusion Network

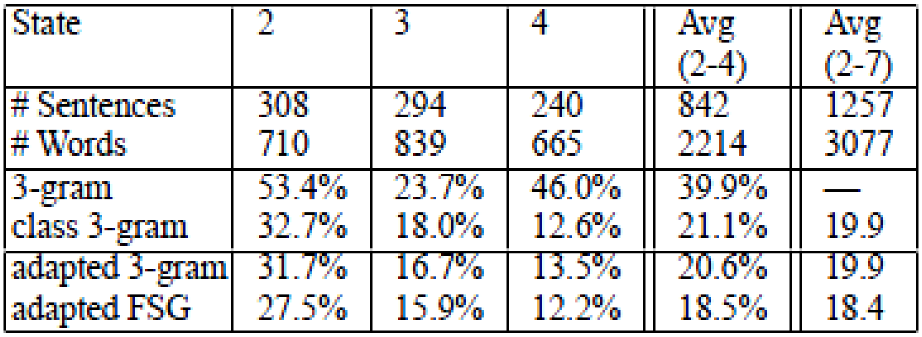
To test the composite system, the CU-HTK system that was used in the March 2000 Hub5 Conversational Telephone Speech evaluation was applied and the acoustic models were trained using the Switchboard and CallHome corpora. The two training criteria used were the maximum likelihood estimation criterion and the maximum mutual information estimation criterion.

The resulting data analysis showed that the use of confidence scores in this manner produced a decrease in error rates compared to the original ROVER system that used a simple voting scheme and the addition of the confusion network gave another small increase in accuracy rates.

2.10 Adaptive Language Models

Solsona et. al. (2002) investigated multiple recognizers within a travel reservation system using state dependent Finite State Grammars (FSG) and context-independent n-grams. The researchers used two recognizers working in parallel with one recognizer based on n-gram statistical language models and the other on finite state grammars. Once the results from each recognizer were combined, an acoustic confidence measure was used to reconcile to one result. The confidence measure was a phone-based likelihood ratio. To select the final result, the sentence with the highest phone score was chosen.

The experiment that was executed was based on a trigram model and used a training corpus that consisted of 19,283 sentences, or 58,595 words, from the June 2000 DARPA Communicator data collection and database. The results from the experiment are shown in table 6. The last column in the table gives the average of the results for states 2 through 7, although states 5 through 7 are not listed in the table. This is due to the fact that states 2 through 4 contained the most number of training sentences and were the focus. The averages of the other states were given for completeness.



State	2	3	4	Avg (2-4)	Avg (2-7)
# Sentences	308	294	240	842	1257
# Words	710	839	665	2214	3077
3-gram	53.4%	23.7%	46.0%	39.9%	—
class 3-gram	32.7%	18.0%	12.6%	21.1%	19.9
adapted 3-gram	31.7%	16.7%	13.5%	20.6%	19.9
adapted FSG	27.5%	15.9%	12.2%	18.5%	18.4

Table 6 Adaptive Language Models Word Error Rates

The 3-gram language model produced an average word error rate of 39.9% (for states 2-4) and was considered as poor performance. The researchers attributed the poor performance to sparse training data. To overcome the sparseness, 15 semantic classes were introduced (class 3-gram) and the error rate was reduced to 21.1%. This shows that a class-based language model is effective for generalizing the language model. A further reduction in error rates was achieved by interpolating the state-independent class 3-gram with a state-dependent trigram (adapted 3-gram). The interpolation produced an average of 20.6%, which is a relative reduction of 2.5% over the class 3-gram model. Lastly, the results from combining the general class trigram model and the state-dependent FSG (adapted FSG) showed a decrease from 21.1% to 18.5%.

Overall, combining results from state-dependent FSGs and context-independent n-gram language models routinely outperforms the baseline and can achieve up to 12% relative reduction in error rates.

While previous research activities using multiple ASR systems have resulted in improvements, they consistently used the same input source or relied on the alignment of recognition results to achieve an optimal result. The research presented here uses neither of those criteria in resolving the results from the multiple ASR systems, as discussed in the next chapter.

3 System Design

Speech recognition is capable of 95% accuracy under optimal conditions, but optimal conditions are not always possible, which is why there isn't mainstream use of speech technology. Distributed Listening attempts to improve the recognition accuracy achieved independent of the environment conditions.

3.1 Problem Statement

Distributed Listening uses multiple perspectives collected from distributed speech recognizers. Distributed Listening also uses a phrase resolution algorithm to reconcile the results of each recognizer. This approach addresses the followings issues found with current speech recognition systems:

1. Less than favorable recognition results in sub-optimal environments. This includes environments with considerable background noise and environments where the system has not been trained for a specific individual.
2. Poor recognition accuracy due to distorted input. Current systems use the same input source for each recognizer, so poor input will result in undesirable recognition rates.

Distributed Listening was developed in response to the aforementioned problems with a hypothesis that **Distributed Listening will perform at worst, as good as the best**

individual recognizer. The remainder of this chapter will describe the physical aspects, both hardware and software, of the system and how those aspects work together as a viable solution to the issues that have been presented.

3.2 Design Principles

The purpose of Distributed Listening is to provide a more accurate speech recognition system. The main features of the system include:

- Multiple, yet distinct, input sources
- An accurate reconciler for the multiple recognition results

These necessary features dictate the architecture of the Distributed Listening system, as described in the following section.

3.3 System Features

Distributed Listening is composed of two significant parts. The first being listeners and the second being interpreters, which rely on the Evidence-Based Phrase Resolution Algorithm (PRA). In addition, Distributed Listening utilizes a speech corpus and a database. Each part is equally important and will be described in detail next.

3.3.1 Listeners

Distributed Listening uses multiple speech recognizers to process the spoken input. Each recognizer is called a listener and is equipped with its own input source. Each listener is a separate, physical computing device with its own memory, processor, and disk space and works independently of the other listeners. Each listener collects input in the form of speech and performs interpretation that is ultimately used by the master interpreter to produce one result that serves as the most likely spoken phrase.

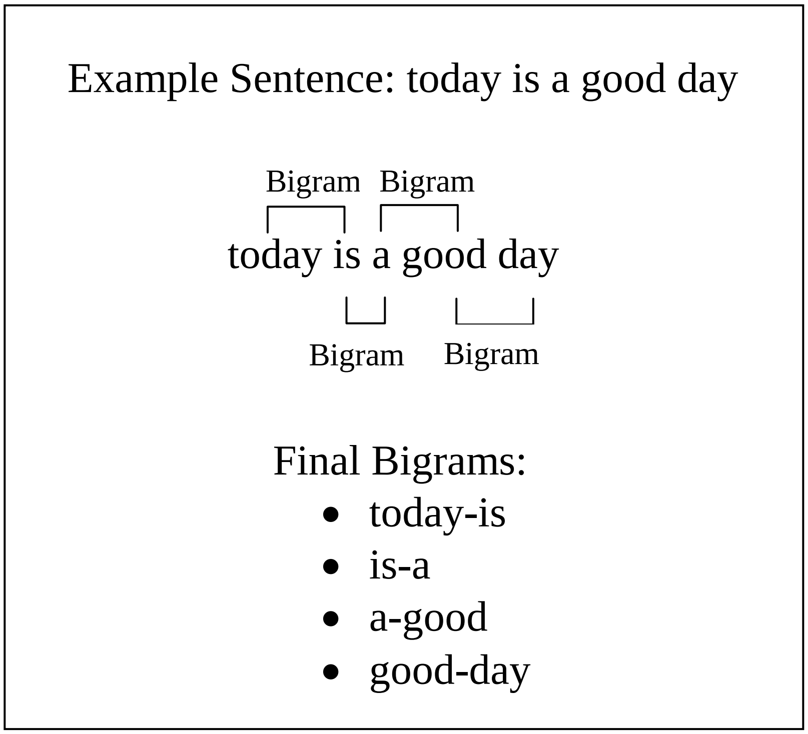
3.3.2 Interpreters

Once input is collected by the listeners, each listener performs interpretation on the input using a resolution algorithm to produce a final set of candidate phrases. Each set of candidate phrases is then processed by the master interpreter to reconcile the variations in the results of the listeners. As a separate entity, the interpretation is not very powerful. Therefore, the interpretation works together with, and relies on, the resolution algorithm. The algorithm is dependent on one additional system feature; a corpus. The corpus will be defined next, followed by a detailed description of the resolution algorithm.

3.3.2.1 Corpus

The corpus is a validation entity and is subject to the domain that utilizes Distributed Listening. It can be increased or decreased as necessary, based on the characteristics of the domain. In essence, the corpus is a database table of known utterances relative to a particular area.

For this system, the utterances are recorded in bigram form, where a bigram is a 2-word pair as illustrated in figure 14. The corpus maintains a unique list of bigrams that people commonly speak, as well as the number of times the bigram is spoken (frequency), relative to the other bigrams in the corpus. This list is maintained as a database table, as defined in figure 15. The bigrams, or 2-word pairs, of the corpus resemble a bigram approach to language modeling and are used with the resolution algorithm to produce the most likely spoken phrase. The actual corpus used in the experimentation phase will be discussed in detail in chapter 4.



Example Sentence: today is a good day

Final Bigrams:

- today-is
- is-a
- a-good
- good-day

Figure 14 Example Bigrams

bigrams (word1, word2, frequency)

Figure 15 Bigrams ER Diagram

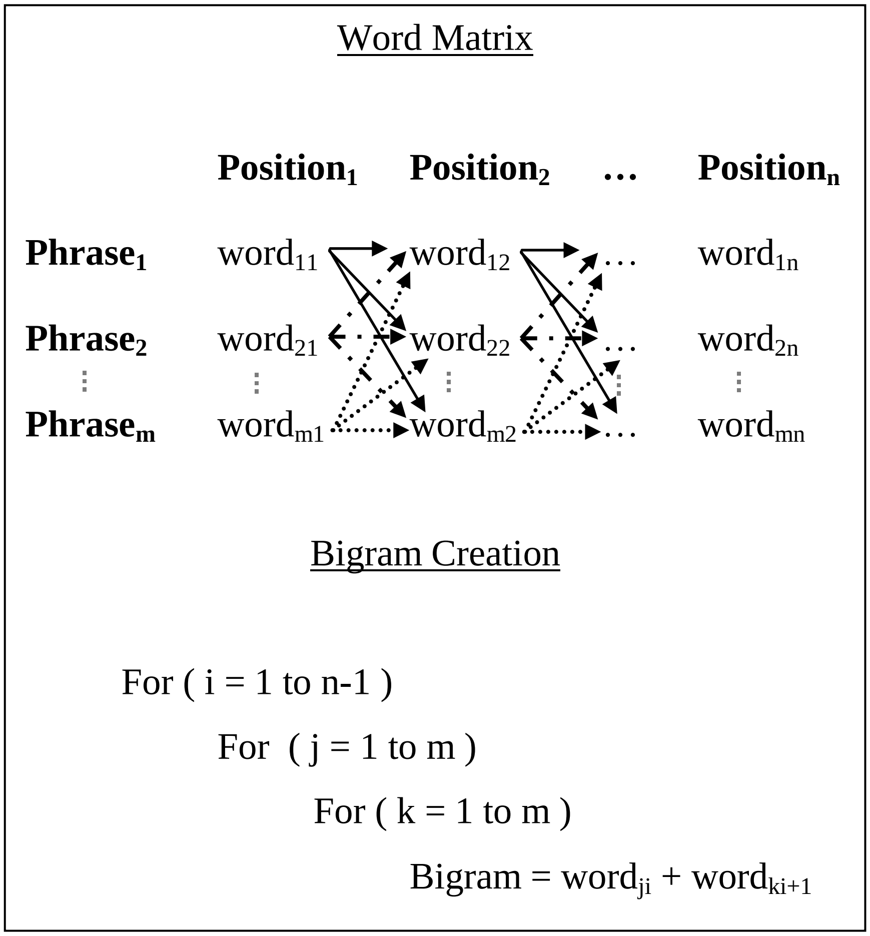
The corpus works in combination with the PRA. The algorithm, and the supporting role of the corpus, will be described next.

3.3.2.2 Evidence-Based Phrase Resolution Algorithm (PRA)

To resolve multiple recognitions from the listeners, the Evidence-Based Phrase Resolution Algorithm (PRA) is used. Recall from chapter 1 that Distributed Listening is analogous to a crime scene investigation, where witnesses correspond to listeners. The stories of the witnesses are the beginning of the evidence collection. Likewise, the algorithm uses as evidence the actual bigrams from the recognitions of the listeners. Each listener's recognition result is broken down into its individual bigrams and the collection of original bigrams from all of the listeners is retained as evidence.

Next, the recognitions go through an iterative process where new bigrams are created by combining a word from one recognition with a word from a separate recognition, based on the position of the word within the recognition. The recognitions are not aligned. Rather, the first word from each listener is combined with the second word from each listener, followed by combining the second word from each listener with the third word from each listener and so on until the end of each individual recognition phrase. This procedure is best depicted as a 2-dimensional word matrix with resulting iterations shown as nested loops (figure 16). For recognitions that are shorter than others, when the end of that recognition is reached, it is no longer used to create bigrams. As a new bigram is created, it is validated against a corpus and the evidence. A bigram is considered valid if it is found within either the evidence or the corpus, otherwise the bigram is considered invalid. If the newly created bigram is deemed valid, it is added to a temporary table that maintains a local count of the frequency of that particular bigram. The valid bigrams that are created are concatenated together to form candidate strings that represent the most likely spoken phrase.

It is important to note that the initial bigrams from the recognitions from the listeners are **not** validated against the corpus. This is so as not to delete the "evidence" presented by the listeners. Should the recognitions from the listeners contain a bigram that is not in the corpus, that bigram should not be discarded as the recognitions from the listeners are the base, or the evidence, of the candidate phrases.



Word Matrix

	Position_1	Position_2	…	Position_n
Phrase_1	word_{11}	word_{12}	…	word_{1n}
Phrase_2	word_{21}	word_{22}	…	word_{2n}
⋮	⋮	⋮	⋮	⋮
Phrase_m	word_{m1}	word_{m2}	…	word_{mn}

Bigram Creation

```
For ( i = 1 to n-1 )
    For ( j = 1 to m )
        For ( k = 1 to m )
            Bigram = $\text{word}_{ji} + \text{word}_{ki+1}$
```

Figure 16 Evidence-Based Phrase Resolution Bigram Creation

Once the iterative process is complete and the candidate strings have been created, each listener computes the local total frequency count of its candidate phrases. Specifically, for each bigram within a candidate string, a running total of the frequency of those bigrams is calculated. The candidate phrase or set of candidate phrases with the highest total is sent to the master interpreter. This iterative process happens in parallel with each listener.

Next, the master interpreter receives the full set of candidate phrases from the listeners. The phrase within the set of phrases with the greatest frequency sum is chosen as the most likely spoken phrase. If there is a tie between candidate phrases, the phrase that is equal to the recognition result of a listener is selected. This is due to the likelihood that one of the distributed listeners heard the actual spoken phrase correctly. This selection process favors results that have evidence from the listeners, as opposed to the results that were created through concatenations. If there is still a tie between candidate phrases, those candidate phrases are validated against the corpus. If a phrase contains bigrams that are not found within the corpus, it is discarded and the remaining valid phrase is chosen as the most likely spoken phrase. If by chance there is still a tie, the total bigram frequency count for each phrase is re-calculated according to the corpus, instead of the local bigram frequency total, and the phrase with the greatest frequency sum is chosen as the most likely spoken phrase. Any additional ties are broken by arbitrarily choosing a phrase as the most likely spoken phrase.

An example of the PRA will help put this into perspective. Assume there are three listeners and that the actual spoken phrase of the person is:

- These kids don’t deserve to be educated they say

The actual spoken phrases as “heard” by the three listeners are as follows:

Listener 1: These kids don’t deserve to be educated basic

Listener 2: These kids don’t deserve to be educated they say

Listener 3: These kids don’t deserve to be educated these to

Also assume there is a corpus called DL Utterances that contains 400,000 unique bigrams and corresponding frequency counts and is comprised of various newspaper text articles. The portion of the DL Utterances corpus that contains the bigrams from this example is shown in table 7.

Bigram		Frequency Count	Bigram		Frequency Count
Word 1	Word 2		Word 1	Word 2	
be	educated	5	these	to	4
deserve	to	18	these	kids	72
don't	deserve	5	they	say	432
educated	they	2	they	to	3
kids	don't	21	to	be	5490

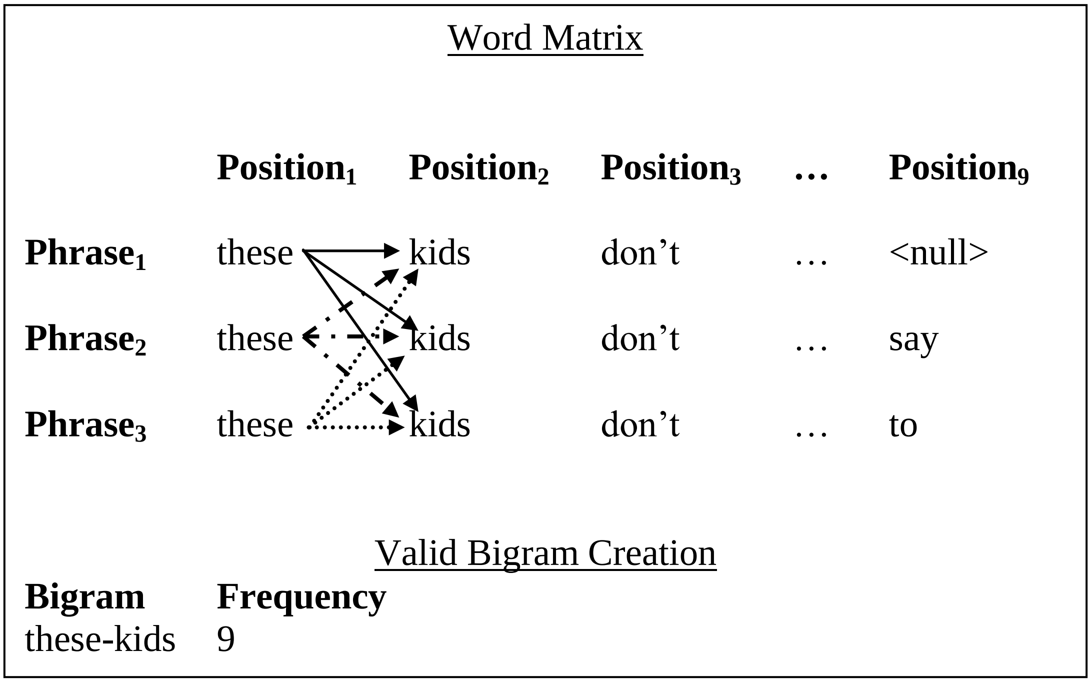
Table 7 PRA – DL Utterances Corpus

The first step of the algorithm is to store the evidence from the recognitions of all three listeners, resulting in an evidence bag that contains 11 unique bigrams (table 8).

Bigram		Bigram		Bigram	
Word1	Word2	Word1	Word2	Word1	Word2
these	kids	kids	don’t	don’t	deserve
deserve	to	to	be	be	educated
educated	basic	educated	they	they	say
educated	these	these	to		

Table 8 PRA Evidence Bigrams

Next, the local bigram frequency count is calculated and begins by combining the first and second words of each listener. As bigrams are created and validated against either the evidence (table 8) or the corpus (table 7), they are put into a temporary table that maintains the unique list of bigrams and their cumulative total frequency count. In keeping with the PRA example, after the first round of iterations for each recognition result, the temporary table has the bigram “these-kids” with a frequency count of 9, as shown in figure 17. Within the figure, <null> indicates an empty word position.



Word Matrix

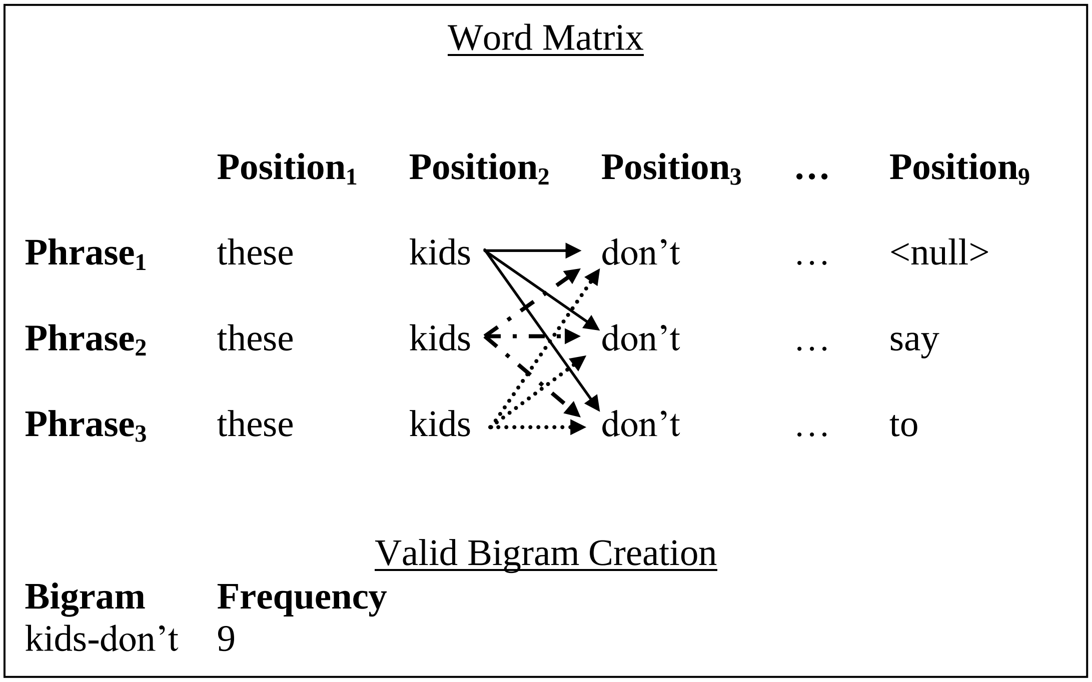
	Position$_1$	**Position$_2$**	**Position$_3$**	**…**	**Position$_9$**
Phrase$_1$	these	kids	don’t	…	<null>
Phrase$_2$	these	kids	don’t	…	say
Phrase$_3$	these	kids	don’t	…	to

Valid Bigram Creation

Bigram	**Frequency**
these-kids	9

Figure 17 PRA Round 1 Valid Bigram Creation and Local Frequency Count

The next round of iterations involves the second and third words from each listener (figure 18).



Word Matrix

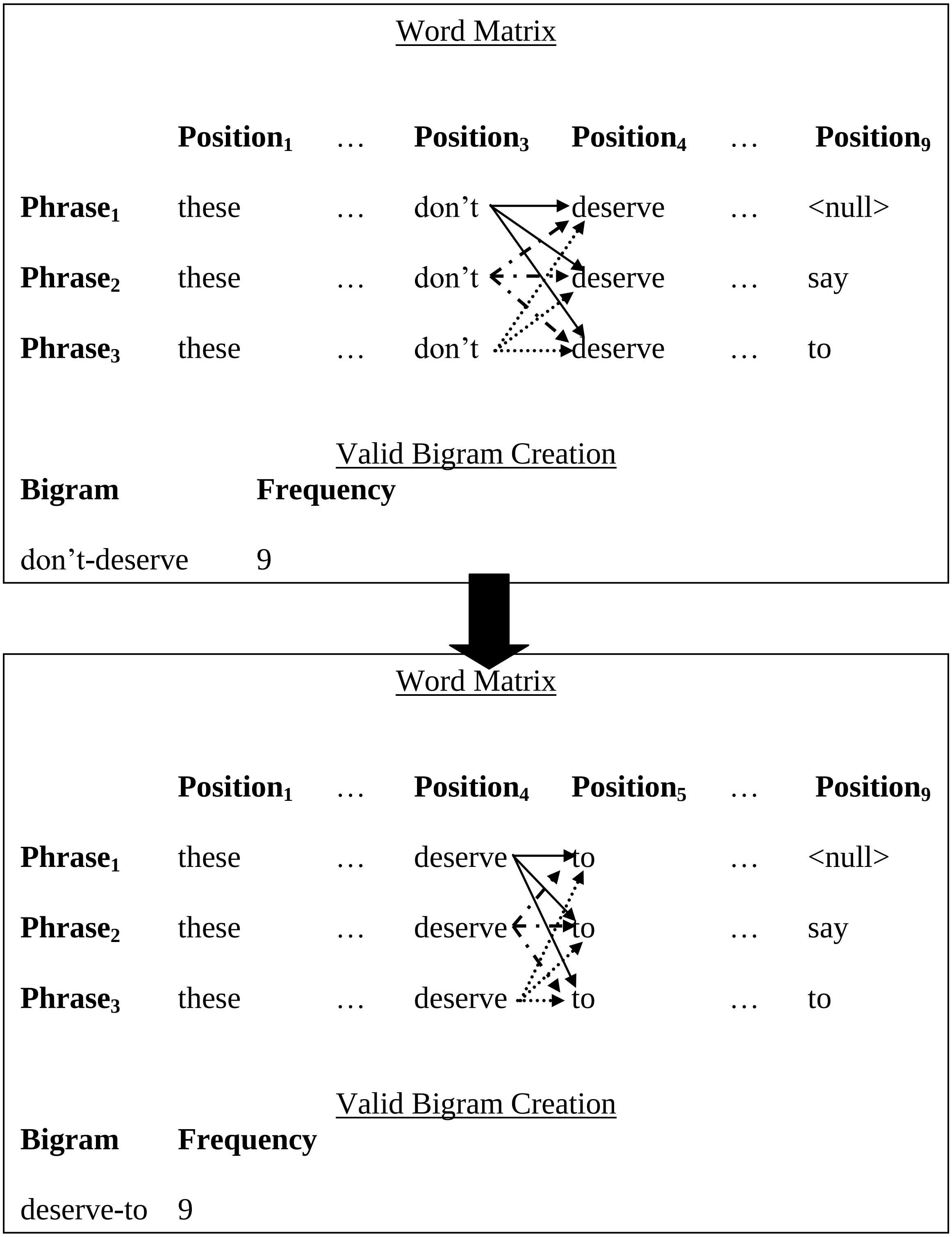
	Position$_1$	**Position$_2$**	**Position$_3$**	**…**	**Position$_9$**
Phrase$_1$	these	kids	don’t	…	<null>
Phrase$_2$	these	kids	don’t	…	say
Phrase$_3$	these	kids	don’t	…	to

Valid Bigram Creation

Bigram	**Frequency**
kids-don’t	9

Figure 18 PRA Round 2 Valid Bigram Creation and Local Frequency Count

The next two rounds of iterations and valid bigram creation are shown in figure 19.



Word Matrix

	Position$_1$	…	**Position$_3$**	**Position$_4$**	…	**Position$_9$**
Phrase$_1$	these	…	don’t	deserve	…	<null>
Phrase$_2$	these	…	don’t	deserve	…	say
Phrase$_3$	these	…	don’t	deserve	…	to

Valid Bigram Creation

Bigram	**Frequency**
don’t-deserve	9

Word Matrix

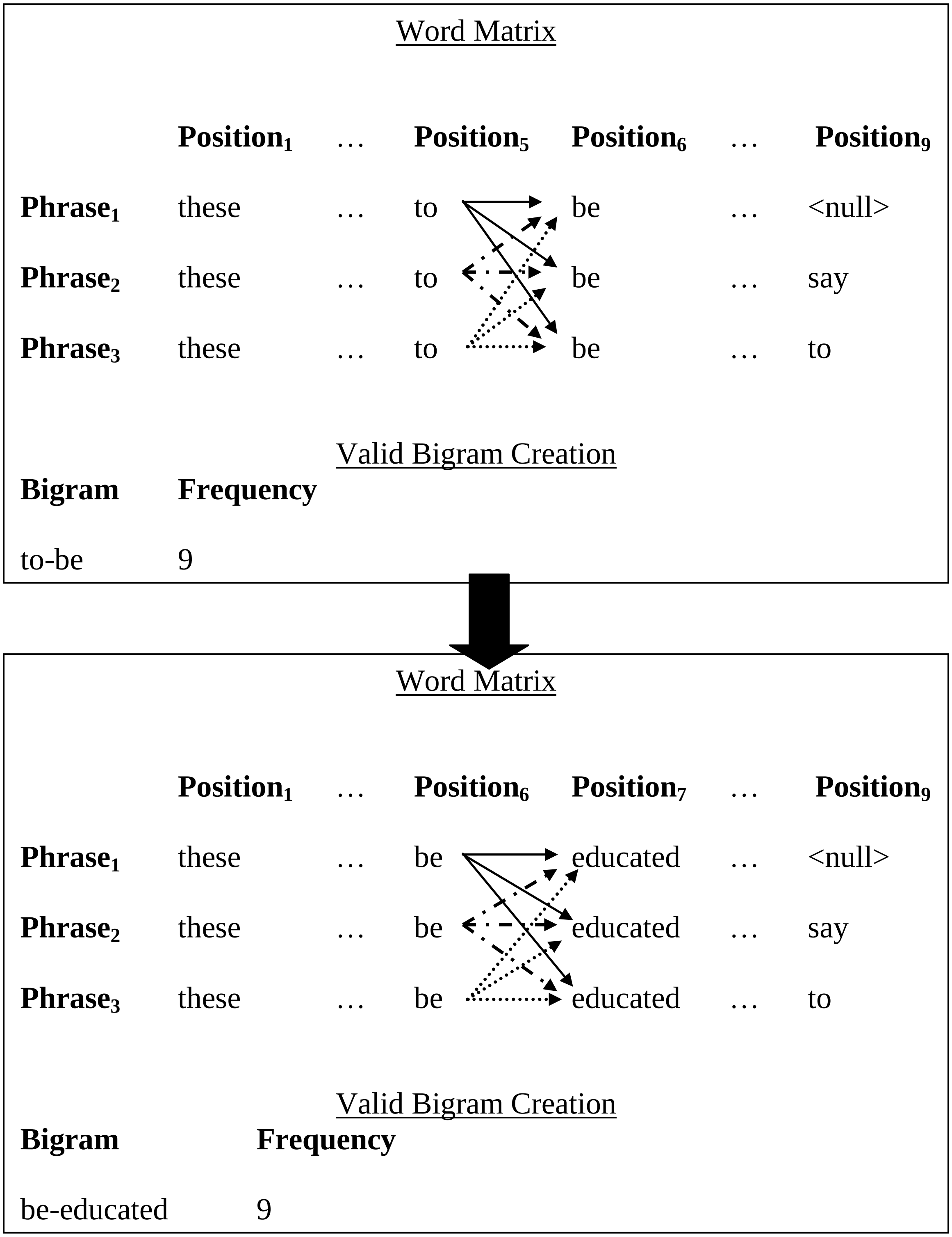
	Position$_1$	…	**Position$_4$**	**Position$_5$**	…	**Position$_9$**
Phrase$_1$	these	…	deserve	to	…	<null>
Phrase$_2$	these	…	deserve	to	…	say
Phrase$_3$	these	…	deserve	to	…	to

Valid Bigram Creation

Bigram	**Frequency**
deserve-to	9

Figure 19 PRA Rounds 3 and 4 Valid Bigram Creation and Local Frequency Count

Likewise, the bigrams for rounds 5 and 6 are valid, as illustrated in figure 20



Word Matrix

	Position$_1$	…	**Position$_5$**	**Position$_6$**	…	**Position$_9$**
Phrase$_1$	these	…	to	be	…	<null>
Phrase$_2$	these	…	to	be	…	say
Phrase$_3$	these	…	to	be	…	to

Valid Bigram Creation

Bigram	**Frequency**
to-be	9

Word Matrix

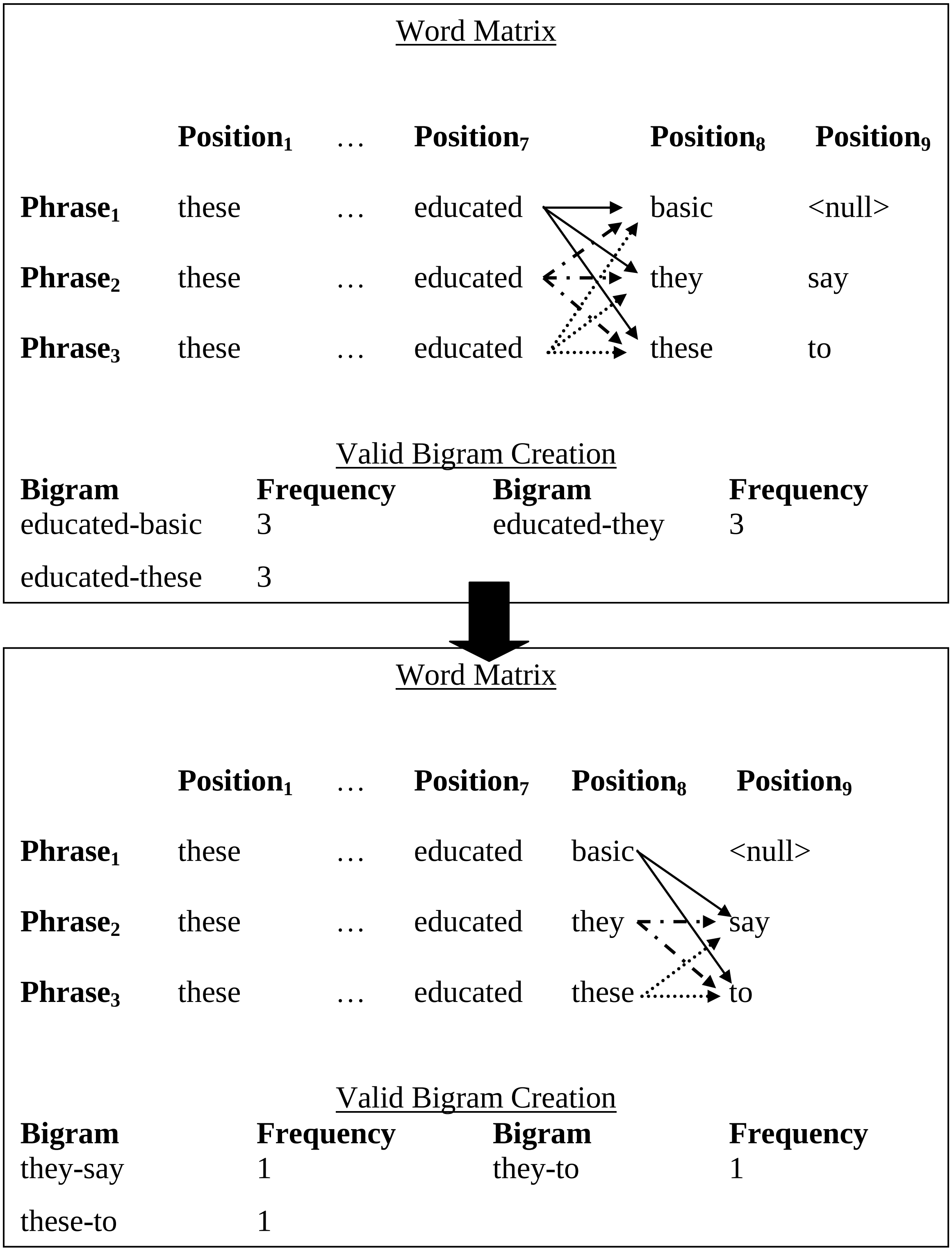
	Position$_1$	…	**Position$_6$**	**Position$_7$**	…	**Position$_9$**
Phrase$_1$	these	…	be	educated	…	<null>
Phrase$_2$	these	…	be	educated	…	say
Phrase$_3$	these	…	be	educated	…	to

Valid Bigram Creation

Bigram	**Frequency**
be-educated	9

Figure 20 PRA Rounds 5 and 6 Valid Bigram Creation and Local Frequency Count

The final two rounds and resulting valid bigrams are shown in figure 21.



Word Matrix

	Position$_1$	…	**Position$_7$**	**Position$_8$**	**Position$_9$**
Phrase$_1$	these	…	educated	basic	<null>
Phrase$_2$	these	…	educated	they	say
Phrase$_3$	these	…	educated	these	to

Valid Bigram Creation

Bigram	**Frequency**	**Bigram**	**Frequency**
educated-basic	3	educated-they	3
educated-these	3		

Word Matrix

	Position$_1$	…	**Position$_7$**	**Position$_8$**	**Position$_9$**
Phrase$_1$	these	…	educated	basic	<null>
Phrase$_2$	these	…	educated	they	say
Phrase$_3$	these	…	educated	these	to

Valid Bigram Creation

Bigram	**Frequency**	**Bigram**	**Frequency**
they-say	1	they-to	1
these-to	1		

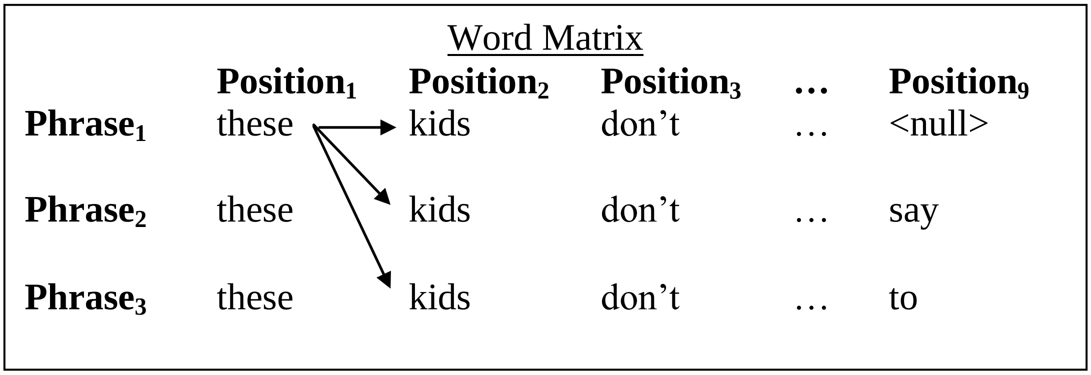
Figure 21 PRA Rounds 7 and 8 Bigram Creation and Local Frequency Count

Table 9 lists the resulting bigrams and local frequency counts from this process. Recall that only those bigrams that were validated against either the evidence or the corpus were counted, and therefore the local bigram and frequency count includes the evidence bigrams.

Bigram		**Frequency Count**	**Bigram**		**Frequency Count**
Word1	**Word 2**		**Word 1**	**Word 2**	
be	educated	9	kids	don't	9
deserve	to	9	these	kids	9
don't	deserve	9	these	to	1
educated	these	3	they	say	1
educated	basic	3	they	to	1
educated	they	3	to	be	9

Table 9 PRA Local Bigrams and Corresponding Frequency Counts

Now that the evidence has been collected and the local bigram frequency count has been calculated, the iterative process begins in parallel with each listener. The remainder of this example will focus on listener 1. The purpose of this iterative process is to build unique candidate strings that will ultimately be passed to the master interpreter to represent the most likely spoken phrase. The beginning of the candidate string(s) consists of the first word of listener 1, concatenated with the second word of each listener. This concatenation, as shown in figure 22, results in one unique candidate string: these kids.

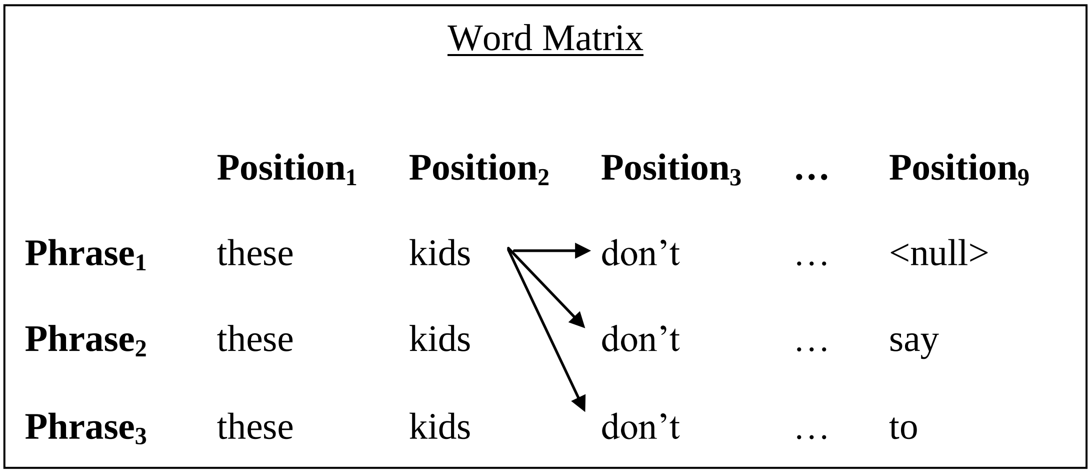


Word Matrix

	Position$_1$	**Position$_2$**	**Position$_3$**	**…**	**Position$_9$**
Phrase$_1$	these	kids	don’t	…	<null>
Phrase$_2$	these	kids	don’t	…	say
Phrase$_3$	these	kids	don’t	…	to

Figure 22 Listener 1-Round 1 PRA Iteration

Further concatenations to build up the candidate strings follow a simple matching procedure. If the last word of the current candidate string is equal to the first word of a bigram in the current round, a concatenation can be made. In keeping with the example, there is currently one candidate string: “these kids”. The next round of valid and unique bigrams for listener 1 consists of “kids-don’t”, as shown in figure 23. To build up the candidate string, the algorithm takes the last word of the current candidate string (i.e., kids) and compares it to the first word of the bigram in the current round (i.e., kids). Since the two words are equal, a concatenation can be made (Figure 24). The resulting candidate phrase is now: “these kids don’t”.



Word Matrix

	Position$_1$	**Position$_2$**	**Position$_3$**	**…**	**Position$_9$**
Phrase$_1$	these	kids	don’t	…	<null>
Phrase$_2$	these	kids	don’t	…	say
Phrase$_3$	these	kids	don’t	…	to

Figure 23 Listener 1-Round 2 PRA Iteration

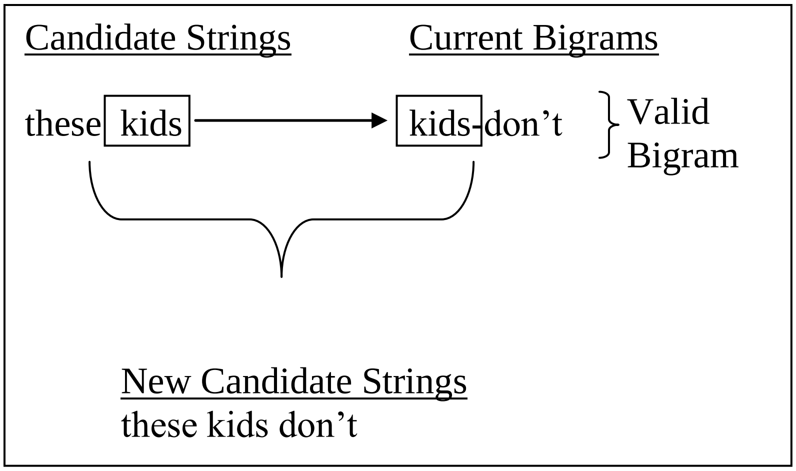
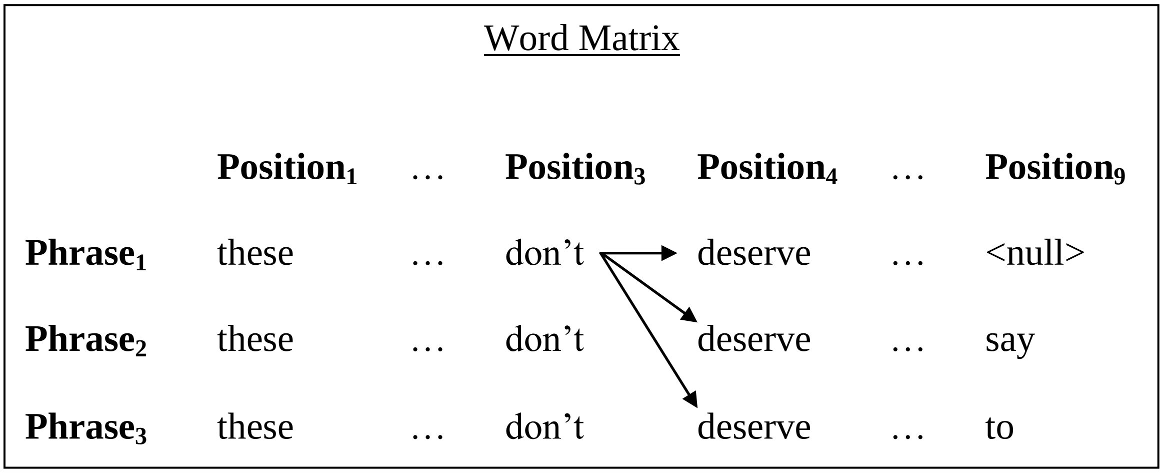


Figure 24 Listener 1-Round 2 Candidate Phrase Concatenations

The next round of iterations involves word 3 of listener 1 and word 4 of each listener (figure 25). The new unique bigram is: don't-deserve.



Word Matrix

	Position$_1$	…	**Position$_3$**	**Position$_4$**	…	**Position$_9$**
Phrase$_1$	these	…	don't	deserve	…	<null>
Phrase$_2$	these	…	don't	deserve	…	say
Phrase$_3$	these	…	don't	deserve	…	to

Figure 25 Listener 1-Round 3 PRA Iterations

The newly created bigram is valid, according to table 9, and is subsequently used to continuing building the candidate strings, as shown in figure 26.

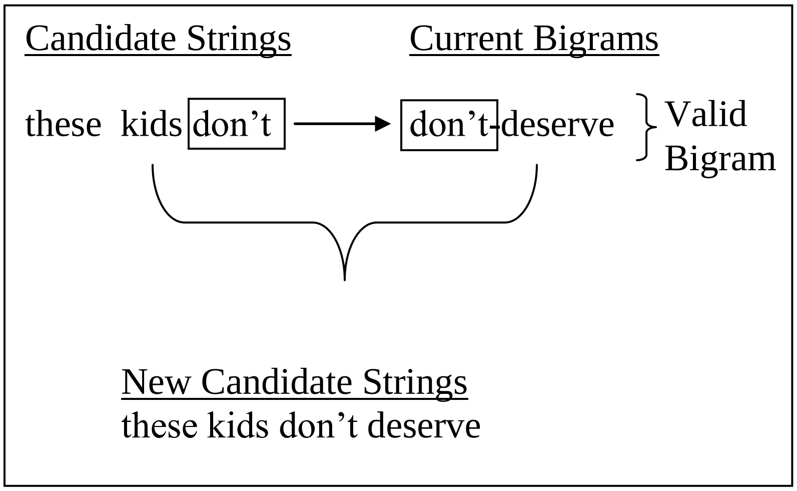
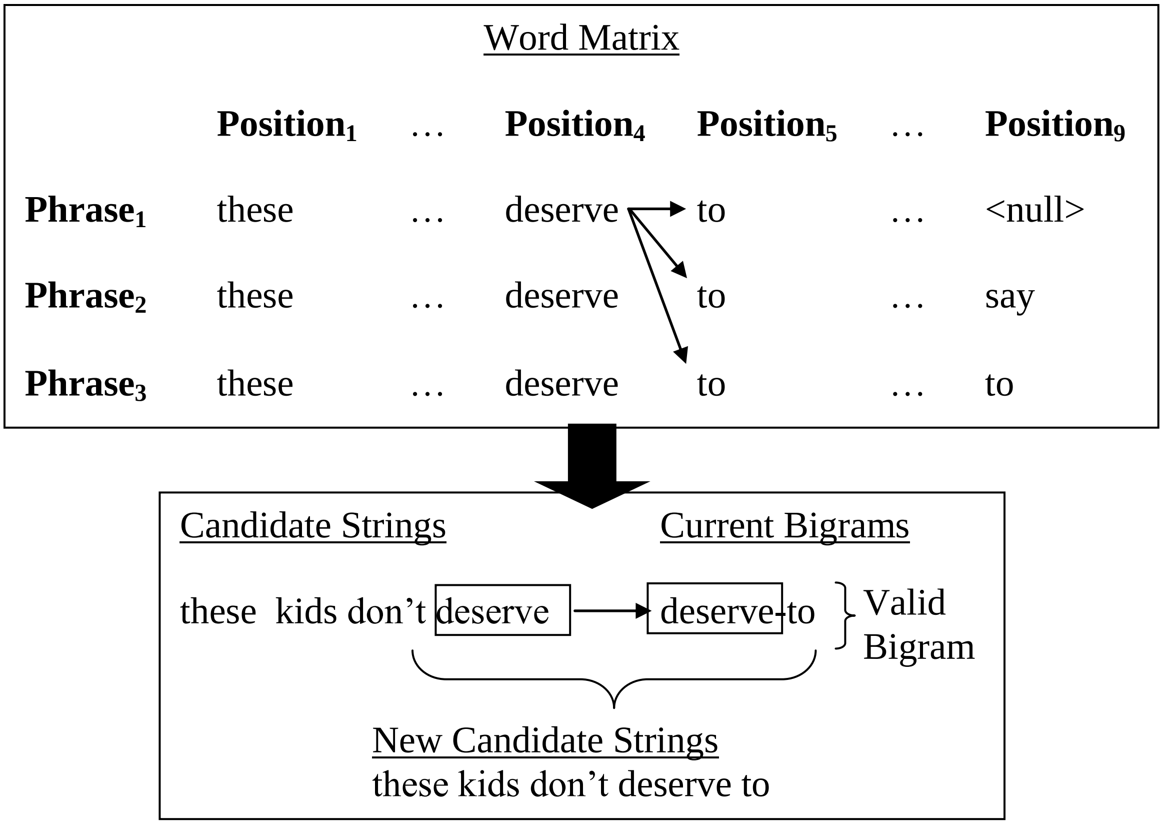


Figure 26 Listener 1-Round 3 PRA Concatenations

The next round of iterations proceeds in the same manner, as illustrated in figure 27.

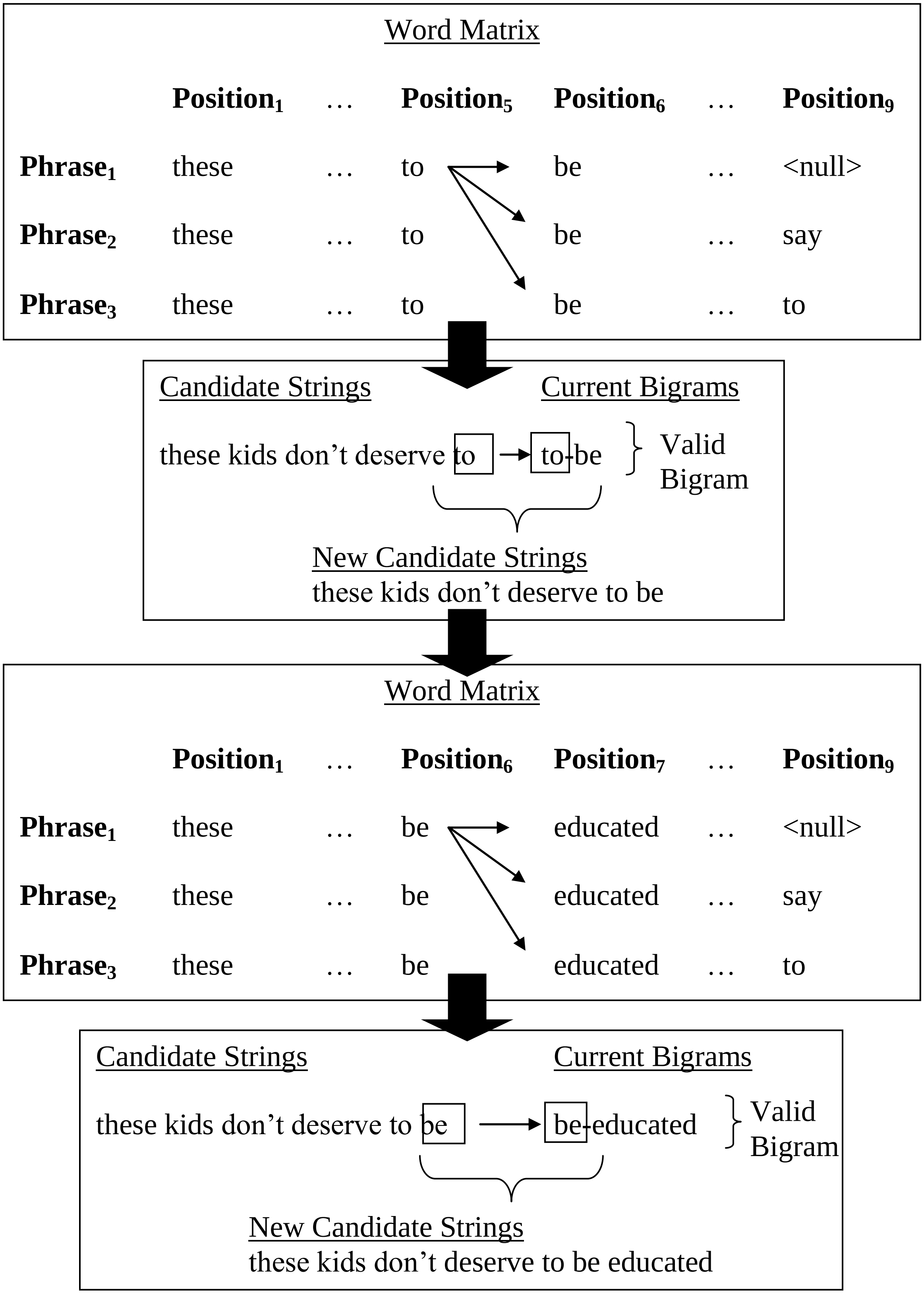


Word Matrix

	Position_1	…	**Position_4**	**Position_5**	…	**Position_9**
Phrase_1	these	…	deserve	to	…	<null>
Phrase_2	these	…	deserve	to	…	say
Phrase_3	these	…	deserve	to	…	to

Figure 27 Listener 1-Round 4 PRA Iteration and Concatenation

Likewise, rounds 5 and 6 complete iterations as shown in figure 28.



Word Matrix

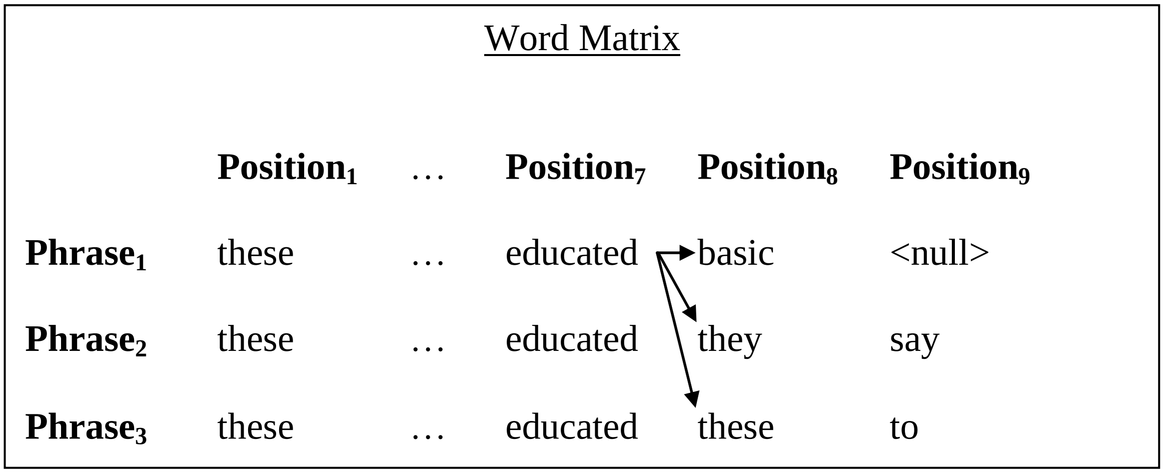
	Position$_1$	…	Position$_5$	Position$_6$	…	Position$_9$
Phrase$_1$	these	…	to	be	…	<null>
Phrase$_2$	these	…	to	be	…	say
Phrase$_3$	these	…	to	be	…	to

Word Matrix

	Position$_1$	…	Position$_6$	Position$_7$	…	Position$_9$
Phrase$_1$	these	…	be	educated	…	<null>
Phrase$_2$	these	…	be	educated	…	say
Phrase$_3$	these	…	be	educated	…	to

Figure 28 Listener 1-Rounds 5 and 6 PRA Iterations

Due to the fact that the first 7 words of the all three listeners are exactly the same, the iterative process produces only one candidate string until the 8th word is reached. Up to this point, the one candidate string is: “these kids don’t deserve to be educated”. The next subsequent round of iterations involves word 7 of listener 1, along with word 8 of each listener, as displayed in figure 29.



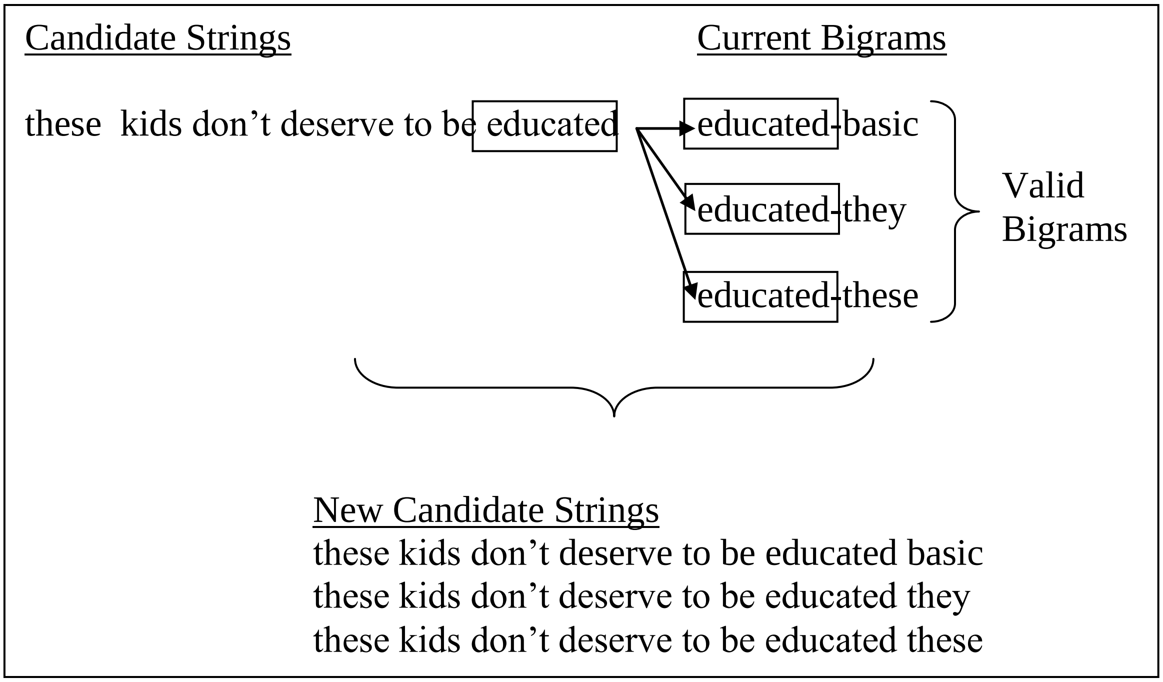
Word Matrix

	Position_1	…	**Position_7**	**Position_8**	**Position_9**
Phrase_1	these	…	educated	basic	<null>
Phrase_2	these	…	educated	they	say
Phrase_3	these	…	educated	these	to

Figure 29 Listener 1-Round 7 PRA Iteration

As displayed in figure 30, this round of iterations produces three candidate strings:

1. these kids don’t deserve to be educated basic
2. these kids don’t deserve to be educated they
3. these kids don’t deserve to be educated these



Candidate Strings

these kids don’t deserve to be educated

Current Bigrams

educated-basic

educated-they

educated-these

Valid Bigrams

New Candidate Strings

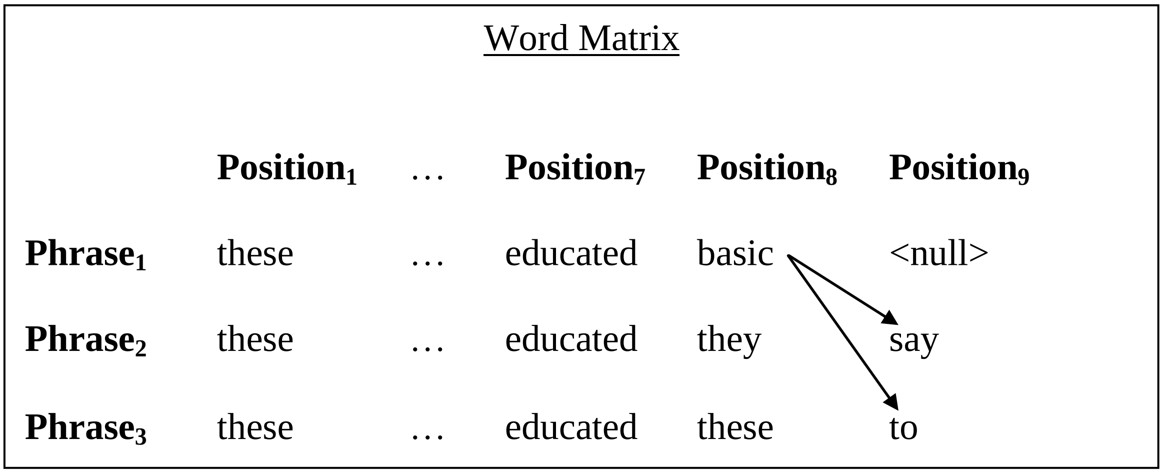
these kids don’t deserve to be educated basic

these kids don’t deserve to be educated they

these kids don’t deserve to be educated these

Figure 30 Listener 1-Round 7 PRA Concatenations

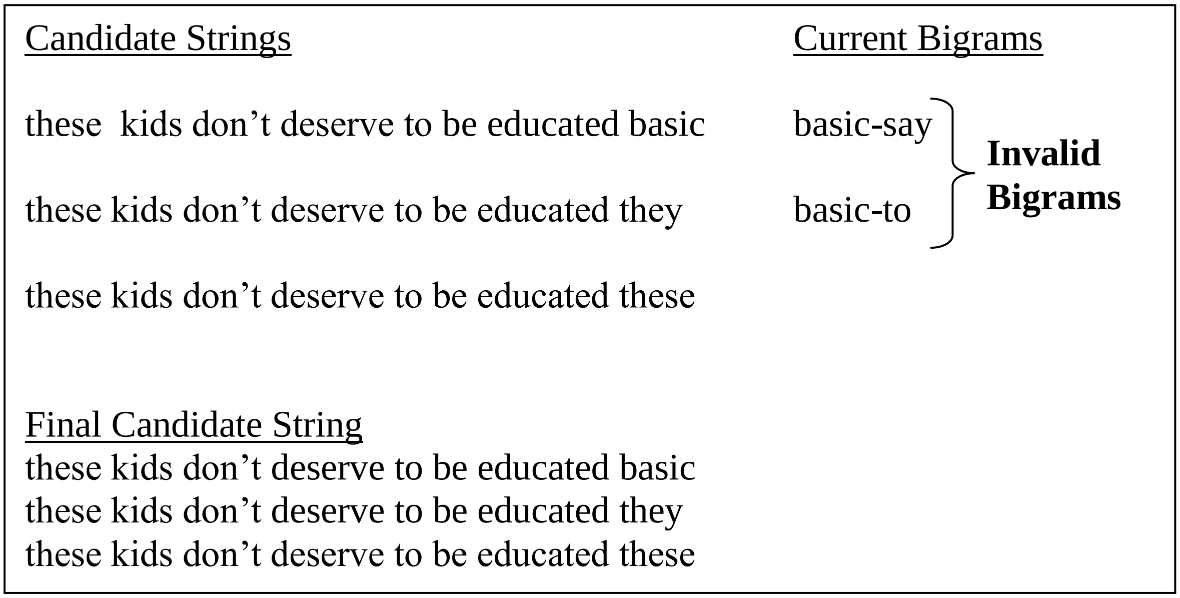
The next, and final, round of iterations for listener 1 produces two unique bigrams: basic-say and basic-to (figure 31). According to table 9, neither of those two bigrams is considered valid, and therefore, will not be used as concatenations to further build the candidate strings (figure 32).



Word Matrix

	Position$_1$	…	**Position$_7$**	**Position$_8$**	**Position$_9$**
Phrase$_1$	these	…	educated	basic	<null>
Phrase$_2$	these	…	educated	they	say
Phrase$_3$	these	…	educated	these	to

Figure 31 Listener 1-Round 8 PRA Iteration



Candidate Strings

these kids don’t deserve to be educated basic

these kids don’t deserve to be educated they

these kids don’t deserve to be educated these

Current Bigrams

basic-say

basic-to

Invalid Bigrams

Final Candidate String

these kids don’t deserve to be educated basic
these kids don’t deserve to be educated they
these kids don’t deserve to be educated these

Figure 32 Listener 1-Round 8 PRA Concatenations

The final candidate strings produced by listener 1 are as follows:

1. these kids don’t deserve to be educated basic
2. these kids don’t deserve to be educated they
3. these kids don’t deserve to be educated these

The next step of the algorithm is to compute the local bigram frequency total of the candidate strings that have been created, according to table 9, and send the candidate string(s) with the highest cumulative frequency count to the master interpreter. The set of candidate strings produced by listener 1 all have a total bigram frequency count of 57 and is shown in table 10.

Listener 1	
Local Bigram Frequency	**Candidate Phrase**
57	these kids don't deserve to be educated basic
57	these kids don't deserve to be educated they
57	these kids don't deserve to be educated these

Table 10 Listener 1 Set of Candidate Phrases and Frequency Counts

Listeners 2 and 3 proceed in parallel in the same manner as described for listener 1. At the end of this iterative process, once the bigram creation and concatenation procedure has been exhausted, listener 2 produces four candidate strings and corresponding frequency counts as shown in Table 11.

Listener 2	
Bigram Frequency	**Candidate Phrase**
58	these kids don't deserve to be educated they say
58	these kids don't deserve to be educated they to
57	these kids don't deserve to be educated basic
57	these kids don't deserve to be educated these

Table 11 Listener 2 Set of Candidate Phrases and Frequency Counts

At the completion of the parallel processing for listener 3, there are three candidate strings. Those candidate strings and corresponding local frequency counts are shown in table 12.

Listener 3	
Bigram Frequency	**Candidate Phrase**
58	these kids don't deserve to be educated these to
57	these kids don't deserve to be educated basic
57	these kids don't deserve to be educated they

Table 12 Listener 3 Set of Candidate Phrases and Frequency Counts

Each listener sends only the phrase with the highest frequency sum to the master interpreter. If a listener produced more than one phrase with the same frequency total, those phrases are also sent to the master interpreter. The final set of candidate phrases and frequency counts that will be sent to the master interpreter are listed in table 13.

Bigram Frequency	Candidate Phrase	Listener Number
57	these kids don't deserve to be educated basic	1
57	these kids don't deserve to be educated they	1
57	these kids don't deserve to be educated these	1
58	these kids don't deserve to be educated they say	2
58	these kids don't deserve to be educated they to	2
58	these kids don't deserve to be educated these to	3

Table 13 Master Interpreter Set of PRA Candidate Phrases

The final set of candidate strings consists of 6 phrases. It is now up to the master interpreter to select the phrase that will represent the most likely spoken phrase. Three of the six phrases that comprise the final set of candidate strings have a lower bigram frequency count of 57, than the other three phrases at 58 and are discarded by the master interpreter. The remaining three phrases all have the same local bigram frequency count and the master interpreter must break the tie between the three candidate strings, in order to select the string that will serve as the most likely spoken phrase. To accomplish this, the master interpreter checks for equality between the candidate strings and the recognition results from the listeners. Two of the three candidate phrases are the same as the results from the listeners and make up the new set of candidate phrases:

1. these kids don’t deserve to be educated they say
2. these kids don’t deserve to be educated these to

The next step of the algorithm is to validate the two candidate phrases against the corpus. Since all of the bigrams within the candidate phrases are maintained in the corpus (table 7), neither candidate phrase is discarded.

Finally, the sum of the frequencies of the bigrams is re-calculated for each of the two candidate strings according to the corpus (table 7). The final two phrases and their corresponding frequency counts are listed in table 14, with the winner shown in bold. The higher sum of those two phrases is 6,045 and corresponds to the following phrase: these kids don’t deserve to be educated they say. This phrase is returned by Distributed Listening as the most likely spoken phrase, which is 100% correct when compared to the actual spoken phrase of the user.

Bigram Frequency	Candidate Phrase
6,045	**these kids don't deserve to be educated they say**
5,615	these kids don't deserve to be educated these to

Table 14 PRA Most Likely Spoken Phrase

The full system design of Distributed Listening is comprised of listeners, interpreters, and a reconciliation algorithm that are all supported by a corpus and a temporary storage medium. The theoretical design and preliminary results suggest that this system design was practical and possible, yet a full experiment was needed to ultimately rate the effectiveness of the system design. A number of experiments were run for that purpose and the experimental setup will be discussed in the next chapter.

4 Experiment Design

The goal of the research experiments was to effectively evaluate Distributed Listening with regard to recognition accuracy rates when combining recognitions of individual listeners through a distributed approach. Recall that the hypothesis is that Distributed Listening will only perform as bad as the best individual listener. The experiments were necessary to support and validate the premise of Distributed Listening and will be discussed in the remainder of this chapter.

4.1 Internal Review Board

The first aspect of the experiment that had to be addressed was the approval from the Internal Review Board (IRB). The IRB is an institutional entity that protects the rights of human research participants. Since Distributed Listening relies on spoken input from humans, IRB approval was necessary before the experiment could begin. The spoken inputs used in the experiment, as described in section 4.2, were collected from the internet and were readily available and open to the public, therefore the final IRB approval designation was exempt under 45 CFR 46.110(b)(4). This IRB request was very straight-forward as there was no need to find users with specialized knowledge, there was no Graphical User Interface that required a user satisfaction evaluation, nor was there a need to evaluate the resulting system with regard to usability. Details of the IRB approval can be found in appendix D.

4.2 Data

The data needed for the experiment were in the form of spoken utterances and were taken at random from an online broadcast of the National Public Radio. In lieu of having a variety of users read the same set of phrases to the listeners, 56 random phrases were selected from the broadcast and used instead (see table 15). Each phrase was saved as an individual MP3 file. The average length of the utterances in words was 13, with the shortest phrase containing 4 words and the longest phrase containing 25 words. Using the recordings in this manner ensured consistency, eliminated the chance of having filler words, such as “ah” and “um”, embedded into an utterance, and kept the speaking rate uniform. In addition, the 56 recordings underwent a gain procedure using MP3Gain (MP3Gain 2010). MP3Gain is a software application that optimizes audio recordings so that a set of files have the same average loudness level without sacrificing quality or re-encoding. This is to ensure that when playback of the audio switches from one file to the next, the volume level will remain constant. An inherent problem to this type of normalization is clipping, where certain files are clipped so that they do not exceed the maximum allowable decibel level. The clipping creates a rough scratchy sound during loud parts of the audio recording. Most MP3 files will not have clipping at 89.0 dB, which was the setting used to normalize the 56 files.

Transcription	Word Length	Seconds
your handwriting has changed	4	1.797
it's become increasingly physical	4	2.199
how long can he maintain his equilibrium	7	2.947
women's rights advocates are cautiously hopeful now	7	2.939
it’s a disease that affects how you move	8	2.445
an arm doesn't swing quite the same way	8	2.328
there are medications that make a real difference	8	2.950
spend or cut taxes that is the question	8	3.355
but are there any bright spots out there	8	2.276
the temporary job market may provide an answer	8	2.888
the authorities have been slow to address it	8	2.270
a lot has happened before you first notice it	9	3.271
and you'd stop the disease before it even starts	9	2.458
would stem cell research figure into the genetic connection	9	3.764
and it does appear that there is a relationship	9	2.670
it helps with some problems and not with others	9	2.987
these kids don't deserve to be educated they say	9	2.947
an egyptian court convicted a man of sexual harassment	9	3.085
and we also saw that a number of researchers left	10	2.779
i also understand the moral objections that some people have	10	4.070
but he was young and this had happened seemingly overnight	10	3.151
genetics load the gun and the environment pulls the trigger	10	3.242
let them rot in their dead end low class jobs	10	3.180
that's why those real time classroom scenes are so startling	10	4.406
then they all take a deep breath and enter the arena	11	3.093
and that it wasn't immune from last year's deteriorating economic environment	11	4.085
a finger that wiggles and you can't really get it to stop	12	3.566
is well in motion before any of these symptoms first become clear	12	4.469
the only reason i really did was because of my family history	12	3.968
and that was the prevailing wisdom for a very very long time	12	3.363
but they don't necessarily need or want to know more than that	12	2.976
but just when you're getting a warm utopian feeling something bad happens	12	4.668
but the more we learn about the disease the more complicated it becomes	13	3.370
in lots of ways it made the research task that much more complex	13	4.752
and now it's nominated for an academy award for best foreign language feature	13	4.034
for a brief spell they seem younger more open and ready to learn	13	5.252

Transcription	Word Length	Seconds
i'll say the hero of the movie threatens to become its bad guy	13	4.551
and by then the lesson has been derailed and there are snickers all around	14	4.785
what finally rouses most of his students is an assignment to write self portraits	14	5.398
he writes on the blackboard and a student makes him stop and define a word	15	3.997
he tells his colleagues that it's the job of the teacher to bring kids out	15	4.231
pink slips seem to be raining down in just about every sector and every zip code	16	4.843
i noticed that i didn't think that my arm was swinging quite the same way when i jogged	18	4.939
i guess i choose to believe that i'll be able to do this for a very long time	18	4.252
the class is a semi improvised look inside a high school in a diverse working class paris neighborhood	18	6.211
the toy maker based here in los angeles said in november it would shed about a thousand jobs	18	5.047
they show you that at least until the system can be changed the battles will be moment to moment	19	7.294
but i think it's pretty clear that it had a negative effect on the way in which the field progressed	20	6.127
they went to other places that were more open to stem cell research whether that was in europe or in singapore	21	5.886
he had been in declining health for the past year dealing with the long term effects of the stroke he suffered	21	5.376
the first thing i remember noticing was this odd buzzing tingling sensation in my left leg and to some extent my left arm	22	9.297
republicans disagree complaining that the bill's tax cuts fall short and that it spends too much on things they say won't create jobs	23	7.848
sales of both product lines fell more than twenty percent last year as fewer people bought toys leading up to the holiday season	23	6.302
i used to feel that my cell phone was vibrating and i'd reach for it and then i'd find that there was nothing there	24	6.795
for a lot of people it's a dilemma because your faith might be telling you one thing and your body is telling you another	24	7.126
and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true	25	7.352

Table 15 Actual Spoken Phrases

4.3 Environment

The experiment was conducted in the human centered computing lab within the computer science and software engineering department at Auburn University. The lab was a semi-private room in that it was home to a few graduate students, but was not open to all of the students of the department.

Occasionally, background noise such as a door closing, a phone ringing, the hum from the air conditioning and heating unit, and a microwave signal indicating the end of a cooking cycle would be present. To ensure that the environment background noise remained constant throughout the duration of the experiment, a decibel meter was used to measure the noise level of the lab. A decibel (dB) is a measurement level for sound magnitude and has one of three common weightings (Pierce 1981). The decibel meter used in the experiment was capable of measuring either an A-weighting A-curve or C-weighting C-curve frequency characteristics. C-weighting is commonly used for musical material whereas A-weighting responds to frequencies in the 500-to-10,000 Hz range, which is the range most sensitive for human ears and was used in the experiment. The decibel meter is capable of measuring sound ranging from 50 dB to 126 dB, where average conversation is measured at 60 dB (Durrant and Lovrinic 1995). The average sound measurement of the lab was consistently below 50 dB and implied that there was no significant measure of background noise that would interfere with the speech recognition attempts.

4.4 Materials

A combination of hardware and software was used to implement the design and experimentation of Distributed Listening, including the listeners and the speech recognition software. Each of the components listed in this section and subsequent sub-sections are necessary and act as a team.

4.4.1 Listeners

Distributed Listening, in theory, is capable of processing recognitions from any number of listeners. For the scope of this experiment, three listeners were utilized as follows:

1. Listener 1: Toshiba Satellite A355-S6935 with an Intel® Core™ 2 Duo CPU T6400 and 2.99 GB RAM running Microsoft Windows XP Professional, Service Pack 3. This listener will also be referred to as Satellite.
2. Listener 2: Dell Latitude D830 with an Intel® Core™ 2 Duo CPU T7500 and 2.00 GB of RAM running Microsoft Windows XP Professional, Service Pack 3. Listener 2 will also be referred to as Dell.
3. Listener 3: Toshiba Satellite with a Genuine Intel® CPU T2600 and 3.24 GB RAM running Microsoft Windows XP Tablet Edition, Service Pack 3. This listener will also be referred to as Tablet.

Microsoft Windows XP has an option in the Speech Control Panel to train profiles to improve recognition accuracy. Each of the listeners has this option and it was not utilized in order to support the over-arching goal of Distributed Listening in that the system improves speech recognition accuracy in sub-optimal speaker-independent environments.

4.4.2 Dragon NaturallySpeaking 10

Dragon NaturallySpeaking is a commercially available speech recognition software solution created by Nuance Communications that includes a recognition engine. It is a speaker dependant recognition system that is optimized when profiles are created and trained for the way specific users speak and the way they use words in regard to a vocabulary and language model.

Since a premise of Distributed Listening is to increase accuracy on untrained systems, the user profile that was created for the experiment was not trained or optimized. Dragon NaturallySpeaking 10 was loaded onto each of the listeners and was the interface between capturing the audio recordings and transcribing them.

Dragon NaturallySpeaking 10 has tool called DragonPad, which is a built in word processing feature, optimized for dictation. While Dragon NaturallySpeaking will work with other word processors, it was a natural selection to use the provided tool.

4.4.3 Corpus

The corpus that populated the database was taken from the open portion of the American National Corpus (OANC) (Ide and Suderman 2007). The open portion of the OANC contains approximately 15 million words from both spoken and written sources. The words from the OANC are presented in text files. Those text files were parsed using a script, the bigrams were found, and a MySQL database table was populated with the bigrams and a frequency count for each bigram. Because the OANC included spoken and transcribed words, there were bigrams that were removed. Those included bigrams that had filler words, like “um” and “ah”, repeated words that came from stuttering, and grammatically incorrect bigrams like “it’s is”. It’s important to note that not all repeated words are assumed incorrect. For example, the sentence “it’s been a very very long trip” contains the bigram “very-very” that can be assumed was deliberately used to put emphasis on the type of trip that was taken. In addition, punctuation marks were removed and hyphens were replaced with spaces within the OANC text files. The final corpus contained 456,981 unique bigrams with a total frequency count of 3,291,722.

4.5 Procedure

Three separate tests were run, with each test having 56 trials. The differences in the three tests were the ways in which the audio was collected by the listeners, otherwise the procedure was the same and adhered to the following steps:

1. Each listener received the 56 audio recordings in the same order.
2. The audio was processed using the Dragon NaturallySpeaking 10 software.
3. The transcribed recognitions of each listener were processed through the Distributed Listening algorithm.
4. For each of the 56 recordings, the algorithm produced the most likely spoken phrase, which was subsequently saved to a text file for further review.

The difference in this procedure occurred in step 1. There were three separate methods used for capturing audio by the listeners, as described in the next three subsections.

4.5.1 Stereo Mix Option

The first test captured the audio using the stereo mix option of the laptops. Stereo mix is an option that allows a computer to hear sound that is playing through the computer’s sound card. The computers microphone is not capturing the sound playing through the speakers; rather the computer is internally capturing sound playing through the sound card. This was done by loading the 56 separate recordings into Windows Media Player and playing them sequentially, with each listener. As the recordings played through Windows Media Player, Dragon NaturallySpeaking 10 captured the recognitions. The 56 recognitions were saved to a text file, to be used with the recognitions from the other two listeners.

4.5.2 External Speaker and External Microphone Option

The second test mimicked the way in which a typical person would speak to a computer, meaning using a typical desktop microphone plugged into the standard “mic in” jack of the laptop. When a person speaks into a microphone, placement of the microphone, as well as the volume of the speech, is important. Therefore, each laptop microphone was configured for optimal volume level, and that level of speech was recorded using the decibel meter. After which, the 56 recordings were loaded onto a SONY ICD-P17 digital recorder and an external speaker was plugged into the ear jack of the recorder. The volume of the speaker was then set to the optimized level that was found using the decibel meter (approximately 70 dB). The external speaker was then placed approximately two inches away from the external microphone (as is the case when a person speaks into a desktop microphone), Dragon Naturally Speaking was started, and the recognitions were captured and saved to a text file (figure 33).

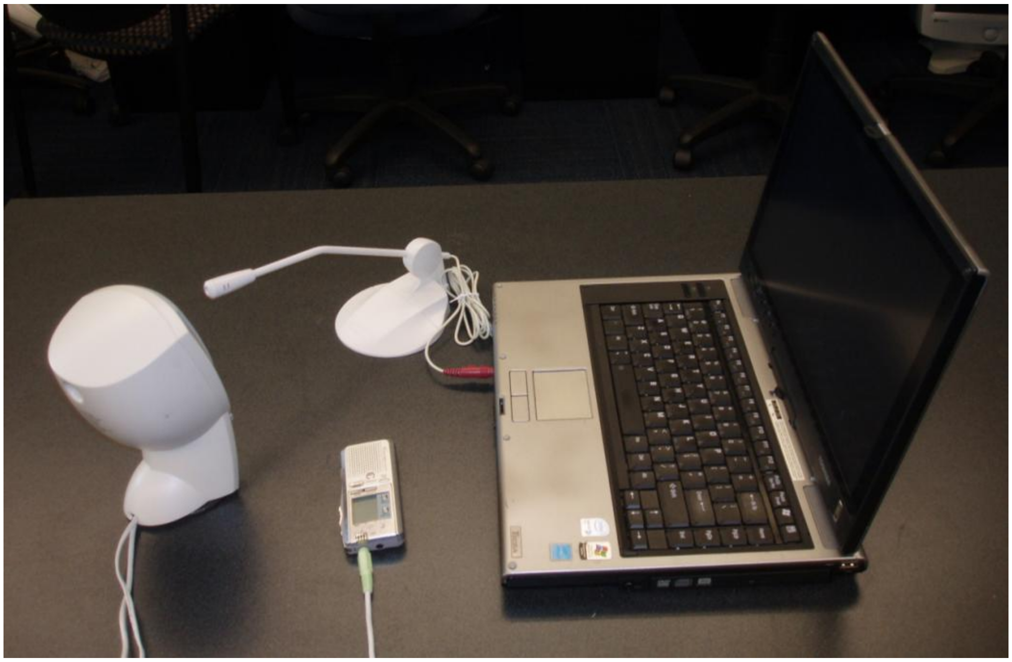


Figure 33 External Speaker/Microphone Experiment Setup

4.5.3 3.5mm Cable Option

The last test involved a 3.5mm male-to-male audio cable. One end of the cable was plugged into the ear jack of the digital recorder and the other end was plugged into the microphone jack of the laptop, using the line-in option. Line-in allows a computer to capture audio devices that are connected to the computer. Once the cable was connected and the settings selected, Dragon NaturallySpeaking 10 was loaded and the 56 files were played through the digital recorder (figure 34). Dragon NaturallySpeaking captured the recognitions that were ultimately saved to a text file and used with the recognitions from the other two listeners.

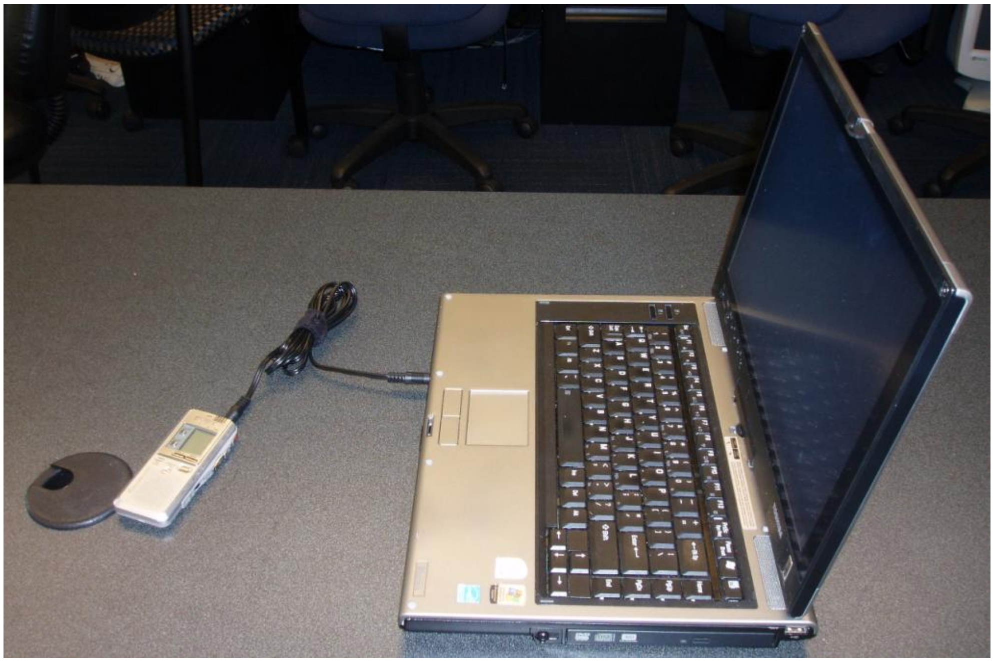


Figure 34 3.5mm Cable Experiment Setup

The three separate experiments each produced very different results and will be discussed in detail in the next chapter.

5 Research Findings

The first metric used to determine the success of Distributed Listening was the accuracy percentage of how often the result from Distributed Listening matched the actual spoken phrase, compared to the same accuracy percentage of the individual listeners. The actual formula used was:

$$\text{Accuracy} = \frac{\text{Instances Where Result Matched Actual Spoken Phrase}}{\text{Total Number of Actual Spoken Phrases}}$$

The three tests described previously in section 4.5 were further broken down into categories. After the completion of the 56 trials of each test, the comparative analysis had 6 designations:

1. Overall: This category compared the overall accuracy of the individual listeners with the result returned by Distributed Listening across all 56 trials.
2. Overall Interpretation: For all 56 trials, this category looked at the individual recognition results of the listeners, as well as the result of the Distributed Listening system, and counted those results that had a *semantic interpretation* that matched the actual spoken phrase as correct. If the semantic meaning of a recognition or Distributed Listening result resolved to the meaning of the actual spoken phrase, then that result had a correct semantic interpretation and was considered equal to the actual spoken phrase for calculating accuracy.

3. Valid: This category put a definition on the actual spoken phrases. If all of the bigrams of the actual spoken phrase were in the corpus, then the phrase was deemed valid. The individual listeners and the Distributed Listening result that corresponded to a valid actual spoken phrase were used to compute the accuracy in this category.
4. Valid Interpretation: This category looked at the subset of phrases that were deemed valid and if a *semantic interpretation* of the result of an individual listener or Distributed Listening matched the actual spoken phrase, it was also counted as correct when computing accuracy.
5. Invalid: This category was made up of those actual spoken phrases, and corresponding results, whose bigrams did not appear in the corpus.
6. Invalid Interpretation: From the subset of invalid actual spoken phrases, if a recognition or result had a *semantic interpretation* that resolved to the actual spoken phrase, it was also counted as correct.

The second metric used to test the success of Distributed Listening was the word error rate (WER), as calculated by the following formula:

$$\text{WER} = \frac{S + D + I}{N}$$

where:

S = The number of substitutions
D = The number of deletions
I = The number of insertions
N = The number of words in the correct word sequence

The WER is based on the Levenshtein distance, also called the edit distance, and is a common metric used for establishing the accuracy of speech recognition (McCowan et. al. 2005). The number returned by the formula is an indication of how similar two character sequences are, based on the number of character additions, substitutions, and deletions it takes to turn one character sequence into the other, with the minimum number of changes. In recent years, there have been research activities to establish a more accurate metric. Some scholars argue that the current WER is hard to interpret since it is quite possible to have a WER that can be greater than 1or a WER that can be negative. The fact that a recognition result can be significantly shorter or longer than the actual word sequence allows the WER to be greater than or less than unity. Additionally, the WER measure does not address word importance. Despite the validity of the arguments against the WER, a new metric has not been established and therefore, the WER is the metric that will be used for the data analysis of this research task. The reference character sequence for each WER calculation presented in this chapter will be the actual spoken phrase that corresponds to each listener's recognition results and the results of Distributed Listening. No additional evaluation measures will be used. Demographical characteristics were not possible given the nature of the data collection of the spoken phrases. A best guess could have been attempted to determine the gender of the speaker, but without explicitly being told, there was no way to definitively determine age, education level, experience with spoken language systems, reading level, nationality, or gender.

Each of the three tests underwent the same analysis as discussed in the next three subsections.

5.1 Stereo Mix Option

During this trial, the listeners had the best individual results and Distributed Listening returned optimal results, as shown in table 16. Within the overall category, across all 56 utterances, the best individual listener had a recognition accuracy of 48%, which was also the accuracy of Distributed Listening.

	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
Overall	36%	48%	39%	48%
Overall Interpretation	38%	50%	43%	52%
Valid	32%	50%	39%	54%
Valid Interpretation	32%	50%	43%	57%
Invalid	39%	46%	39%	43%
Invalid Interpretation	43%	50%	43%	46%

Table 16 Stereo Mix Option Recognition Accuracy Rates

The overall interpretation category showed an improvement in recognition accuracy over the overall category for all three listeners, as well as Distributed Listening. Listener 1 and listener 2 both improved by 2%, which is the equivalent of one additional result being counted as correct. Listener 3 and Distributed Listening improved by 4%, or two additional results being counted as correct. The best individual listener within the overall interpretation category had an accuracy of 50%, which Distributed Listening exceeded at 52%. The correct semantic interpretations of the listeners and Distributed Listening are displayed in table 17, with the corresponding semantic equivalents shown as underlined text.

Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening	Actual Spoken Phrase
it's a disease that affects how you move	is a disease that affects how you move	it is a disease that affects how you move	it is a disease that affects how you move	it's a disease that affects how you move
but he was young and this happened seemingly overnight	but he was young and this happened seemingly overnight	but he was young and this happened seemingly overnight	but he was young and this happened seemingly overnight	but he was young and this had happened seemingly overnight

Table 17 Correct Semantic Interpretations of the Actual Spoken Phrase

For the valid phrases category, Distributed Listening had an accuracy rate of 54%, which was better than the best individual listener.

The valid interpretation category showed that only listener 3 and Distributed Listening improved over the valid phrases category. Listener 3 improved from 39% to 43%, yet still did not have the best individual accuracy. The best individual listener achieved an accuracy of 50%, compared to Distributed Listening at 54%.

The last two categories, invalid and invalid interpretation, showed that Distributed Listening did not beat the best individual listener and is based on a special case. In one of the trials, listener 1 and listener 3 agreed word-for-word on their recognition results, yet those resulting recognitions were not equal to the actual spoken phrase. In contrast, listener 2 recognized the actual spoken phrase at 100% correctness (table 18). This caused Distributed Listening to return as the most likely spoken phrase, the recognition from the listeners that agreed. Because of this special case, the accuracy result of Distributed Listening in both categories is less than optimal. If that particular trial is omitted from the calculations, then the Distributed Listening accuracy rate would be equal to that of the best individual listener in both the invalid and invalid interpretation categories. The recognition results of the three listeners, the result from Distributed Listening, and the actual spoken phrase for the aforementioned trial are listed in table 18, with the differences of the phrases indicated with underlined text.

Listener 1 (Satellite)	**Listener 2 (Dell)**	**Listener 3 (Tablet)**	**Distributed Listening**	**Actual Spoken Phrase**
they show you that at least until the system can be changed but battles will be moment to moment	they show you that at least until the system can be changed the battles will be moment to moment	they show you that at least until the system can be changed but battles will be moment to moment	they show you that at least until the system can be changed but battles will be moment to moment	they show you that at least until the system can be changed the battles will be moment to moment

Table 18 Agreement Between Listeners on Incorrect Recognitions

For a complete listing of the recognition results from the listeners, the Distributed Listening results, and the actual spoken phrases from all 56 trials, refer to appendix A. Appendix A also indicates which of the actual spoken phrases are considered valid and which are considered invalid.

Distributed Listening returned optimal results with regard to the word error rate. Listener 1 had an average WER of 0.36. Listeners 2 and 3 had average word error rates of 0.31 and 0.34, respectively. Distributed Listening had the lowest average word error rate of

0.26. The word error rates of each listener and Distributed Listening for all 56 trials are displayed in table 19. The phrase numbers in the table correspond to the phrase numbers as listed in appendix A.

Phrase Number	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
1	0.00	0.25	0.25	0.25
2	0.58	0.00	0.42	0.00
3	0.38	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00
6	0.33	0.00	0.00	0.00
7	0.57	0.26	0.09	0.09
8	0.83	0.67	0.63	0.63
9	0.33	0.33	0.33	0.33
10	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00
12	0.89	0.22	0.22	0.22
13	0.44	0.44	0.44	0.44
14	0.31	0.54	0.62	0.54
15	0.00	0.00	0.00	0.00
16	0.40	0.70	0.40	0.40
17	0.00	0.33	0.00	0.00
18	0.10	0.20	0.20	0.20
19	0.00	0.00	0.00	0.00
20	0.08	0.08	0.08	0.08
21	0.40	0.40	0.40	0.40
22	0.00	0.00	0.00	0.00
23	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00
25	0.56	0.56	0.56	0.56
26	0.00	0.00	0.00	0.00
27	0.83	0.75	1.50	0.75
28	0.00	0.00	0.00	0.00
29	0.00	0.00	0.00	0.00
30	0.06	0.06	0.06	0.06
31	0.55	0.55	0.55	0.55
32	0.33	0.33	0.33	0.33
33	0.44	0.00	0.00	0.00
34	0.79	0.14	0.14	0.14
35	0.78	0.00	0.56	0.00

Phrase Number	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
36	0.40	0.00	0.00	0.00
37	0.00	0.00	0.00	0.00
38	0.00	0.00	0.20	0.00
39	0.79	1.00	0.79	0.79
40	0.38	0.00	0.00	0.00
41	0.00	0.00	0.00	0.00
42	0.15	0.38	0.15	0.15
43	2.10	1.20	1.70	1.70
44	0.16	0.00	0.16	0.16
45	0.35	0.35	0.22	0.35
46	1.13	0.83	1.09	1.09
47	1.67	1.17	1.67	1.67
48	0.18	0.73	1.00	0.18
49	0.50	0.88	0.38	0.50
50	0.94	1.00	0.94	0.94
51	0.00	0.00	0.88	0.00
52	0.00	0.00	0.00	0.00
53	0.75	2.25	0.75	0.75
54	0.13	0.63	0.88	0.13
55	0.00	0.00	0.00	0.00
56	0.57	0.00	0.57	0.00
Average	**0.36**	**0.31**	**0.34**	**0.26**

Table 19 Stereo Mix Word Error Rates

5.2 External Speaker and External Microphone Option

During this test, the individual listeners had the poorest recognition results. For a full listing of recognition results from this experiment, including the variability of the 3 listeners, see appendix B. Table 20 displays the actual accuracy results of the 56 trials, across the 6 categories. Notice that listener 1 (Satellite) produced 0% accuracy across all 6 categories. Individually, the other two listeners did not achieve accuracy rates that were much better than those of listener 1. Yet, across the 6 categories of this test, Distributed Listening consistently met the recognition accuracy of the best individual listener. Specifically, for each of the 6 categories, Distributed Listening matched the accuracy of the best individual listener.

	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
Overall	0%	7%	2%	7%
Overall Interpretation	0%	7%	2%	7%
Valid	0%	4%	0%	4%
Valid Interpretation	0%	4%	0%	4%
Invalid	0%	11%	4%	11%
Invalid Interpretation	0%	11%	4%	11%

Table 20 External Speaker/Microphone Option Recognition Accuracy Rates

This test is the only test that did not show an improvement in recognition accuracy for any of the listeners or Distributed Listening when semantic interpretation of the actual spoken phrases was taken into account.

The poor recognition accuracy of the listeners and the resulting percentages show that when nonsensical recognitions are combined with syntactically and grammatically correct recognitions and a reliable corpus, optimal results can still be produced.

It is also worth mentioning that one of the 56 trials did not produce a conclusive result (table 21). This is due to the fact that the number of permutations of a recognition result, and therefore the resulting potential candidate phrases, from just one listener numbered greater than 45,349,600. Even in a real-time parallel processing environment, an exhaustive listing of candidate phrases is prohibitive with regard to time. This is attributed to the accuracy of the individual listeners. When the individual listeners differ at every word, by position, and there are several words per recognition result, it is not feasible to produce an accurate result and Distributed Listening is not a practical solution in such environments. The accuracy results shown in table 20 are calculated out of 55 trials, since one round of trials was inconclusive.

Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening	Actual Spoken Phrase
though the feature set aside as he was the one that has been known	as for the picture is satisfied when they first what would be wrong with that and then know if	as for the future have to set aside when they are well with you will not admit to it	inconclusive	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true

Table 21 External Speaker/Microphone Inconclusive Round

The average word error rates of the three listeners and Distributed Listening are expected to be greater than one, because of the inconsistencies of the individual recognizers. Given that the individual listeners had such poor recognition accuracy and didn’t readily resemble the corresponding actual spoken phrase, it is expected that the WER will be quite high. Likewise, the resulting candidate phrases that were returned from Distributed Listening did not readily resemble the corresponding actual spoken phrase and had extremely high word error rates. Listener 1, which had 0% accuracy, had the highest average WER at 3.47. This WER was considerably larger than the rates of listeners 2 and 3, and over twice as great as that of Distributed Listening. Listener 2 had an average WER of 1.64, listener 3 had an average WER of 1.97, and Distributed Listening returned the lowest average WER of 1.60. Because of the inconclusive round, the average word error rates are calculated out of 55 trials and the inconclusive round is not listed in table 22.

Phrase Number	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
1	2.00	1.25	0.63	0.63
2	3.58	2.08	2.33	2.08
3	3.25	2.75	3.38	2.75
4	2.00	0.00	0.00	0.00
5	3.67	1.75	3.33	1.75
6	3.44	0.33	1.56	0.33
7	4.00	1.70	1.91	1.91
8	3.50	1.83	0.83	0.83
9	3.56	0.67	1.56	0.78
10	3.92	0.00	1.17	0.00
11	4.08	1.00	0.83	0.83
12	3.78	1.44	2.00	1.44
13	5.67	2.78	3.00	3.00
14	4.00	0.92	2.00	0.92
15	3.54	0.46	1.00	0.46
16	3.90	1.80	0.80	0.80
17	4.62	2.38	2.33	2.33
18	1.40	1.25	1.05	1.05
19	4.30	0.00	0.50	0.00
20	4.00	1.96	2.50	1.96
21	3.00	4.30	3.40	3.00
22	1.00	0.44	0.44	0.44
23	2.50	2.20	1.50	1.50
24	5.00	0.00	1.88	0.00
25	2.44	1.22	0.78	0.78
26	3.00	0.39	0.89	0.39
27	4.25	3.00	1.75	1.75
28	2.57	2.14	1.24	2.10
29	1.00	0.62	2.00	0.62
30	2.83	1.50	2.39	3.06
31	3.45	1.27	1.18	1.18
32	2.80	2.33	2.67	2.33
34	3.14	2.64	2.93	2.64
35	4.78	0.33	1.78	1.22
36	2.40	0.70	2.30	0.70

Phrase Number	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
37	4.14	1.71	0.71	1.71
38	3.80	1.60	2.00	1.87
39	5.36	2.07	4.21	2.07
40	3.92	2.23	3.54	2.23
41	4.42	2.92	4.00	3.42
42	3.54	1.38	2.08	2.08
43	4.80	3.50	3.10	3.50
44	3.26	1.05	2.42	1.05
45	2.48	1.91	2.17	2.17
46	4.22	3.17	3.74	3.09
47	3.83	2.72	2.94	2.94
48	5.09	4.00	5.18	5.18
49	3.50	1.63	2.25	1.63
50	2.13	3.13	2.50	2.56
51	3.00	1.00	1.00	1.00
52	1.50	0.75	1.38	0.75
53	6.75	2.25	2.25	2.25
54	4.50	0.63	0.88	0.63
55	1.89	2.33	1.56	1.56
56	2.14	0.57	0.57	0.57
Average	**3.47**	**1.64**	**1.97**	**1.60**

Table 22 External Speaker/Microphone Word Error Rates

5.3 3.5mm Cable Option

The accuracy rates of this experiment are only slightly better than those of the external speaker/microphone experiment and are listed in table 23. For the overall category, the best individual listener achieved an accuracy of 7%, which Distributed Listening exceeded at 9%.

	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
Overall	5%	7%	7%	9%
Overall Interpretation	7%	7%	7%	9%
Valid	7%	7%	4%	7%
Valid Interpretation	11%	7%	4%	7%
Invalid	4%	7%	11%	11%
Invalid Interpretation	4%	7%	11%	11%

Table 23 3.5mm Cable Option Accuracy Rates

The overall interpretation category, when compared to the overall category, showed an increase in just one of the listeners. Interestingly, listener 1 increased accuracy from 5% to 7%, yet the result from Distributed Listening did not reflect that increase. Although listener 1 had a semantic result that resolved to the same meaning as the actual spoken phase, listeners 2 and 3 agreed on aspects of their recognition results, which caused Distributed Listening to return a result that included the agreed upon bigrams, as shown in table 24. The semantic equivalents are indicated with underlined text in the table. Notwithstanding the round listed in table 24, Distributed Listening still had the best accuracy of this category at 9%, with the best individual listener achieving 7%.

Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening	Actual Spoken Phrase
authorities have been slow to address it	the authority has been slow to attract	the authority to go to	the authority has been slow to it	the authorities have been slow to address it

Table 24 3.5mm Cable Correct Semantic Interpretation of the Actual Spoken Phrase

The valid category had a best individual listener at 7% accuracy, which Distributed Listening met. The valid interpretations category showed an improvement over the valid category with listener 1 from 7% to 11%, yet Distributed Listening again did not reflect that increase. In this category, 11% accuracy is better than Distributed Listening at 7%. As shown previously in table 24, the semantic interpretation of listener 1 for one of the trials resolved to the correct meaning of the actual spoken phrase, but the agreement between listeners 2 and 3 on misrecognized bigrams caused Distributed Listening to return a result that included the agreed upon bigrams.

The last two categories of table 23, invalid and invalid interpretation, show that Distributed Listening was as good as the best individual listener. For both categories, the best individual listener achieved an accuracy of 11%, which was the same accuracy of Distributed Listening. The complete list of recognition results for this experiment is displayed in appendix C.

Distributed Listening had the lowest word error rate out of this dataset, at 1.55. Listeners 1, 2, and 3, had word error rates of 2.53, 1.70, and 1.57, respectively. Notice that the average WER of Distributed Listening is extremely close to that of Listener 3. Because of the variability of the recognition results of the individual listeners, this is expected. The WER is calculated based on the number of character insertions, deletions, and substitutions needed to change one string into a reference string. Given the variability of the recognition results, as Distributed Listening created candidate strings from recognitions that were overwhelmingly different from the actual spoken phrase, those candidate strings had a number of insertions, deletions, and substitutions. The complete list of word error rates is listed in table 25.

Phrase Number	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
1	0.00	0.00	0.00	0.00
2	2.83	0.83	0.75	0.75
3	2.38	1.00	1.38	1.38
4	0.00	0.00	0.00	0.00
5	4.25	2.58	0.50	0.50
6	1.22	2.89	0.33	0.33
7	2.13	1.83	1.00	1.00
8	1.96	1.79	1.42	1.29
9	1.39	1.22	1.11	1.22
10	1.08	2.92	0.50	0.50
11	1.83	1.75	0.00	0.00
12	1.67	1.22	1.89	1.22
13	3.22	2.67	3.44	3.44
14	2.38	3.00	2.62	2.62
15	1.08	2.38	0.31	0.31
16	2.10	1.20	0.80	0.80
17	3.52	0.71	0.86	0.71
18	2.00	1.00	0.50	0.50
19	1.60	0.40	0.00	0.40
20	2.54	0.75	1.63	0.75
21	2.00	1.00	2.00	1.00
22	0.00	0.00	0.44	0.00
23	3.90	0.30	1.00	0.30
24	1.38	2.13	3.88	5.13
25	1.00	1.89	0.67	5.78
26	3.06	1.06	0.28	3.06
27	3.33	0.83	1.58	0.83
28	2.38	1.05	1.19	2.10
29	2.23	0.00	0.69	0.00
30	2.00	2.39	1.17	1.17
31	2.09	0.82	1.09	1.09
32	2.87	1.60	2.67	1.60
33	2.88	1.40	2.04	1.40
34	4.00	2.71	1.93	1.93
35	2.78	1.11	1.22	1.22
36	2.10	2.50	2.10	2.50
37	4.29	2.00	3.00	3.00

Phrase Number	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening
38	3.07	1.47	2.53	1.47
39	4.07	2.29	2.00	2.00
40	3.38	2.92	3.54	3.38
41	4.67	3.00	2.25	2.25
42	3.15	1.08	1.69	1.08
43	4.30	4.00	3.20	3.20
44	3.47	1.16	1.11	1.16
45	3.13	1.43	2.17	1.26
46	4.22	3.83	1.96	1.96
47	3.89	2.50	2.94	2.50
48	5.00	4.00	3.27	3.27
49	3.88	2.75	3.38	2.75
50	2.81	3.06	2.44	2.44
51	3.75	1.25	1.25	1.25
52	2.25	1.88	0.75	0.75
53	2.25	2.25	2.50	2.50
54	0.50	1.50	3.25	1.50
55	1.89	1.56	1.33	1.44
56	0.57	0.57	0.57	0.57
Average	**2.53**	**1.70**	**1.57**	**1.55**

Table 25 3.5mm Cable Word Error Rates

5.4 n-Listeners Combinations

The previous research findings were calculated using 3 listeners, yet Distributed Listening, in theory, is capable of using any number of multiple listeners. Using that principle, it is noteworthy to present data analysis using a combination of the Distributed Listening results in additional multi-listeners environment and will be presented in the following three subsections. The baseline comparison for the data analysis will be the accuracy rates of the stereo mix experiment, as the accuracy rates from that experiment were the best.

5.4.1 Stereo/Speaker Combination

The most likely spoken phrases as returned by Distributed Listening from the 3-listeners experiments using the stereo mix option and the external speaker/microphone option were theoretically combined to simulate a 6-listener experiment. The results, per round of the 56-trials, were processed by calculating the cumulative local bigram frequency count of each phrase and selecting the phrase with the greatest total as the most likely spoken phrase. The individual bigram frequencies for this experiment were the same counts created from the 3-listeners experiments and the same categories used in the 3-listeners experiments were used in the resulting data analysis. The accuracy results of this experiment are shown in table 26. In the table, the column Stereo DL represents the most likely spoken phrases that were returned by Distributed Listening during the stereo mix experiment. The column Speaker DL represents the results of Distributed Listening from the external speaker/microphone experiment. The last column of the table corresponds to the results returned from this experiment.

	Stereo DL	Speaker DL	Distributed Listening
Overall	48%	7%	48%
Overall Interpretation	52%	7%	52%
Valid	54%	4%	54%
Valid Interpretation	57%	4%	57%
Invalid	43%	11%	43%
Invalid Interpretation	46%	11%	46%

Table 26 Stereo/Speaker Combination Recognition Accuracy Rates

Recall that the external speaker/microphone experiment had one round of trials that was inconclusive. In that case, the result from the stereo mix experiment was chosen by default.

The Distributed Listening results of this experiment remained consistent from the stereo mix experiment. For each category of the experiment, the accuracy did not change from the stereo mix experiment.

At first glance, there appears to be no merit to combining additional ASR system with regard to accuracy. In contrast, this experiment showed that combining ASR systems that have relatively good recognition results with systems that have extremely poor recognition results does not degrade the resulting accuracy. The external speaker/microphone experiment resulted in the poorest accuracy results, with one listener reporting an overall accuracy of 0%. In all likelihood, the accuracy could improve if all of the combined ASR systems have relatively good individual recognition results.

The conclusion of the WER calculations is similar to that of the accuracy results. As shown in table 27, the average WER remained the same when compared to the stereo mix experiment. This is definitely promising as the poor results from the external speaker/microphone experiment did not increase the WER.

Phrase Number	Stereo DL	Speaker DL	Distributed Listening
1	0.25	0.63	0.25
2	0.00	2.08	0.00
3	0.00	2.75	0.00
4	0.00	0.00	0.00
5	0.00	1.75	0.00
6	0.00	0.33	0.00
7	0.09	1.91	0.09
8	0.63	0.83	0.83

Phrase Number	Stereo DL	Speaker DL	Distributed Listening
9	0.33	0.78	0.33
10	0.00	0.00	0.00
11	0.00	0.83	0.00
12	0.22	1.44	0.22
13	0.44	3.00	0.44
14	0.54	0.92	0.54
15	0.00	0.46	0.00
16	0.40	0.80	0.40
17	0.00	2.33	0.00
18	0.20	1.05	0.20
19	0.00	0.00	0.00
20	0.08	1.96	0.08
21	0.40	3.00	0.40
22	0.00	0.44	0.00
23	0.00	1.50	0.00
24	0.00	0.00	0.00
25	0.56	0.78	0.56
26	0.00	0.39	0.00
27	0.75	1.75	0.75
28	0.00	2.10	0.00
29	0.00	0.62	0.00
30	0.06	3.06	0.06
31	0.55	1.18	0.55
32	0.33	2.33	0.33
33	0.00	inconclusive	0.00
34	0.14	2.64	0.14
35	0.00	1.22	0.00
36	0.00	0.70	0.00
37	0.00	1.71	0.00
38	0.00	1.87	0.00
39	0.79	2.07	0.79
40	0.00	2.23	0.00
41	0.00	3.42	0.00
42	0.15	2.08	0.15
43	1.70	3.50	1.70
44	0.16	1.05	0.16
45	0.35	2.17	0.35
46	1.09	3.09	1.09
47	1.67	2.94	1.67
48	0.18	5.18	0.18
49	0.50	1.63	0.50
50	0.94	2.56	0.94

Phrase Number	Stereo DL	Speaker DL	Distributed Listening
51	0.00	1.00	0.00
52	0.00	0.75	0.00
53	0.75	2.25	0.75
54	0.13	0.63	0.13
55	0.00	1.56	0.00
56	0.00	0.57	0.00
Average	**0.26**	**1.60**	**0.26**

Table 27 Stereo/Speaker Combination Word Error Rates

5.4.2 Stereo/Cable Combination

Like the stereo/speaker combination, the stereo/cable theoretical experiment combined the Distributed Listening results from the stereo mix and 3.5mm cable experiments and combined them to select the most likely spoken phrase in a 6-listener environment. The selection was based on the sum of the local bigram frequency counts, with the candidate phrase with the greatest sum returned as the most likely spoken phrase. The local bigram frequency counts are the same as those that were calculated during the respective 3-listerner experiments.

Refer to table 28 for a comprehensive view of the accuracy results. Columns Stereo DL and Cable DL of the table refer to the respective results of the stereo mix and 3.5mm cable experiments. The Distributed Listening column refers to the results of this experiment.

	Stereo DL	Cable DL	Distributed Listening
Overall	48%	9%	50%
Overall Interpretation	52%	9%	52%
Valid	54%	7%	57%
Valid Interpretation	57%	7%	57%
Invalid	43%	11%	43%
Invalid Interpretation	46%	11%	46%

Table 28 Stereo/Cable Combination Accuracy Rates

The results for both the overall and valid categories increased compared to the results from the stereo experiment. In both cases, this is due to the first round of recognitions (table 29). As more listeners correctly "heard" the actual spoken phrase, the Evidence-Based Phrase Resolution Algorithm correctly reconciled the recognitions.

L1	L2	L3	L4	L5	L6	DL	Actual Spoken Phrase
it's a disease that affects how you move	is a disease that affects how you move	it is a disease that affects how you move	it's a disease that affects how you move	it's a disease that affects how you move	it's a disease that affects how you move	it's a disease that affects how you move	it's a disease that affects how you move

Table 29 Stereo/Cable Combination Round 1

The remaining categories showed that Distributed Listening remained consistent when compared to the stereo mix experiment.

The results of the first round of trials are also reflected in the average WER (table 30). The average WER improved compared to the stereo mix experiment. The stereo mix experiment produced an average WER of .26, whereas Distributed Listening produced an average WER of .25.

Phrase Number	Stereo DL	Cable DL	Distributed Listening
1	0.25	0.00	0.00
2	0.00	0.75	0.00
3	0.00	1.38	0.00
4	0.00	0.00	0.00
5	0.00	0.50	0.00
6	0.00	0.33	0.00
7	0.09	1.00	0.09
8	0.63	1.29	0.63
9	0.33	1.22	0.33
10	0.00	0.50	0.00
11	0.00	0.00	0.00
12	0.22	1.22	0.22
13	0.44	3.44	0.44
14	0.54	2.62	0.54
15	0.00	0.31	0.00
16	0.40	0.80	0.40
17	0.00	0.71	0.00
18	0.20	0.50	0.20
19	0.00	0.40	0.00
20	0.08	0.75	0.08
21	0.40	1.00	0.40
22	0.00	0.00	0.00
23	0.00	0.30	0.00
24	0.00	5.13	0.00
25	0.56	5.78	0.56
26	0.00	3.06	0.00
27	0.75	0.83	0.83
28	0.00	2.10	0.00
29	0.00	0.00	0.00
30	0.06	1.17	0.06
31	0.55	1.09	0.55
32	0.33	1.60	0.33
33	0.00	1.40	0.00

Phrase Number	Stereo DL	Cable DL	Distributed Listening
34	0.14	1.93	0.14
35	0.00	1.22	0.00
36	0.00	2.50	0.00
37	0.00	3.00	0.00
38	0.00	1.47	0.00
39	0.79	2.00	0.79
40	0.00	3.38	0.00
41	0.00	2.25	0.00
42	0.15	1.08	0.15
43	1.70	3.20	1.70
44	0.16	1.16	0.16
45	0.35	1.26	0.35
46	1.09	1.96	1.09
47	1.67	2.50	1.67
48	0.18	3.27	0.18
49	0.50	2.75	0.50
50	0.94	2.44	0.94
51	0.00	1.25	0.00
52	0.00	0.75	0.00
53	0.75	2.50	0.75
54	0.13	1.50	0.13
55	0.00	1.44	0.00
56	0.00	0.57	0.00
Average	**0.26**	**1.55**	**0.25**

Table 30 Stereo/Cable Combination Word Error Rates

Due to the poor recognition results of both the external speaker/microphone and 3.5mm cable experiments, it was not meaningful to combine them into a theoretical experiment. Rather, it was more noteworthy to create a 9-listeners experiment as described next.

5.4.3 9-Listener Combination

The final theoretical experiment combined the Distributed Listening results from all of the 3-listeners experiments to simulate a 9-listeners environment. The sum of the local bigram frequency counts was used to determine the most likely spoken phrase in this experiment and the local frequency counts were the same as those calculated during the individual 3-listeners experiments. In the instance where a round from a 3-listeners setup was inconclusive, only the conclusive results from that round were used. The results from this experiment are listed in table 31. Columns Stereo DL, Speaker DL, and Cable DL represent the most likely spoken phrase results of the stereo mix, external speaker/.microphone, and 3.5mm cable experiments, respectively.

	Stereo DL	Speaker DL	Cable DL	Distributed Listening
Overall	48%	7%	9%	50%
Overall Interpretation	52%	7%	9%	52%
Valid	54%	4%	7%	57%
Valid Interpretation	57%	4%	7%	57%
Invalid	43%	11%	11%	43%
Invalid Interpretation	46%	11%	11%	46%

Table 31 Stereo/Speaker/Cable Combination Recognition Accuracy Rates

Like the 6-listener stereo/cable experiment, the 9-listener test resulted in an increase in the overall and valid categories compared to the stereo mix experiment for Distributed Listening and is likewise due to the first round of recognitions as shown in table 32. Within the overall category, the accuracy of Distributed Listening at 50% is better than the accuracy from the stereo mix experiment at 48%. Within the valid category, Distributed Listening exceeds the accuracy of the stereo mix experiment by 3%.

Recognizer	Recognition
Listener 1 – Satellite (Stereo)	it's a disease that affects how you move
Listener 2 – Dell (Stereo)	is a disease that affects how you move
Listener 3 – Tablet (Stereo)	it is a disease that affects how you move
Listener 4 – Satellite (Speaker)	the thing that affects how you
Listener 5 – Dell (Speaker)	it is believed that affect how you move
Listener 6 – Tablet (Speaker)	this disease that affects how you move
Listener 7 – Satellite(Cable)	it's a disease that affects how you move
Listener 8 – Dell (Cable)	it's a disease that affects how you move
Listener 9 – Tablet (Cable)	it's a disease that affects how you move
Distributed Listening	it's a disease that affects how you move
Actual Spoken Phrase	it's a disease that affects how you move

Table 32 Stereo/Speaker/Cable Combination Round 1

The remaining categories stayed consistent from the stereo mix experiment in that Distributed Listening matched the accuracy of the stereo mix experiment for those categories.

The average WER also remained consistent from the stereo mix experiment at .26, as shown in table 33.

Phrase Number	Stereo DL	Speaker DL	Cable DL	Distributed Listening
1	0.25	0.63	0.00	0.00
2	0.00	2.08	0.75	0.00
3	0.00	2.75	1.38	0.00
4	0.00	0.00	0.00	0.00
5	0.00	1.75	0.50	0.00
6	0.00	0.33	0.33	0.00
7	0.09	1.91	1.00	0.09
8	0.63	0.83	1.29	0.83
9	0.33	0.78	1.22	0.33
10	0.00	0.00	0.50	0.00
11	0.00	0.83	0.00	0.00
12	0.22	1.44	1.22	0.22
13	0.44	3.00	3.44	0.44

Phrase Number	Stereo DL	Speaker DL	Cable DL	Distributed Listening
14	0.54	0.92	2.62	0.54
15	0.00	0.46	0.31	0.00
16	0.40	0.80	0.80	0.40
17	0.00	2.33	0.71	0.00
18	0.20	1.05	0.50	0.20
19	0.00	0.00	0.40	0.00
20	0.08	1.96	0.75	0.08
21	0.40	3.00	1.00	0.40
22	0.00	0.44	0.00	0.00
23	0.00	1.50	0.30	0.00
24	0.00	0.00	5.13	0.00
25	0.56	0.78	5.78	0.56
26	0.00	0.39	3.06	0.00
27	0.75	1.75	0.83	0.83
28	0.00	2.10	2.10	0.00
29	0.00	0.62	0.00	0.00
30	0.06	3.06	1.17	0.06
31	0.55	1.18	1.09	0.55
32	0.33	2.33	1.60	0.33
33	0.00	inconclusive	1.40	0.00
34	0.14	2.64	1.93	0.14
35	0.00	1.22	1.22	0.00
36	0.00	0.70	2.50	0.00
37	0.00	1.71	3.00	0.00
38	0.00	1.87	1.47	0.00
39	0.79	2.07	2.00	0.79
40	0.00	2.23	3.38	0.00
41	0.00	3.42	2.25	0.00
42	0.15	2.08	1.08	0.15
43	1.70	3.50	3.20	1.70
44	0.16	1.05	1.16	0.16
45	0.35	2.17	1.26	0.35
46	1.09	3.09	1.96	1.09
47	1.67	2.94	2.50	1.67
48	0.18	5.18	3.27	0.18
49	0.50	1.63	2.75	0.50
50	0.94	2.56	2.44	0.94
51	0.00	1.00	1.25	0.00
52	0.00	0.75	0.75	0.00
53	0.75	2.25	2.50	0.75
54	0.13	0.63	1.50	0.13
55	0.00	1.56	1.44	0.00

Phrase Number	Stereo DL	Speaker DL	Cable DL	Distributed Listening
56	0.00	0.57	0.57	0.00
Average	**0.26**	**1.60**	**1.55**	**0.26**

Table 33 Stereo/Speaker/Cable Combination Word Error Rates

The preceding experiments individually produced optimal results, and a collective, comparative analysis as set forth in the following section will further display the effectiveness of the Distributed Listening system.

5.5 Discussion

Each of the three physical experiments and three theoretical experiments produced excellent results. Distributed Listening as a composite system consistently met or out-performed the results from the individual listeners.

Overall, for all six experiments, Distributed Listening never performed worse than the best individual listener and in fact, exceeded the accuracy of the best individual listener in the 3.5mm cable experiment and exceeded the baseline metric in the stereo/cable and stereo/speaker/cable experiments within the overall category.

Distributed Listening exceeded the best individual listener within the overall interpretation category for both the stereo mix and 3.5mm cable experiments and matched the best individual listener and the baseline metric for the remaining experiments.

Within the valid category, Distributed Listening exceeded the best individual listener for the stereo mix experiment and exceeded the baseline metric for both the stereo/cable and stereo/speaker/cable experiments. The remaining experiments showed that Distributed Listening was no worse than the best reported accuracy.

The valid interpretation category for the six experiments had the most inconsistent results. While Distributed Listening exceeded the best individual listener of the stereo mix experiment, the accuracy of Distributed Listening for the 3.5mm cable experiment was worse than the best individual listener. This is surprising since it is expected that adding in semantic equivalents will at a minimum not decrease accuracy, and at best, improve accuracy. That was the case for the stereo mix experiment, but not the 3.5mm experiment. As discussed in section 5.3 and shown in table 24, this unexpected result is attributed to the agreement between listeners on misrecognized bigrams.

The invalid category across all six experiments presented interesting and unexpected results from Distributed Listening. In five of the six experiments, Distributed Listening returned results that were only as good as the best individual listener. The remaining experiment (stereo mix) showed that Distributed Listening was worse than the best individual listener. Distributed Listening did not exceed the best individual listener during any of the tests for the invalid category. The same can be said for the invalid interpretation category. Not only did Distributed Listening not exceed an individual listener, Distributed Listening was worse than the best during the stereo mix experiment. This can be attributed to the contents of the corpus. Recall that the results in the invalid category are the ensuing recognitions from the actual spoken phrases that do not have bigrams that are maintained in the corpus. Distributed Listening builds the most likely spoken phrase based on combining and concatenating bigrams. If those bigrams are not in the corpus or the original evidence, the phrases that contain those words do not pass validation. Using the corpus as both a validation tool and as a mechanism to break ties requires that the corpus contains words relevant to the domain to have optimal results.

With regard to the WER calculations, Distributed Listening again reported excellent results. For each of the three physical experiments, Distributed Listening had the lowest average WER. For each of the three theoretical experiments, Distributed Listening was as good as the baseline metric.

When the results of Distributed Listening are taken as a whole, including the poor results from the invalid and invalid interpretation categories, Distributed Listening completely conforms to the theoretical expectations that were established at the start of the research task. The resulting data analysis from the experiments fully supports the hypothesis and establishes that Distributed Listening is a practical composite architecture of distributed ASR systems.

6 Conclusion

Distributed Listening is a novel approach to improving the accuracy of spoken language systems by using a distributed approach through the use of multiple, yet independent, recognizers with distinct and separate input sources. Through the exhaustive experimentation phase, it was demonstrated that Distributed Listening is indeed a viable solution that complements existing technologies and therefore, adds significant contributions to the area of spoken language systems, as described in the next section.

6.1 Contributions

The Distributed Listening research project first addressed the details of the input source of the spoken utterances, one of the common implementation details of systems that use multiple speech recognizers. Previous research activities used one input source that split the speech signal across the multiple speech recognizers. Distributed Listening provides each recognizer with its own independent and physically separate input source, to address this limitation.

An additional component of Distributed Listening that adds to the field is the inherent ability to simulate the process of human hearing and deduction. Recall from previous discussions that Distributed Listening is analogous to the deductive reasoning of detectives in combination with the psychological process of dichotic listening. Further developments that mimic that way people process speech will add to the technological advancements of spoken language systems.

Lastly, Distributed Listening addresses the need for speaker dependant systems. While systems that are trained will always produce, on average, better accuracy results than untrained, speaker independent systems, this research showed that reasonable recognition rates are possible when systems have not been trained for specific individuals.

6.2 Directions for Future Research

There are several areas within Distributed Listening that can benefit from further investigation. As the research progressed, several promising ideas were developed that warrant exploration. This section will explain those ideas.

6.2.1 Architectures

There is reason to look at alternative architecture models for Distributed Listening. There are three models that came forth during this project. The first of which is the homogeneous model.

The homogeneous model uses the same grammar or language model for each listener. Although all of the listeners are identical in capturing the input, this architecture allows for the different perspectives of the utterances to also be captured. The research presented here follows this model, yet there is a need to compare this model with the other two models.

The second proposed architecture is the heterogeneous model. Within the heterogeneous model, each listener uses a different grammar or language model. Each listener will keep its own input source and produce a recognition result. This model implies a distributed grammar/language model and allows for flexibility as very large grammars and vocabularies can be distributed across several listeners.

The final proposed architecture is the hybrid model, which contains a homogeneous architecture of heterogeneous distributed listening nodes. The hybrid architecture, as shown in figure 35, gives the embedded environment the ability to recognize multiple languages, as well as accommodate translations of inter-mixed spoken language.

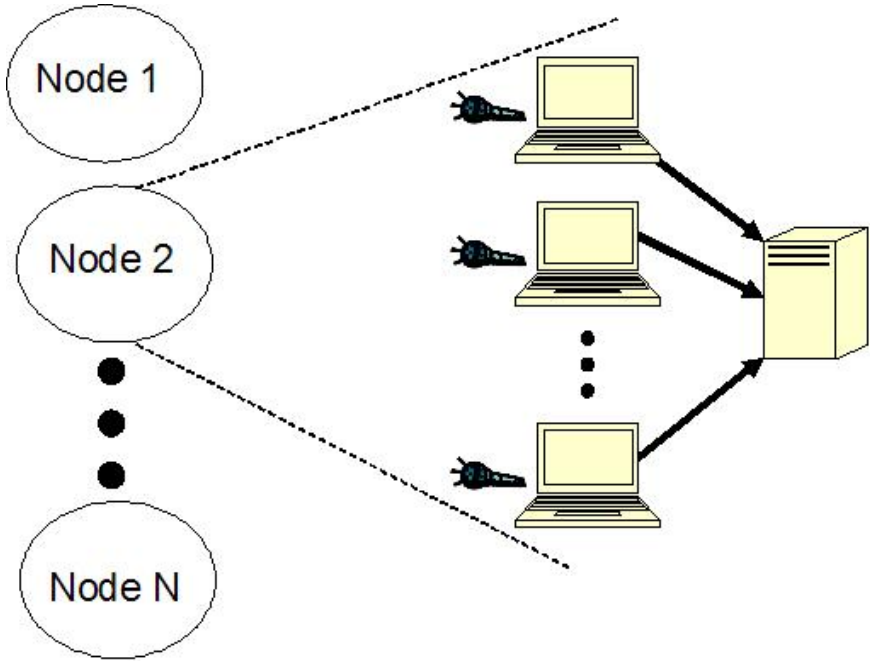


Figure 35 Hybrid Distributed Listening Architecture

The heterogeneous and hybrid models each add another dimension to the Distributed Listening system and a comparative analysis of the three models should show continued recognition accuracy improvements with ASR systems that use multiple recognizers.

In addition to the proposed models, a parallel listening architecture warrants further discussion as described next.

6.2.2 Parallel Processing

The motivation for this research was to simulate the way in which people hear and process speech with multiple independent listeners. An applicable adaptation of that motivation would be to provide an alternative solution for those students who have hearing impairments in the post-secondary classroom. Within the classroom, the increase in speech recognition accuracy will make the communication between hearing impaired students and instructors more effective by improving the automatic transcription of the spoken lecture, especially in a post-secondary school environment. Currently, students with hearing impairments face a number of educational obstacles that can be exacerbated when the transcription results of a lecture are less than optimal. For instance, hearing impaired students have a learning curve that stems from the translation of the sentence structure of written English words to the established syntax of American Sign Language. Unfortunately for these students, American Sign Language does not follow the grammar rules and sentence structures of the English language (Liddell 1980). Additionally, inaccuracies in the transcription result can directly impact the educational future of the student. Wrong or missing information is time consuming in that the student would spend significant time re-learning information. Post-secondary education for hearing impaired students faces different challenges then secondary education as it is increasingly difficult to keep up with language that is usually more technical and rapid, which results in more note taking. For students who have been able to cope by relying on lipreading, it is now impossible to lipread and take notes simultaneously (Reed 1984). The success of Distributed Listening has the ability to directly impact the success of hearing impaired students in an educational environment.

To effectively evaluate Distributed Listening in that environment, a real-time parallel processing scheme is needed. The results presented in this project stemmed from parallel listening through the programming language. An extension of this work that Distributed Listening will further benefit from is to process the interpretations in a real-time parallel processing listening environment.

6.3 Summary

Distributed Listening was created with a hypothesis that the system would perform at worst, as good as the best individual listener. The overall interpretation and valid interpretation categories showed the best results for Distributed Listening, which is in keeping with the underlying theme of this research. The motivation of this research was to mimic the way in which humans process speech. When several people are privy to the same spoken input and you rely on those people to repeat what was heard, the versions that are relayed frequently contain **interpretations** of what was actually said and are prone to inherent errors. The results are only as accurate as the people relaying the information. Adding the computer component eliminates the inconsistencies inherent to people, thereby increasing the effectiveness of the results. The overall interpretation and valid interpretation accuracy results reconciled to the correct meaning of the actual spoken phrases. When the correct **meaning** is accurately conveyed, this mimics the behavior of people.

Once the experiment was complete, the data analysis showed that Distributed Listening performs better with regard to speech recognition than an individual ASR system. Generally, the recognition rates of Distributed Listening, when compared with the recognition rates of the individual standard recognition systems, met or out-performed the individual systems.

The experiments that were executed supported the premise and hypothesis of Distributed Listening and confirm that Distributed Listening is a viable alternative to existing methods of ASR systems that use multiple recognizers.

References

1. Barry, T., Solz, T., Reising, J. and Williamson, D. **The simultaneous use of three machine speech recognition systems to increase recognition accuracy**, In Proceedings of the IEEE 1994 National Aerospace and Electronics Conference, vol.2, pp. 667 - 671, 1994.

2. Baum, L.E. **An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov process**. Inequalities 3, 1-8, 1972.

3. Bruder, G.E., Stewart, J.W., McGrath, P.J., Deliyannides, D., Quitkin, F.M. **Dichotic listening tests of functional brain asymmetry predict response to fluoxetine in depressed women and men**. Neuropsychopharmacology, 29(9), pp. 1752-1761, 2004.

4. Brutti, A., Coletti, P., Cristoforetti, L., Geutner, P., Giacomini, A., Gretter, R., Maistrello, M., Matassoni, M., Omologo, M., Steffens, F. and Svaizer, P., **Use of Multiple Speech Recognition Units in a In-car Assistance Systems**, chapter in "DSP for Vehicle and Mobile Systems", Kluwer Publishers, 2004.

5. Cristoforetti, L., Matassoni, M., Omologo, M. and Svaizer, P., **Use of parallel recognizers for robust in-car speech interaction**, In Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing [ICASSP 2003], Hong-Kong, 2003.

6. Deng, L. and Huang, X., **Challenges in adopting speech recognition**, Communications of the ACM, vol. 47, no. 1, pp. 69-75, January 2004.

7. Duchnowski, P. **A New Structure for Automatic Speech Recognition**. Diss. Massachusetts Institute of Technology, 1993.

8. Durrant, J. and Lovrinic, J., **Bases of Hearing Science**, Williams & Wilkins, 1995.

9. Evermann, G. and Woodland, P., **Posterior Probability Decoding, Confidence Estimation and System Combination**, In Proceedings of NIST Speech Transcription Workshop, 2000.

10. Fiscus, J. G., **A post-processing system to yield reduced error word rates: Recognizer output voting error reduction (ROVER)**. In IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 347–354, 1997.

11. Furui, S., **Digital Speech, Processing, Synthesis, and Recognition,** Marcel Dekker, Inc., 1989.Furui, S., **Recent progress in spontaneous speech recognition and understanding**, In Proceedings of the IEEE Workshop on Multimedia Signal Processing, 2002.

12. Gilbert, J.E. and Zhong, Y., **Speech User Interfaces for Information Retrieval**, In Proceedings of 12th Annual ACM Conference on Information & Knowledge Management, New Orleans, Louisiana, pp. 77-82, 2003.

13. Goel, V., Kumar, S. and Byrne, **Segmental Minimum Bayes-Risk ASR Voting Strategies**, In Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP), Beijing, China, 2000, pp 139-142.

14. Ide, Nancy, and Suderman, Keith (2007). **The Open American National Corpus (OANC)**. http://www.AmericanNationalCorpus.org/OANC.

15. Jurafsky, D. and Martin, J., **Speech and Language Processing**, Prentice Hall, 2000.

16. Kodama, Y., Utsuro, T., Nishizaki, H., Nakagawa, S., **Experimental Evaluation on Confidence of Agreement among Multiple Japanese LVCSR Models**, In Proceedings of the 7th European Conference on Speech Communication and Technology, Aalborg, Denmark, 2001, pp 2549—2552.

17. Liddell, S., **American Sign Language Syntax**, Mouton Publishers, 1980.

18. McCowan, I., Moore, D., Dines, J., Gatica-Perez, D., Flynn, M., Wellner, P. and Bourlard, H., **On the Use of Information Retrieval Measures for Speech Recognition Evaluation**, Idiap Publications, 2005.

19. MP3Gain, [Online]. Available: http://mp3gain.sourceforge.net/, 2010.

20. Natural Language Software Registry, [Online]. Available: http://registry.dfki.de/, 2004.

21. Pierce, A., **Acoustics**, McGraw-Hill, 1981.

22. Price, P., **The Growing Impact of Speech Technology on Society**, In AAAS Session Presentation on Language Processing for Science and Society, San Diego, CA, 2010.

23. Reed, M., **Educating Hearing-Impaired Children**, Open University Press, 1984.

24. Schwenk, H. and Gauvain, J., **Combining Multiple Speech Recognizers using Voting and Language Model Information**, In Proceedings of the IEEE International Conference on Speech and Language Processing (ICSLP), Pekin, pp. II:915–918, 2000.

25. Solsona, R. A., Fosler-lussier, E., Kuo, H., Potamianos, A., Zitouni, I., **Adaptive Language Models for Spoken Dialogue Systems**, In Proceedings of the ICASSP, 2002.

26. Young, S.R., Hauptmann, A.G. , Ward, W.H. , Smith, E.T. and Werner, P., **High level knowledge sources in usable speech recognition systems**, Communications of the ACM, vol. 32, no. 2, pp. 183-194, 1989.

27. Young, S.R., Use **of dialog, pragmatics and semantics to enhance speech recognition**, Speech Communication, vol. 9, pp. 551-564, 1990.

Appendix A: Recognition Engines Results – Stereo Mix

Shading indicates a valid actual spoken phrase, meaning each bigram of the phrase is found within the corpus.

	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening Result	Actual Spoken Phrase
1.	it's a disease that affects how you move	is a disease that affects how you move	it is a disease that affects how you move	it is a disease that affects how you move	it is a disease that affects how you move
2.	a finger that were goals and you can't really get it to stop	a finger that wiggles and you can't really get it to stop	a finger wiggles and you can't really get it to stop	a finger that wiggles and you can't really get it to stop	a finger that wiggles and you can't really get it to stop
3.	an arm doesn't swing quite the same but	an arm doesn't swing quite the same way	an arm doesn't swing quite the same way	an arm doesn't swing quite the same way	an arm doesn't swing quite the same way
4.	your handwriting has changed	your handwriting has changed	your handwriting has changed	your handwriting has changed	your handwriting has changed
5.	is well in motion before any of these symptoms first become clear	is well in motion before any of these symptoms first become clear	is well in motion before any of these symptoms first become clear	is well in motion before any of these symptoms first become clear	is well in motion before any of these symptoms first become clear
6.	a lot has happened before you first notice	a lot has happened before you first notice it	a lot has happened before you first notice it	a lot has happened before you first notice it	a lot has happened before you first notice it

7.	the first thing i remember noticing process on closing tingling sensation in my left leg and to some extent my left arm	the first thing i remember noticing was this on causing tingling sensation in my left leg and to some extent my left arm	the first thing i remember noticing was this on buzzing tingling sensation in my left leg and to some extent my left arm	the first thing i remember noticing was this on buzzing tingling sensation in my left leg and to some extent my left arm	the first thing i remember noticing was this odd buzzing tingling sensation in my left leg and to some extent my left arm
8.	used to feel that my cell phone was vibrating a reach for a 10 minute find that there was nothing there	are used to feel that my cell phone is vibrating i'd reach for the men i'd find that there was nothing there	are used to feel that my cell phone was vibrating i'd reach for the and i'd find that there was nothing there	are used to feel that my cell phone was vibrating i'd reach for the and i'd find that there was nothing there	i used to feel that my cell phone was vibrating and i'd reach for it and then i'd find that there was nothing there
9.	i noticed that i didn't think that my arm was swinging quite the same way when i draw	i noticed that i didn't think that my arm was swinging quite the same way when i draw	i noticed that i didn't think that my arm was swinging quite the same way when i draw	i noticed that i didn't think that my arm was swinging quite the same way when i draw	i noticed that i didn't think that my arm was swinging quite the same way when i jogged
10.	the only reason i really did was because of my family history	the only reason i really did was because of my family history	the only reason i really did was because of my family history	the only reason i really did was because of my family history	the only reason i really did was because of my family history
11.	and that was the prevailing wisdom for a very very long time	and that was the prevailing wisdom for a very very long time	and that was the prevailing wisdom for a very very long time	and that was the prevailing wisdom for a very very long time	and that was the prevailing wisdom for a very very long time
12.	to stop the disease before it even starts	and you stop the disease before it even starts	and you stop the disease before it even starts	and you stop the disease before it even starts	and you'd stop the disease before it even starts

13.	with stem cell research figure into the genetic connection	with stem cell research figure into the genetic connection	with stem cell research figure into the genetic connection	with stem cell research figure into the genetic connection	would stem cell research figure into the genetic connection
14.	but the more we learn about the disease more complicated it becomes	to the more we learn about the disease more complicated it becomes	the more we learn about the disease more complicated it becomes	to the more we learn about the disease more complicated it becomes	but the more we learn about the disease the more complicated it becomes
15.	in lots of ways it made the research task that much more complex	in lots of ways it made the research task that much more complex	in lots of ways it made the research task that much more complex	in lots of ways it made the research task that much more complex	in lots of ways it made the research task that much more complex
16.	and we also saw the number of researchers left	and we also saw the number for searchers left	and we also saw the number of researchers left	and we also saw the number of researchers left	and we also saw that a number of researchers left
17.	they went to other places that were more open to stem cell research whether that was in europe or in singapore	they went to other places that were more open to some sort research whether that was in europe or in singapore	they went to other places that were more open to stem cell research whether that was in europe or in singapore	they went to other places that were more open to stem cell research whether that was in europe or in singapore	they went to other places that were more open to stem cell research whether that was in europe or in singapore
18.	but i think it's pretty clear that it had a negative effect on the way in which the fuel progressed	but i think it's pretty clear that it had a negative effect on the way in which the feel progress	but i think it's pretty clear that it had a negative effect on the way in which the feel progress	but i think it's pretty clear that it had a negative effect on the way in which the feel progress	but i think it's pretty clear that it had a negative effect on the way in which the field progressed

19.	i also understand the moral objections that some people have	i also understand the moral objections that some people have	i also understand the moral objections that some people have	i also understand the moral objections that some people have	i also understand the moral objections that some people have
20.	for a lot of people it's a dilemma because your faith might be telling you one thing in your body is telling you another	for a lot of people it's a dilemma because your faith might be telling you one thing in your body is telling you another	for a lot of people it's a dilemma because your faith might be telling you one thing in your body is telling you another	for a lot of people it's a dilemma because your faith might be telling you one thing in your body is telling you another	for a lot of people it's a dilemma because your faith might be telling you one thing and your body is telling you another
21.	but he was young and this happened seemingly overnight	but he was young and this happened seemingly overnight	but he was young and this happened seemingly overnight	but he was young and this happened seemingly overnight	but he was young and this had happened seemingly overnight
22.	and it does appear that there is a relationship	and it does appear that there is a relationship	and it does appear that there is a relationship	and it does appear that there is a relationship	and it does appear that there is a relationship
23.	genetics load the gun and the environment pulls the trigger	genetics load the gun and the environment pulls the trigger	genetics load the gun and the environment pulls the trigger	genetics load the gun and the environment pulls the trigger	genetics load the gun and the environment pulls the trigger
24.	there are medications that make a real difference	there are medications that make a real difference	there are medications that make a real difference	there are medications that make a real difference	there are medications that make a real difference
25.	that helps with some problems are not with others	that helps with some problems are not with others	that helps with some problems are not with others	that helps with some problems are not with others	it helps with some problems and not with others

26.	i guess i choose to believe that i'll be able to do this for a very long time	i guess i choose to believe that i'll be able to do this for a very long time	i guess i choose to believe that i'll be able to do this for a very long time	i guess i choose to believe that i'll be able to do this for a very long time	i guess i choose to believe that i'll be able to do this for a very long time
27.	but don't necessarily need or want to know more than	but i don't necessarily need or want to know more than	ago necessarily need or want to know more than	but i don't necessarily need or want to know more than	but they don't necessarily need or want to know more than that
28.	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered
29.	and now it's nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language feature
30.	the class as a semi improvised look inside a high school in a diverse working class paris neighborhood	the class as a semi improvised look inside a high school in a diverse working class paris neighborhood	the class as a semi improvised look inside a high school in a diverse working class paris neighborhood	the class as a semi improvised look inside a high school in a diverse working class paris neighborhood	the class is a semi improvised look inside a high school in a diverse working class paris neighborhood
31.	and they all take a deep breath and entered the arena	and they all take a deep breath and entered the arena	and they all take a deep breath and entered the arena	and they all take a deep breath and entered the arena	then they all take a deep breath and enter the arena

32.	he writes on the blackboard and a student makes them stop and define the word	he writes on the blackboard and a student makes them stop and define the word	he writes on the blackboard and a student makes them stop and define the word	he writes on the blackboard and a student makes them stop and define the word	he writes on the blackboard and a student makes him stop and define a word
33.	is so the teacher has to set aside his plan to save first what would be wrong with that and they know it isn't true	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true
34.	by the lesson has been derailed and there were snickers all around	and by then the lesson has been derailed and there were snickers all around	and by then the lesson has been derailed and there were snickers all around	and by then the lesson has been derailed and there were snickers all around	and by then the lesson has been derailed and there are snickers all around
35.	these kids don't deserve to be educated basic	these kids don't deserve to be educated they say	these kids don't deserve to be educated these to	these kids don't deserve to be educated they say	these kids don't deserve to be educated they say
36.	let them rot in their dead end world class jobs	let them rot in their dead end low class jobs	let them rot in their dead end low class jobs	let them rot in their dead end low class jobs	let them rot in their dead end low class jobs
37.	how long can he maintain his equilibrium	how long can he maintain his equilibrium	how long can he maintain his equilibrium	how long can he maintain his equilibrium	how long can he maintain his equilibrium
38.	he tells his colleagues that it's the job of the teacher to bring kids out	he tells his colleagues that it's the job of the teacher to bring kids out	he tells his colleagues that's the job of the teacher to bring kids out	he tells his colleagues that it's the job of the teacher to bring kids out	he tells his colleagues that it's the job of the teacher to bring kids out

39.	what rouses most of the students is an assignment to write self portraits	life rouses most of the students is an assignment to write self portraits	what rouses most of the students is an assignment to write self portraits	what rouses most of the students is an assignment to write self portraits	what finally rouses most of his students is an assignment to write self portraits
40.	for a brief spell they see him younger more open and ready to wear	for a brief spell they seem younger more open and ready to learn	for a brief spell they seem younger more open and ready to learn	for a brief spell they seem younger more open and ready to learn	for a brief spell they seem younger more open and ready to learn
41.	but just when you're getting a warm utopian feeling something bad happens	but just when you're getting a warm utopian feeling something bad happens	but just when you're getting a warm utopian feeling something bad happens	but just when you're getting a warm utopian feeling something bad happens	but just when you're getting a warm utopian feeling something bad happens
42.	i'll save the hero of the movie threatens to become its bad guy	i'll save a hero of the movie threatens to become its bad guy	i'll save the hero of the movie threatens to become its bad guy	i'll save the hero of the movie threatens to become its bad guy	i'll say the hero of the movie threatens to become its bad guy
43.	that's what he does with real time classroom scenes are so strong that	that's what he does with real time classroom scenes are so startling	that's what he does with real time classroom scenes are so start what	that's what he does with real time classroom scenes are so start what	that's why those real time classroom scenes are so startling
44.	they show you that at least until the system can be changed but battles will be moment to moment	they show you that at least until the system can be changed the battles will be moment to moment	they show you that at least until the system can be changed but battles will be moment to moment	they show you that at least until the system can be changed but battles will be moment to moment	they show you that at least until the system can be changed the battles will be moment to moment

45.	republicans disagreed complaining that the bill's tax cuts fall short and that spends too much on things they say will create jobs	republicans disagreed complaining that the bill's tax cuts fall short and that spends too much on things they say will create jobs	republicans disagreed complaining that the bill's tax cuts fall short and that it spends too much on things they say will create jobs	republicans disagreed complaining that the bill's tax cuts fall short and that spends too much on things they say will create jobs	republicans disagree complaining that the bill's tax cuts fall short and that it spends too much on things they say won't create jobs
46.	shows both product lines fell more than 20 percent last year as fewer people but tories leading up to the holiday season	skills in both product lines fell more than 20 percent last year as fewer people bought toys leading up to the holiday season	seals the both product lines fell more than 20 percent last year as fewer people but tories leading up to the holiday season	seals the both product lines fell more than 20 percent last year as fewer people but tories leading up to the holiday season	sales of both product lines fell more than twenty percent last year as fewer people bought toys leading up to the holiday season
47.	toymaker v los angeles said in november it would shut about a thousand jobs	toy maker based in los angeles said in november it would shut about a thousand jobs	toymaker v los angeles said in november it would shut about a thousand jobs	toymaker v los angeles said in november it would shut about a thousand jobs	the toy maker based here in los angeles said in november it would shed about a thousand jobs
48.	in that it wasn't immune from last year's deteriorating economic environment	admitted wasn't immune from last year's deteriorating economic environment	embedded wasn't immune from last year's deteriorating economic environment	in that it wasn't immune from last year's deteriorating economic environment	and that it wasn't immune from last year's deteriorating economic environment
49.	spread to cut taxes that is the question	spirit and cut taxes that is the question	sprint or cut taxes that is the question	spread to cut taxes that is the question	spend or cut taxes that is the question

50.	slips into berating down and just about every sector and every zip code	thinks lip synch derailing down and just about every sector and every zip code	slips into berating down and just about every sector and every zip code	slips into berating down and just about every sector and every zip code	pink slips seem to be raining down in just about every sector and every zip code
51.	but are there any bright spots out there	but are there any bright spots out there	but are there any bright spots up in	but are there any bright spots out there	but are there any bright spots out there
52.	the temporary job market may provide an answer	the temporary job market may provide an answer	the temporary job market may provide an answer	the temporary job market may provide an answer	the temporary job market may provide an answer
53.	to become increasingly physical	to become increasingly difficult	to become increasingly physical	to become increasingly physical	it's become increasingly physical
54.	the authorities have been slow to address at	authorities have been slow to address at	authorities have been slow to address	the authorities have been slow to address at	the authorities have been slow to address it
55.	an egyptian court convicted a man of sexual harassment	an egyptian court convicted a man of sexual harassment	an egyptian court convicted a man of sexual harassment	an egyptian court convicted a man of sexual harassment	an egyptian court convicted a man of sexual harassment
56.	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful now	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful now	women's rights advocates are cautiously hopeful now

Appendix B: Recognition Engines Results – External Speaker/Microphone

Shading indicates a valid actual spoken phrase, meaning each bigram of the phrase is found within the corpus.

	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening Result	Actual Spoken Phrase
1.	the thing that affects how you	it is believed that affect how you move	this disease that affects how you move	this disease that affects how you move	it's a disease that affects how you move
2.	it is illegal to give up	a dearth of legal and you can't really do dot	over the legal can't really get it	a dearth of legal and you can't really do dot	a finger that wiggles and you can't really get it to stop
3.	on the going quite	a form that is going quite in the	an armed point 5	a form that is going quite in the	an arm doesn't swing quite the same way
4.	your handwriting and	your handwriting has changed	your handwriting has changed	your handwriting has changed	your handwriting has changed
5.	well in motion with his and	a rolling motion before any of the will become clear	no one motion before any clear	a rolling motion before any of the will become clear	is well in motion before any of these symptoms first become clear
6.	a lot has we are	a lot has happened before you first notice	a lot has before you first note	a lot has happened before you first notice	a lot has happened before you first notice it

7.	the thing i remember is a leg	the first thing i remember noticing what was on london england vision in my left leg of my left arm	the first thing i remember noticing what one thing dangling from fiction in my left leg in a pilot are	the first thing i remember noticing what one thing dangling from fiction in my left leg in a pilot are	the first thing i remember noticing was this odd buzzing tingling sensation in my left leg and to some extent my left arm
8.	i used to feel that myself and a	i used to feel that my cell phone is vibrating a week or another i'm not there	i used to feel that my cell phone is vibrating a reach for a do not find that there was nothing there	i used to feel that my cell phone is vibrating a reach for a do not find that there was nothing there	i used to feel that my cell phone was vibrating and i'd reach for it and then i'd find that there was nothing there
9.	i noticed that my mom will	i noticed that i didn't think that my arm was in quite the same way when i draw	i noticed that i didn't think that my mom in quite the way i draw	i noticed that i didn't think that my mom was in quite the same way when i draw	i noticed that i didn't think that my arm was swinging quite the same way when i jogged
10.	only reason i'm	the only reason i really did was because of my family history	the only reason i really didn't think about my family history	the only reason i really did was because of my family history	the only reason i really did was because of my family history
11.	not with them is	and that was the prevailing wisdom works very very well	and that was the prevailing wisdom for a very very	and that was the prevailing wisdom for a very very	and that was the prevailing wisdom for a very very long time
12.	ethos of the for the	can you spot the disease before they can start	and you thought the defeat before the start	can you spot the disease before they can start	and you'd stop the disease before it even starts

13.	and so it is	is a well researched figure in the genetic	with stem cell research figure in a way	with stem cell research figure in a way	would stem cell research figure into the genetic connection
14.	more we learn about	the more we learn about the disease more complicated become	the more we learn about the need for complicated	the more we learn about the disease more complicated become	but the more we learn about the disease the more complicated it becomes
15.	lots of ways we are	in lots of ways and means the research task that much more complex	in lots of ways to leave the research task that much more	in lots of ways and means the research task that much more complex	in lots of ways it made the research task that much more complex
16.	the office of	and we also saw the number of lot	and we also saw the number of researchers are	and we also saw the number of researchers are	and we also saw that a number of researchers left
17.	little is known as	they went to the place of the more open about what is in europe or in	they went to the place of a more open system without we in europe will report	they went to the place of a more open system without we in europe will report	they went to other places that were more open to stem cell research whether that was in europe or in singapore
18.	i think it's pretty neat that it had a negative effect on the way in which	but i think it's pretty clear that it had a negative impact on the way in which	but i think it's pretty clear that it had a negative effect on the way in which	but i think it's pretty clear that it had a negative effect on the way in which	but i think it's pretty clear that it had a negative effect on the way in which the field progressed

19.	i also understand	i also understand the moral objections that some people have	i also understand the moral objection to come people have	i also understand the moral objections that some people have	i also understand the moral objections that some people have
20.	a lot of people he is not	a lot of people it was the one because and let me tell you one thing in your body when you are	a lot of people think of when you think might be going with you and your body will not	a lot of people it was the one because and let me tell you one thing in your body when you are	for a lot of people it's a dilemma because your faith might be telling you one thing and your body is telling you another
21.	yeah and if it happened overnight	jan and i are not	at the end of happen overnight	yeah and if it happened overnight	but he was young and this had happened seemingly overnight
22.	it does appear as if there is a relationship	it does appear that there is a relationship	it does appear that there is a relationship	it does appear that there is a relationship	and it does appear that there is a relationship
23.	you know the guys in the environmental trigger	genetic flow the gun and the environment	kinetic flow the gun and the environmental trigger	kinetic flow the gun and the environmental trigger	genetics load the gun and the environment pulls the trigger
24.	there are	there are medications that make a real difference	there are medications that make up the	there are medications that make a real difference	there are medications that make a real difference
25.	some problems with others	it helped him problems are not with others	it helps that some problems are not with other	it helps that some problems are not with other	it helps with some problems and not with others

26.	can you believe it is a very	i guess i choose to believe that i'll be able to do this for very long	i guess i choose to believe that i'll be able to do that for her	i guess i choose to believe that i'll be able to do this for very long	i guess i choose to believe that i'll be able to do this for a very long time
27.	there is really no	there surely need or want more	without necessarily need or want to know what	without necessarily need or want to know what	but they don't necessarily need or want to know more than that
28.	been in declining health here in the long term effects of	given the declining health of the year dealing with the long term effect will	he had been in declining health for the past year in the long term effects of stroke	he in declining health of the year long with year in the long term effects of stroke	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered
29.	nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language	and that's nominated for an academy award at our language	and now it's nominated for an academy award for best foreign language	and now it's nominated for an academy award for best foreign language feature
30.	if you improvise inside the high school and working class and	the class and the improvised look inside the high school in the diverse working class parent that	a classic family and provide for the entire high school were working class neighborhood	the you and inside the high school and high class and the diverse working class parent that	the class is a semi improvised look inside a high school in a diverse working class paris neighborhood
31.	you think is the envy of the	they all take a deep breath and interviewing	and they all take a deep breath and interviewing	and they all take a deep breath and interviewing	then they all take a deep breath and enter the arena

32.	month on the blackboard is often defined as	work on the blackboard at make them stop and the following were	from the blackboard at length album to follow the word	work on the blackboard at make them stop and the following were	he writes on the blackboard and a student makes him stop and define a word
33.	though the feature set aside as he was the one that has been known	as for the picture is satisfied when they first what would be wrong with that and then know if	as for the future have to set aside when they are well with you will not admit to it	inconclusive	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true
34.	and by the less than an in your microphone	why the left has been rao in your knickers all well	by then the lesson of thin radio in your personal	and by the has been an in your knickers all	and by then the lesson has been derailed and there are snickers all around
35.	if he is	these kids don't deserve to be educated they fit	if kids don't deserve to be addicted to	if kids don't deserve to be educated to fit	these kids don't deserve to be educated they say
36.	let them log in her job	let them live in their dead end low cost jobs	let them wanted to get into a glass jaw	let them live in their dead end low cost jobs	let them rot in their dead end low class jobs
37.	how long is	how long can we maintain this week with	how long can we maintain equilibrium	how long can we maintain this week with	how long can he maintain his equilibrium
38.	currently jobless thing is that	until his colleagues at the job of the future to clean kids out	he told his colleagues that the job of going into	until his colleagues at that the job of going to clean kids out	he tells his colleagues that it's the job of the teacher to bring kids out

39.	i've only	what rouses most of the students an assignment for white silk	what relative to the field of furniture like to	what rouses most of the students an assignment for white silk	what finally rouses most of his students is an assignment to write self portraits
40.	dell at the end when a	how to spell and younger or older and ready to	tell you what to do and we were	how to spell and younger or older and ready to	for a brief spell they seem younger more open and ready to learn
41.	getting a room with bath	adjust when you're getting a warning to you about how to	just when you will warn you data	adjust when you getting a you to you about how to	but just when you're getting a warm utopian feeling something bad happens
42.	movie at home that are	authentic hero of the movie threatens to become a backup	david hill of the movie to become a bad guy	david hill of the movie to become a bad guy	i'll say the hero of the movie threatens to become its bad guy
43.	one note of caution is	what one does when you're taught in classrooms she was so startled	i flew to new york on classroom she was so startled	what one does when you're taught in classrooms she was so startled	that's why those real time classroom scenes are so startling
44.	ensure that you have been showing the battle zone mode	they show you that if until the system can change the battle moving moment to moment	they show you that until the fifth inning battle moving moment and will	they show you that if until the system can change the battle moving moment to moment	they show you that at least until the system can be changed the battles will be moment to moment

45.	you can disagree complaining that the a short and and want something that will create jobs	republicans disagree completely with the dole tax cut short and that to march on thing that it will create jobs	republicans disagree completely with the dole tax cut short and that are too large on thing that it will reach our	republicans disagree completely with the dole tax cut short and that are that on thing that it will create jobs our	republicans disagree complaining that the bill's tax cuts fall short and that it spends too much on things they say won't create jobs
46.	you people are going for 20 percent of your votes will mean little if	super bowl product line still more than 20 percent of a church pew with people while to what he will	google product line for more than 20 work here is what we will	super bowl product line for more than 20 percent of a what we with people while to what he will	sales of both product lines fell more than twenty percent last year as fewer people bought toys leading up to the holiday season
47.	i think that we should have about a thousand euro	employers are very filtered through november it would shut about it	lawmaker victor one third of november in which about a thousand job	lawmaker victor one third of november in which about a thousand job	the toy maker based here in los angeles said in november it would shed about a thousand jobs
48.	if you are not in charge of economic	has everyone been somewhat deteriorating economic	who would you want your picture when you are	who would you want your picture when you are	and that it wasn't immune from last year's deteriorating economic environment
49.	but the fact that you	sperry will cut taxes that you question	very low cut back to that question	sperry will cut taxes that you question	spend or cut taxes that is the question

50.	it's one thing to be written down and just about every sector and to the growth	thanks lip synch derailing barometric but it works or at your throat	it's one thing to berating her out of just about every sort are are are are	it's one thing to be but it and just about every sort are are the are	pink slips seem to be raining down in just about every sector and every zip code
51.	the other day and bought the	are there any bright spots off the	are there any bright spots off the	are there any bright spots off the	but are there any bright spots out there
52.	temporary job market may provide a	the temporary job market may provide an effort	temporary job market may provide better	the temporary job market may provide an effort	the temporary job market may provide an answer
53.	unique vocal	it's becoming increasingly difficult	it's becoming increasingly difficult	it's becoming increasingly difficult	it's become increasingly physical
54.	but i will let you	the authority has been slow to address it	authorities have been slow to address	the authority has been slow to address it	the authorities have been slow to address it
55.	an egyptian court convicted a man shall	an egyptian court convicted a man	an egyptian court convicted a man of such	an egyptian court convicted a man of such	an egyptian court convicted a man of sexual harassment
56.	we are advocates are cautiously hopeful	women's rights advocates are cautiously hopeful that	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful that	women's rights advocates are cautiously hopeful now

Appendix C: Recognition Engines Results – 3.5mm Auxiliary Cable

Shading indicates a valid actual spoken phrase, meaning each bigram of the phrase is found within the corpus.

	Listener 1 (Satellite)	Listener 2 (Dell)	Listener 3 (Tablet)	Distributed Listening Result	Actual Spoken Phrase
1.	it's a disease that affects how you move	it's a disease that affects how you move	it's a disease that affects how you move	it's a disease that affects how you move	it’s a disease that affects how you move
2.	for the winter encampment to stop	a finger wiggle and you can't really get it full stop	a finger wiggles and you can't really get it full stop	a finger wiggles and you can't really get it full stop	a finger that wiggles and you can't really get it to stop
3.	on the swing quite the thing that	unarmed doesn't swing quite the same	an armed swing quite the same	an armed swing quite the same	an arm doesn't swing quite the same way
4.	your handwriting has changed	your handwriting has changed	your handwriting has changed	your handwriting has changed	your handwriting has changed
5.	if i should put this clear to	if whirling motion up short in the ship's first clear	it's well in motion before any of these symptoms will become clear	it's well in motion before any of these symptoms will become clear	is well in motion before any of these symptoms first become clear
6.	a lot has happened before the first to	a lot has or uke perched notes	a lot has happened before you first notice	a lot has happened before you first notice	a lot has happened before you first notice it

7.	the first thing i remember noticing what are you handling temptation in my left leg in our	the first thing i remember notice you aren't posting links on station in my left leg and stomach stunt pilot are	the first thing i remember noticing what i've been tingling sensation in my left leg and some of my left arm	the first thing i remember noticing what i've been tingling sensation in my left leg and some of my left arm	the first thing i remember noticing was this odd buzzing tingling sensation in my left leg and to some extent my left arm
8.	i just feel that my cell phone is vibrating a reach for an ipod or not they are	i used to feel that myself up to by birdie at reach for it cannot find it there but not there	i used to feel that my cell phone with vibrating a reach for and i find nothing there	i used to feel that my cell phone with vibrating a reach for and i find nothing there but not there	i used to feel that my cell phone was vibrating and i'd reach for it and then i'd find that there was nothing there
9.	i noticed that i didn't think that my arm was swinging our way to draw	i noticed that i didn't think my arm what's going in quite the same way i sure	i noticed that i didn't think my arm swinging quite the same way i draw	i noticed that i didn't think my arm what's going in quite the same way i sure	i noticed that i didn't think that my arm was swinging quite the same way when i jogged
10.	the reason i really did was because of my family	the only reason i rate it just	the only reason i really do because of my family history	the only reason i really do because of my family history	the only reason i really did was because of my family history
11.	if the prevailing wisdom for a very very	and that was the prevailing list for surgery	and that was the prevailing wisdom for a very very long time	and that was the prevailing wisdom for a very very long time	and that was the prevailing wisdom for a very very long time
12.	to stop the disease before they can start	can you stop the disease before they can start	each stop the disease before the storm	can you stop the disease before they can start	and you'd stop the disease before it even starts

13.	what the research figure into it you will	with stem cell research figure in chief that you won't	one of the research figure in you that you won't	one of the research figure in you that you won't	would stem cell research figure into the genetic connection
14.	learn about the disease more complicated	the more we learn about the new work	the more we learn about the need for public	the more we learn about the need for public	but the more we learn about the disease the more complicated it becomes
15.	in lots of ways in the research task that much more	lots way they researched that much work	in lots of ways to name the research task that much more complex	in lots of ways to name the research task that much more complex	in lots of ways it made the research task that much more complex
16.	the office of a number of researchers was	and we also saw the number for searchers walked	and we also saw the number of researchers are	and we also saw the number of researchers are	and we also saw that a number of researchers left
17.	memento of the plaintiff or defendant without we weren't in the	they went other places that were more open to stem cell research without which in europe or in singapore	they went to other places are more open systems research whether that was in europe when singapore	they went other places that were more open to stem cell research without which in europe or in singapore	they went to other places that were more open to stem cell research whether that was in europe or in singapore
18.	but i forget what you said it had a negative effect on the way in which the	but i think it's pretty clear that it had a negative impact on the way in which this	but i think it's pretty clear that it had a negative effect on the way in which the filter graph	but i think it's pretty clear that it had a negative effect on the way in which the filter graph	but i think it's pretty clear that it had a negative effect on the way in which the field progressed

19.	i also understand the moral objection to have the	i also understand the moral objection to some people have	i also understand the moral objections that some people have	i also understand the moral objection to some people have	i also understand the moral objections that some people have
20.	a lot of people think of when you might be telling when your body will not	for a lot of people it's a dilemma because you think might be telling one thing in your body telling you not	a lot of people it's a dilemma because your faith might be telling them your body will	for a lot of people it's a dilemma because you think might be telling one thing in your body telling you not	for a lot of people it's a dilemma because your faith might be telling you one thing and your body is telling you another
21.	mouth and if it happened seemingly overnight	but he was not and if it happened seemingly overnight	but he was not and if it happened overnight	but he was not and if it happened seemingly overnight	but he was young and this had happened seemingly overnight
22.	and it does appear that there is a relationship	and it does appear that there is a relationship	it does appear that there is a relationship	and it does appear that there is a relationship	and it does appear that there is a relationship
23.	you will become an environmental	genetics load the guns and the environment pulled the trigger	genetic load the gun and the environmental trigger	genetics load the guns and the environment pulled the trigger	genetics load the gun and the environment pulls the trigger
24.	there are medications that make a real	there are medications that make out	there are education	it helps that some problems are not with others	there are medications that make a real difference
25.	it helps the phone problems are not with other	it helps with some auto show not without	it helps that some problems are not with others	i guess i choose to believe that i'll be able to do this for a very long	it helps with some problems and not with others

26.	perfectionist and i've enabled it to the	i guess i choose to believe that i'll be able to do that starter	i guess i choose to believe that i'll be able to do this for a very long	but don't necessarily need or want to know more than	i guess i choose to believe that i'll be able to do this for a very long time
27.	assuming you don't want to know what to	but don't necessarily need or want to know more than	they don't necessarily need or want to know	but don't necessarily need or want to know more than	but they don't necessarily need or want to know more than that
28.	you know the kind of health for the past year are the long term effects of the	he had been in declining health for the past year and what the long term effects of stroke each	you don't declining health of your dealing with the long term effects of the stroke he suffered	you don't declining health of health for the past year are the long term effects of the stroke	he had been in declining health for the past year dealing with the long term effects of the stroke he suffered
29.	nominated for an academy award for best language	and now it's nominated for an academy award for best foreign language feature	and nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language feature	and now it's nominated for an academy award for best foreign language feature
30.	the class into semi improvised look inside the high school weren't working or are you	the class and family and provide looking tight high school in the first working class parent	the class is a semi improvised the inside high school in the diverse working class parents need to	the class is a semi improvised the inside high school in the diverse working class parents need to	the class is a semi improvised look inside a high school in a diverse working class paris neighborhood
31.	take a deep breath and interviewing	then they all take a deep breath and interviewing	then we all take a deep breath and interview me	then we all take a deep breath and interview me	then they all take a deep breath and enter the arena

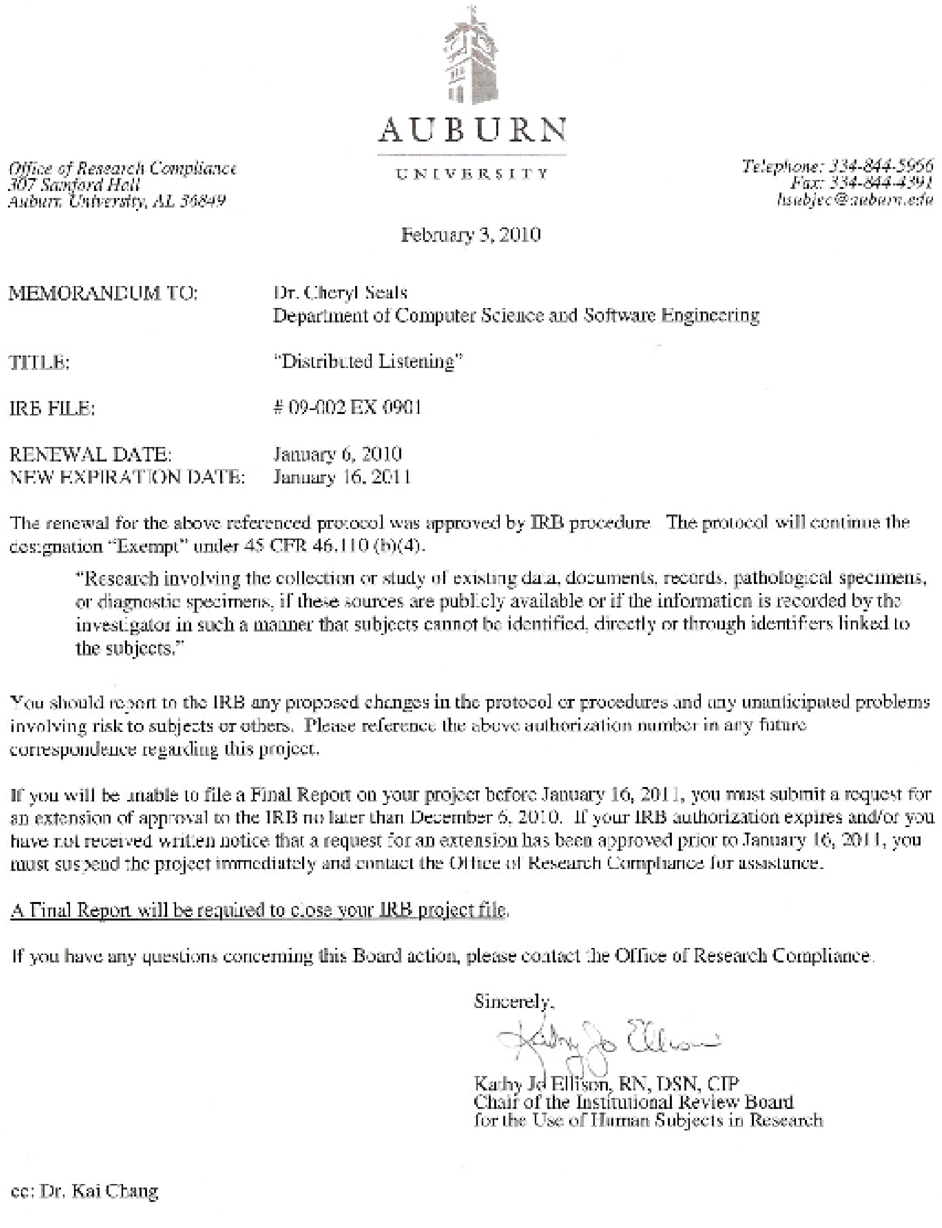
32.	from the blackboard and a student	once on the blackboard and a student makes stop in the photo were	from the blackboard at the next stop in the following work	once on the blackboard and a student makes stop in the photo were	he writes on the blackboard and a student makes him stop and define a word
33.	my favorite feature have to set aside when he wrote to the world that is true to	at so the teacher has to set aside when he first one would be wrong with that and then go would have been true	i thought the feature set aside when they first what would be wrong with that i know what you think	at so the teacher has to set aside when he first one would be wrong with that and then go would have been true	and so the teacher has to set aside his plan and say first what would be wrong with that and then no it isn't true
34.	left within rao and it works well	and by the lesson has been rao nurse who told well	by then the lesson has been real in your sneakers although	by then the lesson has been real in your sneakers although	and by then the lesson has been derailed and there are snickers all around
35.	teach kids to be educated to	these kids don't serve to be educated faith	these kids don't deserve to be defeated if it	these kids don't deserve to be defeated if it	these kids don't deserve to be educated they say
36.	let them live in their quest job	let them live in your skin will cluster out	let them watch you get a world class job	let them live in your skin will cluster out	let them rot in their dead end low class jobs
37.	how well do you think we	how long can we maintain this	how long can mean anything from the	how long can mean anything from the	how long can he maintain his equilibrium
38.	don't you think the job of the future will have to	he told his colleagues that the job of the future from being kicked out	joseph cawley is the job of the feature to link it to	he told his colleagues that the job of the future from being kicked out	he tells his colleagues that it's the job of the teacher to bring kids out

39.	rouses most of us know so we	both houses most of the students an assignment like self quick	row since most of the students an assignment to write self with	row since most of the students an assignment to write self with	what finally rouses most of his students is an assignment to write self portraits
40.	put a spell on her way to the	how to spell a younger and we were	spell over more of the writing from the	spell a spell on her way to the	for a brief spell they seem younger more open and ready to learn
41.	just to let you know if god	just when you're getting warmed up fuel back	but just when you're getting a warm welcome fuel from backup	but just when you're getting a warm welcome fuel from backup	but just when you're getting a warm utopian feeling something bad happens
42.	see this movie to come back to the	authentic hero of the movie threatens to become a bad guy	if you love the movie threatens to become bad for	authentic hero of the movie threatens to become a bad guy	i'll say the hero of the movie threatens to become its bad guy
43.	20 posted on classroom to	it's one of those you'll find quite so stark	it's one of those with you on classroom or so start	it's one of those with you on classroom or so start	that's why those real time classroom scenes are so startling
44.	show me the oktoberfest show about a moment in the	they show you that until the system can be changed to battle and moment to moment to	they show you the until the system can be changed to battle moment to moment	they show you that until the system can be changed to battle and moment to moment to	they show you that at least until the system can be changed the battles will be moment to moment

45.	republican disagree completely with the bush tax cut that to what you say what we ought to	republican disagree complaining that the dole tax cut short and that spends too much on thing that i walk rich off	of the disagree completely with the bill for a short and bit too much on things they say will create jobs	republican disagree completely with the dole tax cut short and that spends too much on things they say will create jobs	republicans disagree complaining that the bill's tax cuts fall short and that it spends too much on things they say won't create jobs
46.	if you are going for 20 percent more accurate about what you	filled with both water flowing through more than 20 percent watcher scooped up about two weeks we will refute	filled with both product lines fell more than 20 percent last year and the people bought the week leading up to the holy	filled with both product lines fell more than 20% last year and the people bought the week leading up to the holy	sales of both product lines fell more than 20 percent last year as fewer people bought toys leading up to the holiday season
47.	i think you have a show about one thousand the	toymaker victor will start in november it would shut about it	toymaker victor walter of november in which about one thousand job	toymaker victor will start in november it would shut about it	the toy maker based here in los angeles said in november it would shed about one thousand jobs
48.	live from the future of economic	has everyone been somewhat deteriorating economic	if it wasn't immune from western european economic	if it wasn't immune from western european economic	and that it wasn't immune from last year's deteriorating economic environment
49.	split effective	at sparing the attack that question	sparing the fact that you	at sparing the attack that question	spend or cut taxes that is the question

50.	it's one thing to be way out of just about every to the	at thing to a finger berating barometric about every starter at co	think the thing to be raining down at the airport sector and for the growth	think the thing to be raining down at the airport sector and for the growth	pink slips seem to be raining down in just about every sector and every zip code
51.	okay but i thought	are there any bright spots out	are there any bright spot for the	are there any bright spot for the	but are there any bright spots out there
52.	but the job market may provide an	the temporary job market lake water	the temporary job market may provide an effort	the temporary job market may provide an effort	the temporary job market may provide an answer
53.	it's becoming increasingly difficult	heat become increasingly pitiful	each become increasingly difficult	each become increasingly difficult	it's become increasingly physical
54.	authorities have been slow to address it	the authority has been slow to attract	the authority to go to	the authority has been slow to attract	the authorities have been slow to address it
55.	an egyptian cleric convicted a man of such the	an egyptian court convicted a man of such	an egyptian court convicted a man of schoharie	an egyptian court convicted a man of such the	an egyptian court convicted a man of sexual harassment
56.	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful	women's rights advocates are cautiously hopeful now

Appendix D: Internal Review Board Approval



AUBURN
UNIVERSITY

Office of Research Compliance
307 Samford Hall
Auburn University, AL 36849

Telephone: 334-844-5966
Fax: 334-844-4391
hsubjec@auburn.edu

February 3, 2010

MEMORANDUM TO: Dr. Cheryl Seals
Department of Computer Science and Software Engineering

TITLE: "Distributed Listening"

IRB FILE: # 09-002 EX 0901

RENEWAL DATE: January 6, 2010
NEW EXPIRATION DATE: January 16, 2011

The renewal for the above referenced protocol was approved by IRB procedure. The protocol will continue the designation "Exempt" under 45 CFR 46.110 (b)(4).

> "Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects."

You should report to the IRB any proposed changes in the protocol or procedures and any unanticipated problems involving risk to subjects or others. Please reference the above authorization number in any future correspondence regarding this project.

If you will be unable to file a Final Report on your project before January 16, 2011, you must submit a request for an extension of approval to the IRB no later than December 6, 2010. If your IRB authorization expires and/or you have not received written notice that a request for an extension has been approved prior to January 16, 2011, you must suspend the project immediately and contact the Office of Research Compliance for assistance.

A Final Report will be required to close your IRB project file.

If you have any questions concerning this Board action, please contact the Office of Research Compliance.

Sincerely,

Kathy Jo Ellison

Kathy Jo Ellison, RN, DSN, CIP
Chair of the Institutional Review Board
for the Use of Human Subjects in Research

cc: Dr. Kai Chang