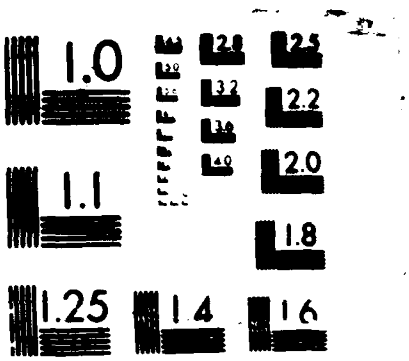
AD-A179 762 AN EVALUATION OF SPEECH RECOGNITION TECHNOLOGY(U) 1/2
BATTELLE MEMORIAL INST RESEARCH TRIANGLE PARK NC
M G JOOST ET AL. 05 DEC 86 DAAG29-81-D-0100

UNCLASSIFIED NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

2

DEPARTMENT OF INDUSTRIAL ENGINEERING
BOX 7906
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695

AN EVALUATION OF SPEECH RECOGNITION TECHNOLOGY

Michael G. Joost
Taryn S. Moody
Robert D. Rodman

An Evaluation of Speech Recognition Technology

by

Michael G Joost
Taryn S Moody
Robert D Rodman

North Carolina State University

for

Product Manager, Army Communicative Systems

5 December 1986

Contract No. DAAG29-81-D-0100
Delivery Order 2439
Scientific Services Program

The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

SECURITY CLASSIFICATION OF THIS PAGE

AD-A179 262

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp Date Jun 30 1986

1a REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
~~May not be released by other than sponsoring organization without approval of US Army Research Office.~~

4 PERFORMING ORGANIZATION REPORT NUMBER(S)
Delivery Order 2439

5. MONITORING ORGANIZATION REPORT NUMBER(S)
TCN 86-564

6a. NAME OF PERFORMING ORGANIZATION
Dr. Michael G. Joost

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
U.S. Army Research Office

6c. ADDRESS (City, State, and ZIP Code)
Dept. of Industrial Engineering
NCSU, Box 7906
Raleigh, NC 27695

7b. ADDRESS (City, State, and ZIP Code)
P.O. Box 12211
Research Triangle Park, NC 27709-2211

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
U.S. Army Communicative Systems

8b. OFFICE SYMBOL (If applicable)
AMCPM-ACS

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)
(Charles H. Clark, Jr.)
P.O. Box 4337
Ft. Eustis, VA 23604-0337

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
An Evaluation of Speech Recognition Technology

12. PERSONAL AUTHOR(S)
M.G. Joost, T.S. Moody, R.D. Rodman

13a. TYPE OF REPORT
FINAL REPORT

13b. TIME COVERED
FROM 7 Jul'86 TO 7 Nov'86

14. DATE OF REPORT (Year, Month, Day)
1986 December 5

15. PAGE COUNT
130

16. SUPPLEMENTARY NOTATION Task was performed under a Scientific Services Agreement issued by Battelle, Research Triangle Park Office, 200 Park Drive, P.O. Box 12297, Research Triangle Park, NC 27709

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Speech recognition, recognition performance, effects of noise.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The last 15 years has seen the development of speech technology at a very rapid rate. Unfortunately, the fact and fiction of recognition are not always easily separated. This confusion is not only evident among users, but also often among system integrators.

This paper outlines the technology today and provides results from one set of benchmark tests. Three tests were performed with live speakers in three noise environments. The tasks used: a) a sixteen word discrete vocabulary, b) a 37 word connected speech vocabulary, and c) a 30 word connected speech vocabulary which was very tightly constrained (syntaxed). The noise environments included a "quiet" background noise, a noisy background of loud voices, and the sounds associated with a normal (loud) vehicle repair shop.

(OVER)

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION

22a NAME OF RESPONSIBLE INDIVIDUAL

22b TELEPHONE (Include Area Code)

22c OFFICE SYMBOL

83 APR edition may be used until exhausted
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

19. ABSTRACT (Continued)

The results indicate that virtually all systems tested could be made to perform well with specific, well-motivated speakers and under all noise conditions. Conditions not requiring advanced features (eg. large vocabularies or connected speech) may turn these features into a liability through increased error. In spite of this, however, the technology is sufficiently mature to support many field applications.

Table of Contents

List of Figures

List of Tables

Acknowledgements

The authors wish to thank IBM, Intel, Interstate Electronics, ITT, Kurzweil AI, Texas Instruments, Verbex, and Votan for their participation which made this study possible. Appreciation is also due to Charles Clark and the Product Manager, Army Communicative Systems for initiating and supporting this task.

I. Abstract

The last 15 years has seen the development of speech technology at a very rapid rate. Unfortunately, the fact and fiction of recognition are not always easily separated. This confusion is not only evident among users, but also often among system integrators.

This paper outlines the technology today and provides results from one set of benchmark tests. Three tests were performed with live speakers in three noise environments. The tasks used: a) a sixteen word discrete vocabulary, b) a 37 word connected speech vocabulary, and c) a 30 word connected speech vocabulary which was very tightly constrained (syntaxed). The noise environments included a "quiet" background noise, a noisy background of loud voices, and the sounds associated with a normal (loud) vehicle repair shop.

The results indicate that virtually all systems tested could be made to perform well with specific, well-motivated speakers and under all noise conditions. Conditions not requiring advanced features (eg. large vocabularies or connected speech) may turn these features into a liability through increased error. In spite of this, however, the technology is sufficiently mature to support many field applications.

II. Introduction

This report is the outgrowth of questions regarding the state of the art in speech recognition both capabilities and limitations. Based on the claims of vendors, it appears that substantial progress has been made recently in the effectiveness of speech recognition systems. There are, however, disappointingly few large-scale applications from which data may be drawn for comparison of the various systems. The shortage of information is especially evident when comparable data on noise, speaker, or recognizer effects are needed. This study attempts to provide some baseline data which would be useful in the evaluation process. Three distinct scenarios are addressed and extrapolation beyond these limits must be pursued with great care. The term "performance evaluation" has been used, though it is necessary to recognize that there are currently no performance standards for speech recognition systems. As a result, there are no "standard" tasks, and a certain amount of quibbling over the appropriateness of a given test, task, or scenario is inevitable.

A. Objectives

Given the above constraints, the primary objectives of this study were to:

a) Perform a survey of the vendor literature,

b) Assess the performance of as many systems as practical, and

c) Provide first-pass data on differential system performance.

The recognition systems evaluated in this study were: the Interstate Vocalink S4000; the ITT Multibus CSR; the Verbex Series 4000; the TI Speech Development System; the Votan VPC 2100; the IBM Voice Communication Adapter; the Intel iSBC 570; the Interstate CSRB; and the Kurzweil Voice System.

The vendor literature was surveyed for two purposes. It provided an assessment of the vendor perception of the state of the speech recognition art. Additionally, it served to identify the vendors who currently have product offerings. As with any survey, there may be some product which has inadvertently been omitted, but every effort has been made to solicit information from every known potential vendor.

Several performance assessments were mandated. For comparison with other studies, testing in a quiet environment was necessary. To be of more practical value, however, it was essential to test performance in two additional noise environments. The first emulated an industrial environment and was taped in an automotive service shop, while the second was somewhat more innocuous, consisting of vocal noise at a fast food restaurant order counter. This last environment approximates a noisy classroom or other area with verbal interference.

Finally, many of the existing studies of recognizer system performance use taped speech under the rationale that each system receives identical input. Taped speech and its means of entry into the recognition system differ in several marked respects from live speech delivered orally into a microphone. Live speech was chosen for this study to preserve the more realistic performance environment, and within speaker variation was dealt with statistically.

III. Background

Almost since the advent of the first commercial speech recognizers in the early 1970's, manufacturers of automatic speech recognizers (ASR's) have been claiming high performance for their systems that is often not achieved in actual applications. The net result, therefore, is a perhaps healthy, skepticism of manufacturers' claims. For this reason, concerns have arisen about how to best evaluate a system for a specific application and a given group of users. The system evaluation often takes two forms; systems can be evaluated based on a review of the literature available from manufacturers or other users, or they can be evaluated through tests of the system performance. The former provides essential design information (eg. vocabulary size, language/application support, or price), whereas the latter is required to characterize the ASR's behavior under actual operating conditions (eg. recognition accuracy, speed, training time, application development time).

IV. Marketing Literature Review

In reviewing the marketing material provided by manufacturers, information was extracted addressing several major areas. These include the technology, vocabulary capacity, training support, hardware and software compatibility, and development tools.

A. Technology

The issue of technology includes two dimensions: manner of speaking and speaker dependence. Although all vendors identify the segment to which they belong, no universally accepted definitions for these terms exist. For this reason, a certain amount of confusion results. For clarity, definitions are presented which closely follow those proposed by Pallett (1985).

1. Manner of Speaking

The cadence of speech allowed (in some cases enforced may be a better term) by the technology can be broken into three classes.

Discrete Speech forces the speaker to aid the recognizer by pausing between each utterance. This results in somewhat stilted speech, may be perceived as being slow, and appears hard for some speakers to learn. In spite of this, discrete utterance recognition is the most common implementation today and is quite adequate for many speech input tasks.

Connected Speech, on the other hand, requires that the word be spoken carefully, but does not require that an explicit pause be used to separate each utterance. Although this appears easier for speakers to use, it is achieved through higher processing requirements and thus, is usually somewhat more expensive.

Continuous Speech is most like natural speech. Words are spoken fluently and rapidly as in conversational speech. When this occurs, however, speech sounds are influenced by neighboring sounds (coarticulation).

Evaluating vendor products can become somewhat confusing at this point, since many vendors do not make the distinction between connected and continuous speech. Additionally, there is nothing inherently "better" about a system simply because it allows or promotes the use of one type of speech. It is the application which usually dictates the recognition requirements. Thus, no one system is likely to prove more suitable than others for all applications and it is likely to be a mistake to attempt to identify one system that is to be the standard for all future applications. In general, unless the application really requires connected speech recognition capabilities, selection of a discrete speech recognizer is desirable, because of the additional cues provided by the speaker (ie. pauses) to the discrete system, which usually make it more tolerant of environmental noise.

2. Speaker Dependence

Speaker Dependent recognition relies on matching speech samples to previous utterances of the same speaker. An enrollment or training procedure is followed to allow the system to extract adequate models of the individual's speech patterns.

Speaker Independent recognition, however requires no enrollment for recognition. Rather than using speaker specific models for recognition, general models appropriate for a large population are used. Most existing systems with speaker independent capabilities have relatively small vocabulary sizes (eg. digits plus several control words), and tend to have somewhat lower recognition accuracy than is usually attained by comparable speaker dependent systems.

The vendor products tabulated In Table 1, are more completely described in Appendix I. In several instances, a vendor has claimed continuous capabilities but may be shown as connected to preserve the above definitions. In rare cases, insufficient information was available to make this assessment so a question mark was inserted to identify the uncertainties.

3. Vocabulary Capacity

The question that is probably most often asked relates to the size of the recognition vocabulary. What most individuals tend to forget is that at any instant in time, unless the user is attempting verbal dictation, the probability is very low that more than a relative handful of words are feasible in the existing context. Additionally, there is usually a trade-off that must be made; as the candidate vocabulary gets larger, the probability of recognition error increases. What in many cases is a more pertinent question is how well the system supports subdividing the vocabulary. As is evident from Table 1, there are a wide variety of vocabulary sizes supported by the various systems. This reflects several major design philosophies - provide several relatively small vocabularies which may be switched very rapidly, a larger vocabulary that can be arbitrarily split under program control, or a large vocabulary that relies heavily on the accuracy of the recognition algorithm. As a point of reference, there are very few well-structured applications requiring more than 200-300 words in the vocabulary if the application is thoroughly studied, understood, and designed.

Table 1. Summary of Vendor Literature.

VENDOR	PRODUCT	RECOGNITION TECHNOLOGY					VOCAB.	VOICE	PRICE
		Dep.	Ind.	Disc.	Conn	Cont	SIZE	OUTPUT	($ US)
AT&T	Conversant 1	X	X	X	X		256[1]	?	quote
AUDEC	SSB-1000	X		X			144		250
Calltalk	DVIO Mod. 100	X			X	?	500	?	quote
Dragon Systems	Voicescribe 1000	X		X			1000		995
Dragon Systems	Voicescribe 20000		X	X			20000		quote
IBM	Voice Comm. Adpt.	X		X			5*64[4]	X	1,700[2]
Intel	iSBC 570	X		X			200		2,900
Interstate V P	CSRB	X		X	digits		240	opt.	1,410
Interstate V P	SRB-LC	X		X			400		395
Interstate V P	Vocalink S4000	X			X		100		5,200
ITT DCD	Multibus CSR	X		X	X	?	300	X	37,000
Kurzweil A I	KVS	X		X			1000		6,500
Microphonics	(various)	X		X	X		128		quote
NEC America	(various)	X	X	X	X		<500[3]		quote
Scott Instr.	Coretechs VET 3	X	X	X	X		200	X	8,995
Speech Systems	Phonetic Engine	X			X	?	5000	?	quote
Texas Instr.	TI-Speech	X		X	X		20*50[4]	X	1,155
Toshiba	TOSVOICE			X	X		64	?	quote
Voice Indust.	Verbex 4000	X			X		100		5,500
Voice Cntrl Sys	VCS Technology		X	X			20	?	quote
VOTAN	VSP 1010	X	limited	X	X	?	64+	X	quote
VOTAN	VPC 2100	X	limited	X	X	?	80+[5]	X	quote
Westinghouse	Series 100 VDCS	X			X		200	?	quote
XCOM	Seraphine	X	X	X	X		100	?	3,000

[1] 13 in speaker independent mode.
[2] Bundled price, may also be purchased unbundled.
[3] Less than 20 in independent mode.
[4] Total vocabulary must be divided into subsets of which only one may be active at any time.
[5] This may be increased with fewer training passes or optional expansion vocabulary.

4. Training Support

The type of training required depends, to a large degree, on the type of recognizer. While discrete word recognizers require only individual template(s) of each word, connected systems must also be able to account for coarticulation. Coarticulation is the phenomenon observed at the boundary of words spoken together. Each word is influenced by the word preceding it and is influenced, in turn, by the succeeding word. Thus, a connected recognizer relies not only on templates for each word, but also requires models of how coarticulation affects each word-pair. In a very simplistic manner, every possible word-pair boundary must be modeled. Needless to say, the combinations quickly get very large as the vocabulay size grows making the enrollment process very cumbersome unless the possible combinations are efficiently pared down. As examples, the Interstate S4000 and Verbex 4000 generate a relatively exhaustive script for coarticulation estimation, while ITT relies on a training script developed by the application designer. VOTAN uses only the discrete utterance templates (relying on a strong algorithm) and allows operator selected embedded training of particularly troublesome combinations. Scott Instruments does no coarticulation evaluation instead, the VET 3 internally adjusts word boundaries to allow connected recognition.

All systems (except speaker independent systems which, by definition, require no training but may require speech patterns for adaptation purposes) provide utilities for training the vocabulary. In most cases this is an off-line function that acquires and maintains patterns. For most applications, this is sufficient. These static models, however, may not be adequate when speech patterns are likely to change due to stress, boredom, or iatigue. Under these conditions, dynamic updating of user templates may be required to cope with dynamic change, permanent or transient. This dynamic update feature is available from very few vendors at the current time.

Whether due to adaptation or standard training techniques, speech recognition systems usually use multiple utterances against which new speech signals are compared. To guard against inadvertent contamination of the speech patterns, major differences between patterns usually result in a user query thus avoiding the inclusion of coughs, burps, and etc. The method used to represent these composite patterns varies greatly. Most systems use an "averaging" technique where template updates are combined with and replace previously existing templates. One potential danger of this system is that, as more samples are included in the template, it may become more general and, over time, no longer represent the intended utterance very well. This would tend to result in an increased number of errors. An alternative approach used by VOTAN and Kurzweil consumes a vocabulary entry for each update of a word. This technique reduces the chance that the template becomes so general that recognition is adversely affected at the expense of reducing the maximum number of words in the vocabulary. For example, if the recognizer had a 20 word vocabulary limit, a single update (after initial training) will reduce the usable vocabulary to 10, two updates to 7 (if one word had only a single update), etc. The effects of this may be minimized by understating the available vocabulary so that updates do not affect the apparent vocabulary size.

Finally, to achieve consistent performance, the user needs feedback, especially during training. This feedback helps the user develcp the necessary speech habits and allows rapid determination of the effects of mispronuciation.

5. Compatibility

Both hardware and software compatibility issues arise with speech recognition. At the hardware level, a number of factors need to be considered. Probably the most flexible systems employ a stand-alone architecture communicating with the host via an RS-232 (typ.) interface. Examples include the Verbex 4000 and the ITT CSR. Unfortunately, the application development libraries for many of these systems assume a specific host (ie. the support software will run only on a specific operating system). When that constraint is considered, then the workstations developed around the Intel iSBC 570 or Westinghouse systems may be considered just as flexible.

The next level of compatibility currently revolves around computer bus standards (typically Multibus or PC bus). Within this level are two subdivisions. One, the "low-priced" (typically under $500 and designed to be used in a PC) systems, often use the host CPU to execute the recognition algorithm. In this case, the speech board is primarily a "front-end" amplifier/filter. While this reduces the cost of the recognition subsystem, it usually severely curtails other computing functions. The result is often substantially slower processing (in some cases, the application software must be custom-designed to use speech input. This custom software is obviously a very expensive solution and may substantially exceed the savings anticipated through the use of the low-cost recognition hardware resulting in a higher net cost in all but the most trivial applications.

The other subdivision (typically $1000 and up with a PC or Multibus formfactor) uses the host processor as a file server, providing mass storage and supporting the application software. Systems in this class usually include integrated signal processor chips as well as powerful microprocessors. Because of this, the host impact is minimal and host applications can be integrated with voice without the need to redevelop the application support software.

To assess software compatibility, several questions must be addressed. Perhaps the most obvious is the operating system supported. The largest number of systems require MS/PC DOS. These include those which reside in the PC (eg. IBM, TI, VOTAN, and Interstate CSRB) as well as those which communicate via an RS-232 link but have DOS-resident support software (eg. Verbex 4000, Interstate 4000, and Kurzweil). The latter set may be used with non-DOS systems, but the necessary support software is likely non existent. Other operating systems which provide the necessary support include UNIX or its derivatives (supporting ITT's CSR and Intel's iSBC 570) and Intel's iRMX (supporting the Intel iSBC 570).

Within the appropriate operating system, the level of support also differs dramatically, ranging from a set of subroutines to transaction generators. The use of support subroutines assumes that the necessary application host languages are available. In addition, the use of these routines must be initiated from within the application program which requires program modification. At the second level, the speech system interacts with the application through operating system calls. This also requires access to be initiated from within the application program and, thus, may require substantial programming. Finally, some vendors supply a utility which may be generically called a transparent keyboard. After being appropriately designed, this software utility parallels the operation of the terminal keyboard and allows speech recognition to be used without modification to the application software.

At the most sophisticated end of the software support spectrum are the transaction generators. These packages (most notably available from Intel and Westinghouse) generate the necessary software automatically after acquiring the interaction rules from the application developer. As a result, although these may be initially more expensive, application development speed may make the net system cost more competitive.

In general, there are limitations inherent in all the software support packages provided by the vendors. This may, to a large extent, be due to the youth of the technology with very few established application niches. As these applications are reproduced, commonalties are likely to emerge which will likely encourage the development of more generic application generators. These generators in turn, will promote the spread of the technology to other applications.

6. Development tools

Although the software mentioned above may allow the integration of voice into an application, there is another aspect to the support of speech systems. The Intel and Westinghouse packages encourage full and careful use of the voice channel implementing dialogue structures, editing support, vocabulary selection, and syntaxing. This improves the probability that applications will be properly designed and implemented by forcing the application developer to consider all aspects of the user dialog.

High-level support for the Kurzweil, TI, Verbex 4000, Interstate S4000, and ITT CSR also exists. This support software, while being well designed, is used primarily to support advanced recognition features (eg. syntaxing or training) and the broader issues of a comprehensive verbal dialog are not addressed.

V. Performance Testing

In the design of performance tests for ASR's,a number of issues must be addressed: 1) Selection of words to be tested; 2) Identification of test scenarios; 3) Environmental conditions; 4) The type of speech used for input (live vs. recorded); 5) Parameter settings; and 6) Evaluation procedures.

A. Vocabulary Selection

The selection of words to be tested is typically made based on one or more of the following three factors. First, the words selected may form a phonetically balanced word list so that all phonemes represented in the language are in some way tested. The words can also be selected based on the frequency with which they are typically used in voice input applications (e.g. the so-called TI word list suggested by Doddington & Schalk (1981)). Finally, the words selected for testing may be chosen with an application in mind, in which case the words that will be used in the application will provide the best indication of performance.

B. Scenario

Since an application does not consist of a random sequence of words, scenarios must be designed to implement transactions that exercise the vocabulary in such a way as to be representative of typical use. Depending on the application, the transactions will be of varying lengths and degrees of difficulty. Transactions used for testing must either be representative of an actual application, or general enough for generic testing. Consideration must also be given as to whether to use "syntaxing", and to what degree.

C. Environmental Conditions

If it is intended that the results be extended to a specific task, the environmental conditions in which the recognizer is to be tested should replicate the application environment as closely as possible. These environmental conditions should include: the noise characteristics of the application area; the acoustic properties of the room ir which the application is located; and the type of speech input apparatus that is required for the application.

D. Noise Characteristics

Automatic Speech Recognizer (ASR) performance accuracy is influenced to some extent by background noise levels (Rollins et al., 1983). The ambient noise dB level, the sound frequency distribution, and the variability of the noise are characteristics that may degrade ASR performance. The reason for this potential performance degradation is related to the influence of the background noise on the ASR, the speaker, and the microphone.

1. Effects of Background Noise on ASR's

When background noise levels become too high (eg. 85 dB(A)), the signal-to-noise ratio may not be large enough for the ASR to detect which, if any, word has been spoken (Rollins et al., 1983). Depending on the spectral characteristics of the noise, this may be experienced with or without a noise canceling microphone. In this case, a rejection error is most likely to occur (Rollins et al., 1983), though this is dependent on the discrimination levels set on the ASR. Rejection and misrecognition errors are also likely to occur in inconsistent noise (sound pressure levels varying more than 5 dB(A)), where the front end gain function of the ASR does not accurately reflect the background noise. This problem is most significant when the ASR only calibrates the front end gain once and for a brief period of time.

Certain frequency characteristics of noise can also affect the ASR performance, especially high frequency components (> 10,000 Hz), outside of the normal speech range. Microphones and/or ASR's typically attenuate background noise differently. Noise canceling microphones are more effective below 2,000 Hz, and provide little filtering above this frequency (Larson et al. 1986). Thus, high frequency noises can severely degrade ASR performance, and in some cases prohibit the use of these recognizers. This degradation often occurs because the high frequency noise is not removed from the speech signal, and impedes the identification of speech onset or termination.

2. Effects of Background Noise and Setting on Speech

It has been well documented that speech characteristics are altered by high background noise levels (Pisoni et al. 1985; Lane et al., 1971; Draegert, 1951). The source of the background noise - whether it is from machinery or from other speakers in a room - is also of significance (Webster et al., 1962), as are the room characteristics, the speaker task (reading aloud vs. talking), the frequency components of the background (masking) noise, and the use of hearing protection devices.

Characteristic changes occur in speech that is produced when there is a masking noise present. These effects have been demonstrated with background noises as low as 50 dB(A) (Lane & Tranel, 1971). The changes noted in the speech include an increase in vocal intensity (voice level increases by half the increase in background noise); increases in fundamental frequency; and an increase in syllabic duration and consequent decrease in rate of speech (Lane & Tranel, 1971). When the masking noise is produced by other speakers, the rate of speaking has been found to increase and not decrease (Webster et al., 1962). A tilt in the short term spectrum of consonants and vowels has also been observed (Pisoni et al., 1985).

The size and reverberation characteristics of the room have also been shown to alter speech characteristics. Black (1950), found that speech rate was slower in large rooms (1900 cu. ft.), as compared to small rooms (150 cu. ft.), and that speech was slowest for large live rooms (reverberation time = .2-.3 sec.) as compared to dead rooms (reverberation time = .8-1.0 sec.). It was also found that the intensity of the speech was greater in dead rooms as compared to live rooms, especially in the larger room.

Garber et al. (1976), demonstrated that noise of equal intensity differentially affected the voice level dependent upon the noises ability to interfere with (mask) the speech signal. Noise with a range of 20-20,000 Hz produced a significantly higher vocal intensity when compared to noise ranges of 1800-2500 Hz; 4000-6000 Hz; and 20-250 Hz. Vocal intensities produced from masking noise with a frequency range of 1800-2500 Hz was also noted to be significantly higher than those produced from a masking noise with a frequency range of 4000-6000 Hz. When noise of equal loudness were presented, a similar differentiation occurred. Vocal intensities produced in masking noise in the 20-20,000 Hz and 350-700 Hz ranges were significantly higher than those in the 1800-2500 Hz and 4000-6000 Hz ranges. The vocal intensity noted for the masking noise in the 1800-2500 Hz range was also significantly higher than that found in the 4000-6000 Hz range. In general, Garber et al. found that the more the frequency components of the noise mask the speech signal, the greater the change in vocal intensity produced by the speaker.

Howell and Martin (1975) have demonstrated that speech is degraded by speakers wearing hearing protection devices. The hearing protector affects the speaker's ability to hear his(her) own voice (occlusion effect). The hearing protection device "attenuates the airborne energy, but has little effect on the bone conduction portion, except in the lower frequencies were the perceived voice levels are actually amplified as a result of the occlusion effect" (Berger et.al., 1986, p.368.). This results in the speaker perceiving his own voice as being louder than it actually is as compared to the background noise level and a subsequent reduction in voice level of 2 - 4 dB by the speaker (Kryter, 1946; Howell & Martin, 1975).

Additional research has indicated that the above effects may be altered unsystematically by factors such as speaker training (instructions); speaker task; hearing loss, and sidetone effects (Lane & Tranel, 1971; Siegel & Pick, 1974; Borden, 1979).

E. Input Apparatus

The microphone-ASR system combination must be chosen to satisfy several performance and operational requirements in order to facilitate ASR performance in the application setting. The type of microphone, performance characteristics, reliability, durability, ease of use, and comfort are important criteria that need to be considered (Waller, 1985).

Microphones presently in use with ASRs include: headset microphones; handheld microphones; gooseneck microphones; wireless microphones (typically headset); and telephone systems. The headset microphones can be either one-way (no verbal feedback) or two-way (can be used for both speaking and hearing). The headset microphones also can provide a full range of hearing protection. In part, physical constraints of the application dictate the type of microphone selected. However, performance characteristics are equally important in microphone selection.

The microphone chosen for a particular speech recognition application must satisfy two performance characteristics. The microphone must perform accurately and reliably in the specific ASR application. Humidity, temperature, noise levels, physical workspace layout, and the nature of the ASR task all influence the microphone performance and subsequent recognition rates. Technical constraints of the microphone must also be considered. Frequency response, range, directionality, and stability (ability to tolerate head movement without changing the microphone position relative to the mouth) are necessary to assure reliable and accurate input to the ASR. To date, the headmount microphones most consistently satisfy the necessary requirements for speech recognition (Plice, 1983).

Headmount microphones, although the best microphone at the present time for speech recognition, have some negative attributes hindering their usefulness in ASR applications. The microphone can be uncomfortable, move out of position and may not cancel ambient noise sufficiently for successful recognition.

The specifications of the particular microphone model must be considered in attempting to match the application needs of the task with the appropriate microphone. The microphones must restrict extraneous noise sounds from entering into the microphone while enhancing the entrance of human speech sounds. Therefore the attenuation characteristics of the microphone must be matched with the frequency characteristics of the application noise.

The application environment (temperature, humidity, dust) also must be considered with respect to microphone durability. Excessively high (above 55-60 C. [130-140 degrees F.]) or excessively low temperatures (below -25 C. [-13 degrees F.]) air pressure changes, and high humidity levels have been found to alter microphone performance (Peterson, 1980). The use of microphones in work situations may subject the microphone to damaging bumps, jolts, or vibrations. The microphone chosen for use in ASR applications must be able to withstand these environmental characteristics.

An additional microphone characteristic is its physical stability. The microphone must be able to be maintained in the same relative position to the mouth throughout all applications. The movement of the microphone piece severely degrades the performance of the ASR.

F. Type of Speech Used in Research

When testing speech recognizers, an important question is whether to use live speakers or recorded/digitized data. There are advantages and disadvantages in either of these methods of speech input.

Proponents of using recorded/digitized speech support the need for both a standardized testing procedure as well as a standardized data base of speech input. They suggest that this method is the only fair means of comparison between speakers due to the variability that exists in a speaker's utterances of the same word. Several tests of speech recognizers have been completed using this type of data base (Doddington & Schalk,1981; Baker, 1982; and Nusbaum et al., 1986).

Proponents of the use of live speakers suggest that this is the most effective way to accurately compare systems as they might be used in an actual application setting. The inter and intra-speaker variability is a naturally occurring phenomenon that should be accounted for, not controlled. This method also provides the speaker with an opportunity to "tune his voice" to the specific ASR being tested. This typically occurs through the feedback that the machine generates in both the training and testing procedures. This adaptation to the system is often seen and its effects on system performance are important. Finally, the entry of recorded data into the ASR differs considerably from the entry of live data. When played back, taped data must either go directly into the ASR, hence by-passing the microphone, or played out through a speaker which produces a speech signal different in many essential respects from orally produced speech.

The disadvantages of live speech is that its replication requires a large number of speakers and it requires more time than alternative approaches. However, this type of evaluation is also more likely to provide a more realistic view of the system's actual performance in an application setting.

1. Parameter Setting

Many, though not all recognizers allow the user to set various parameters. These parameters typically are used to set the minimum match score (ie. how well the current utterance matches the 'best' template) and match score difference (as the minimum difference increases, the probability decreases that the 'runner-up' word is the correct match). Based on these parameters, the decisions are made to report an utterance as "recognized" or "rejected". If testing is performed with a specific application in mind, these parameters may be adjusted to suit that application. For generic testing, either several combinations of these parameters may be tried, which increases the testing effort, or a "forced choice" philosophy may be adopted so that the system's ability to discriminate among similar sounding words is most conservatively tested.

VI. Method

A. Scenarios/Vocabulary

Three distinct speech scenarios were used for this study. The first scenario used discrete speech and a 16 word vocabulary (Appendix II). The second and third scenarios used connected speech with 37 and 30 vocabulary words respectively (Appendix III).

The second scenario was designed to measure recognition accuracy for connected speech using limited syntaxing. The sentences constructed for this scenario were designed to be application specific. Twenty four sentences were used for this task; twelve sentences were three words long, six were four words long, and six were five words long. The syntaxing used for the scenario varied among the recognizers tested. The Votan speech recognizer had no syntaxing capabilities (except via subsets), so the 37 unique vocabulary words used were available at all times for recognition. The syntax used for the ITT recognizer was restricted; the number of possible word choices were limited once the first word was recognized (Appendix IV). The syntaxing used on the remaining three systems (Verbex, Interstate 4000, and TI), consisted of a first word choice (15 words), a second word choice (15 words), a third word choice (15 words), an optional fourth word choice (11 words), and an optional fifth word choice (6 words). The recognition of the first word did not limit the possible choices for the subsequent utterances, except that they could only be chosen from the appropriate word list (Appendix V).

The third scenario was designed to test recognition rates for digits using a connected speech task and restricted syntax. Five basic sentences were used for this scenario, but they varied in number length (one to five digits). This resulted in 25 test sentences (Appendix VI). The syntaxing used for the recognizers tested in this scenario was equivalent since once the first word choice was recognized, the second word choice was known (limited to one choice) and all further utterances were known except for the number string spoken. The spoken number string was constructed from the digits zero to nine, and contained sets of one to five digits.

B. Equipment

1. Recognizers

The Interstate Vocalink S4000 (Interstate 4000), ITT Multibus CSR (ITT), and Verbex Series 4000 (Verbex 4000) recognizers were used for all three tasks included in this research project. The TI Speech Development System (TI) and Votan VPC 2100 (Votan) recognizers were used in all but the second connected speech recognition task. Additionally, the following recognizers were also tested in the discrete recognition task: an IBM Voice Communication Adapter (IBM); an Intel iSBC 570 (Intel); an Interstate CSRB; and the Kurzweil Voice Systems speech recognizer (Kurzweil).

All of the ASRs tested in this study are commercially available except for the ITT system. This ASR is still considered to be a prototype research system.

2. Microphones

The following microphones were used during this study: an AT 9100 headset microphone, a Kurzweil headset microphone; a Prologue handheld microphone; a TI handheld microphone; a Shure SM10A headset microphone; and a Shure VR230 headset microphone. The microphones used in this study were, in most cases, supplied by the manufacturer of the speech recognizer being tested. When the manufacturer did not supply a microphone, a Shure microphone from the NCSU laboratory was used. The microphone-speech recognizer combinations used in this study are listed in Table 2.

3. Recording Equipment

This study used a JVC Model CR 6060U videccassette recorder, a Vector Research VR 220A amplifier, two Acoustic Research AR-5 speakers, and two 3/4 inch 3M Professional VHS videocassettes for the noise conditions tested. A GenRad 1565-B sound level meter was also used for initial calibration of the noise and for the sound pressure level readings.

C. Speech Signals

The voices of six speakers were used for this study. The speakers, four male and two female, had varying degrees of familiarity with the use of speech recognizers. They had no known hearing disorders, were native speakers of English, and ranged in age from 24 to 38 years.

For each scenario, the recognizers were tested in the room described below. The first test was completed under background level noise conditions. The second and third tests were completed with masking noise being played through two speakers which were approximately two feet away from the speaker and recognizer. The order of the presentation of the masking noise was randomized and the noise consisted of either an industrial noise condition or a fast food restaurant noise condition.

Table 2. Microphone/Recognizer Combinations Tested.

	Microphone					
Recognizer	AT9100	Kurzweil	Prologue	TI	SM10A	VR230
IBM	X					
Intel					X	
Interstate 4000						X
Interstate CSRB					X	
ITT					X	
Kurzweil		X				
TI				X		
Verbex						X
Votan			X			

D. Environment

All recognizers except the ITT system, were tested in a large classroom with high ceilings (11 ft.). The background level of noise in the room was measured at 45 dB(A) and 61 dB(C) (the A and C weightings are different descriptions of the noise characteristics). The ITT system was tested in a smaller office area, and the background noise level was not measured, but was not noticably different.

The sound pressure level measurements taken during the playing of the industrial noise tape indicated an L_{eq} = 79 dB(A). The mean sound pressure level for the C-weighted readings was 82 dB(C). The range of sound pressure levels was from 62 dB(A) to 84 dB(A) and from 73 dB(C) to 86 dB(C). The standard deviation of the A-weighted sound pressure level was 5.38.

The sound pressure level measurements taken during the playing of the fast food restaurant noise tape was less variable, with a standard deviation for A-weighted sound pressure levels of 1.36, and L_{eq} = 80 dB(A). The mean C-weighted sound pressure level was 84 dB(C). The range of the sound pressure levels was from 78 dB(A) to 83 dB(A) and from 83 dB(C) to 87 dB(C).

The speech recognizer and the speaker were both in the free field area of the room with respect to the speakers, so reverberant characteristics of the room were not included in any data.

VII. Procedure

The procedure used for all scenarios consisted of a training phase and a testing phase. Since the training phase varied between the different types of scenarios, noise conditions, and for the various recognizers, the procedures used in this study are described according to these three factors.

A. Training Phase

1. First Scenario - Discrete Task

Two male and two female speakers were trained in the discrete speech task. Three sets of templates were made for each speaker and for each recognizer tested. The templates were made according to the manufacturer's recommendations for the number of utterances of vocabulary words required for accurate recognition (Table 3). The training procedures used for all speakers have thus been grouped according to similar types of suggested manufacturer's procedures.

a. Votan

The training procedure for the Votan consisted of three utterances of each vocabulary word. Each vocabulary word was spoken once, and then this process was repeated two times. Training for the two noise conditions consisted of this same process, and all utterances were made in the noise condition being tested. The Votan system stores all templates, so there was no feedback to the user relating to the closeness of the templates, and no training utterances were rejected by the system.

b. IBM, Intel, Interstate CSRB, TI

The four systems in this category are similar in that they all provide some feedback to the user relating the similarity between the initial utterance of the word and the subsequent training utterances (updated templates). The recognizer, at times, rejected an updated utterance because it did not match the initial template formed. Therefore, the number of utterances listed are the minimal number of utterances of a word assuming that all utterances were accepted (though this was not always the case). The Interstate CSRB required three utterances of each vocabulary word while the Intel and IBM used four utterances. The words were spoken sequentially; a complete pass was made through the vocabulary before subsequent words were spoken. The TI speech recognizer required five tokens of each vocabulary word. Two utterances of each vocabulary word were completed sequentially and then, three additional utterances of each word were completed. Training for the two noise conditions consisted of the same process as described above with all utterances spoken in the noise condition, with the exception oi the TI which required two initial utterances to be made in no noise with the three updated utterances being spoken with the noise present.

Table 3. Number of Utterances Required for Training Recognizers

	Voice Profile	# utterances background noise		# utterances masking noise		
Recognizer					Silence	Noise
IBM		4		4		
Intel		4		4		
Interstate 4000		9	5	4		
Interstate CSRB		3		3		
ITT	10 min.	3		0		
Kurzweil	1 hour	3		3		
TI		5	2	3		
Verbex		9	5	4		
Votan		3		3		

c. Interstate 4000 and Verbex 4000

The training for both of these systems was identical. Training was controlled by the recognizer and required approximately nine utterances of each vocabulary word. The user was not provided with feedback on the accuracy of the word spoken in comparison to the template of the word. However, the user had the option of rejecting utterances if it was felt the word was not spoken correctly. The presentation of the vocabulary words to be spoken was randomized by the system. Training of templates required that an initial training be completed in which each word was uttered approximately 5 times in "quiet". A second trial was then completed in which the user uttered the vocabulary words four additional times under the noise conditions in which the recognizer would be tested.

d. Kurzweil

The Kurzweil recognizer required that an enrollment process be completed prior to the actual training of the vocabulary words used in this scenario. The enrollment process forms a "voice profile" which the system requires for each user. The enrollment process took approximately one hour and is system controlled. The training process of the words used in this study required three utterances of each vocabulary word which was presented serially. The training process was repeated for each of the noise conditions with all utterances spoken with the appropriate noise background.

e. ITT

The ITT recognizer also required a "voice profile", though this process required approximately ten minutes. The actual training process consisted of three utterances of each vocabulary word. The templates were not remade for each noise condition. They were made for the background noise condition only. However, "silence templates" were recalibrated (adapted) for the two noise conditions.

2. Second Scenario - Connected Speech

Two male and two female speakers were used for this task. Each speaker attempted to make three sets of templates for each recognizer. However, templates could not be made for two speakers using the TI recognizer in any noise condition (apparently due to the excessive memory required to store the speech templates for their slow speech), one speaker was unable to use the TI in the fast food restaurant noise condition (also apparently due to insufficient memory), and one speaker was unable to use the Interstate 4000 in the industrial noise condition (apparently due to the interaction between his voice and the background noise). The templates were made in accordance with the manufacturers' specified procedures and, thus, the training procedure varied between the recognizers.

a. Votan

The Votan recognizer required three utterances for each word in the vocabulary. The procedure used was the same as described under the discrete task training procedures. The Votan does permit extraction of connected speech templates, however, when this was attempted, the system ran out of memory space for the templates.

b. TI

The TI recognizer required the speaker to say a sentence and then to say isolated words from the sentence. The sentences were defined and developed by the recognizer. After all words had been spoken using this process, the words were updated three times using the system defined sentences for updating. This process was repeated for each of the two noise conditions, except that the initial training was done in quiet with the three updated passes being done with the appropriate noise being present.

c. Verbex 4000 and Interstate 4000

The training required for both of these recognizers was, again, identical. The first phase consisted of the speaker saying each word in the vocabulary using discrete speech. This was followed by approximately four utterances of each word being spoken using connected speech with sentence or sentence fragments as the prompt. The updating phase consisted of each speaker making approximately four utterances of each word using connected speech as prompted with sentence or sentence fragments. This process was repeated for each of the noise conditions with the initial phase being completed in silence and the update phase being completed with the masking noise present.

d. ITT

The training for the ITT system was controlled by the ITT representative present during the training and testing of this system. Training was continued until the templates had been fine tuned to the representative's specification. Thus, training varied greatly between the speakers. The ITT recognizer did not require retraining for the two masking noise conditions, only the "silence templates" were updated.

3. Third Scenario - Connected Speech

Two male and one female speakers were used for this scenario. Three sets of templates were made for each speaker for each recognizer tested. The three recognizers tested in this scenario, Interstate 4000, ITT, Verbex 4000, (the only systems which supported adequate syntaxing for this scenario) were trained using the same procedures as described for the second scenario.

B. Testing Phase

1. First Scenario - Discrete Speech

Each speaker repeated the 16 words in the vocabulary ten times in random order. The words were recorded as being correctly recognized; not recognized (rejected); or misrecognized. For each misrecognition, the misrecognized word was recorded. This process was repeated for each of the noise conditions for each recognizer. A random order was used to test the recognizers as well as the effects of noise to minimize any order effects. The background noise condition (no noise) was presented first in all cases.

2. Second and Third Scenarios - Connected Speech

Each of the sentences used in these scenarios was repeated four times in sequential order. This process was repeated for each noise condition and for each recognizer tested. The sentences were recorded as being recognized correctly; rejected (no sentence or sentence fragment recognized); or misrecognized, for further analysis.

VIII. Results

The results from the three speech scenarios used to assess the performance of the speech recognizers are presented separately. Due to the distinct nature of these scenarios, the results cannot be directly compared. Additionally, since the systems were tested using their default parameter settings, some exhibited forced recognition (the recognizer returned a match for all utterances, a substitution error was preferred over a rejection error) while other systems forced minimum separation (a rejection was preferred over a substitution).

A. First Scenario

An Analysis of Variance procedure (Sheffe, 1959; Searle, 1971) was used to initially evaluate the data. The dependent variable was the number of words correctly recognized, and the independent variables consisted of the following: recognizer, noise conditions, speakers, recognizer by noise condition, and recognizer by speaker. This model accounted for 90 % of the total variance.

As noted in Table 4, there were significant differences exhibited due to differences in the recognizer used, the type of noise in the environment, and the speaker providing the signal. Additionally, the interaction between the recognizer and speaker or noise was also significant.

Table 4. Analysis of Variance: Discrete Data

Source	df	SS	F	p
Recognizer	8	31840.30	37.68	< .0001
Noise	2	5218.35	24.70	< .0001
Speaker	3	1975.51	6.23	< .0011
Recognizer * Noise	16	8135.65	4.81	< .0001
Recognizer * Speaker	24	5543.41	2.19	< .0088

Key:

df Degrees of Freedom
SS Sum of squares of deviations
F Computed F ratio
p Probability that the observed F ratio is due to chance -
Significance is arbitrarily defined at p <= .05

These effects were further analyzed using a Tukey's Studentized Range Test (Tukey, 1952; Dunnett, 1980). Table 5 presents a matrix of the significant differences (p<=.05) that were obtained between the recognizers based on the number of correct recognitions. For example, the Votan performed significantly better than the Verbex and Interstate 4000 systems. The mean recognition rates obtained for the individual recognizers and the Tukey Analysis can be found in Appendix VII and Appendix VIII, respectively.

A Tukey test for the noise effect demonstrated that the mean correct recognition rate for the no noise condition was significantly higher than the mean recognition rates for both the industrial noise condition and the fast food restaurant noise condition (Figures 1, 2, and 3 with details in Appendix IX). Analysis of the speaker effect indicated significantly higher correct recognition rates for speaker 3 as compared to speakers 2 and 4. The results also demonstrated significantly higher recognition rates for speaker 1 as compared to speaker 4 (Figures 4, 5, and 6).

Table 5. Differences Between Recognizers

[Tukey Test (alpha = .05)]

(+ = row significantly better than column
- = column significantly better than row)

Recognizer	1	2	3	4	5	6	7	8	9
1			-	-	-	-	-	-	-
2			-	-	-	-	-	-	-
3	+	+							+
4	+	+							+
5	+	+							
6	+	+							+
7	+	+							
8	+	+							+
9	+	+	-	-		-		-	

Key:

1 = Verbex
2 = *Interstate* 4000
3 = Votan
4 = TI
5 = IBM
6 = Intel
7 = Interstate CSRB
8 = ITT
9 = Kurzweil

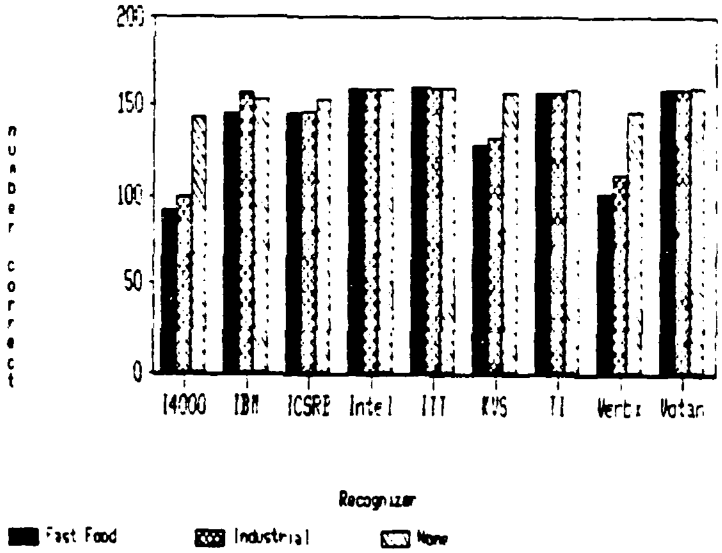


Figure 1. Correct Recognition by Noise (160 maximum)

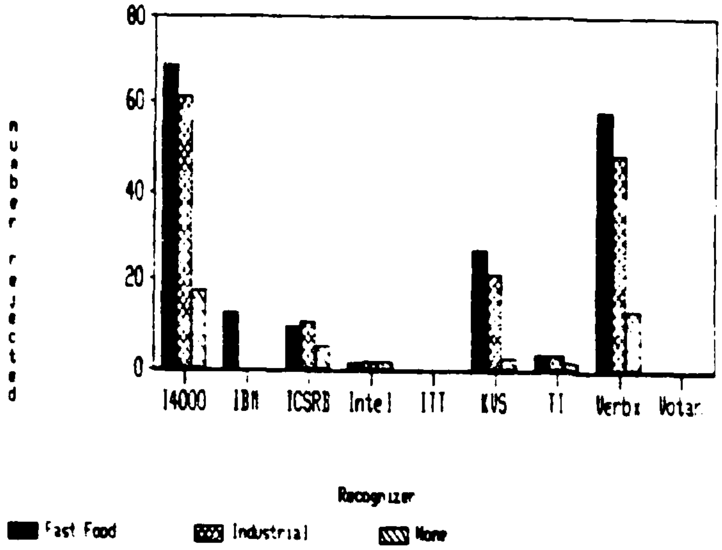


Figure 2. Rejections by Noise

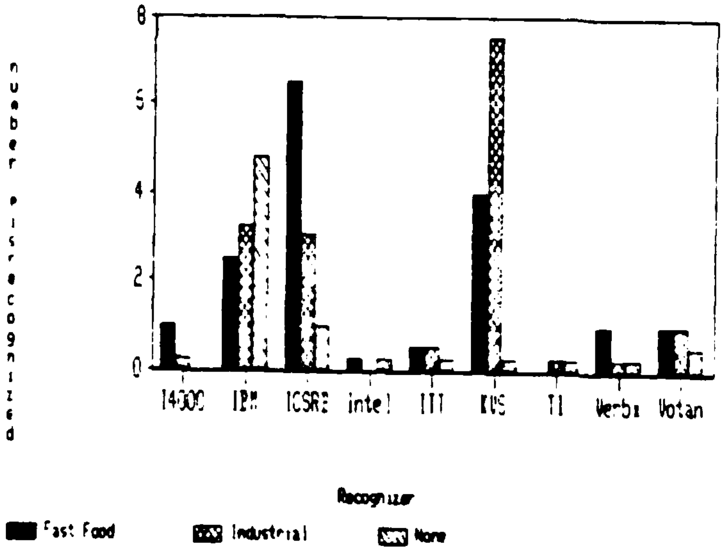


Figure 3. Misrecognitions by Noise

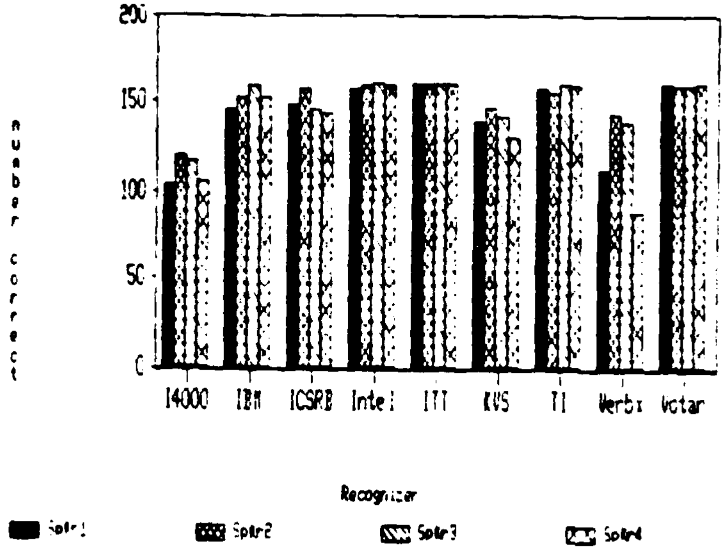


Figure 4. Correct Recognition by Speaker

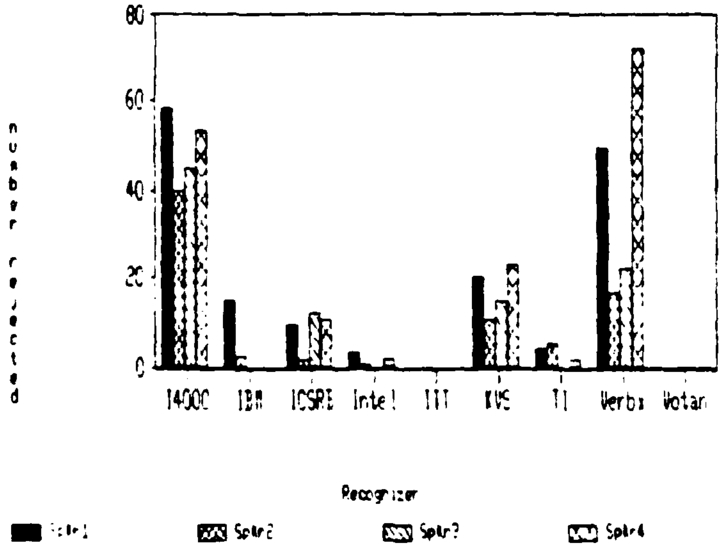


Figure 5. Rejection by Speaker

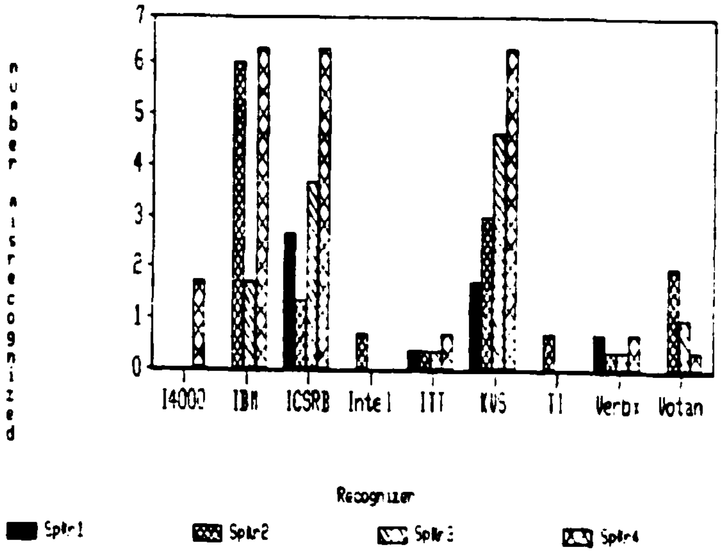


Figure 6. Misrecognition by Speaker

A summary of the interaction effect of recognizer by noise as analyzed by the Tukey test are shown in Tables 6 and 7, with the plus and minus signs having the same meaning as in the previous table. This summary is limited to differences in which the noise condition was held constant. There are no results given for the no noise condition as there were no significant differences in the correct word recognition rates between recognizers for the no noise conditicn. The complete results of this analysis are located in Appendix X.

Table 6. Differences Between Recognizers: Industrial Noise Discrete Speech

[Tukey Test (alpha = .05)]

(+ = row significantly better than column
- = column significantly better than row)

Recognizer	1	2	3	4	5	6	7	8	9
1			-	-	-	-	-	-	
2			-	-	-	-	-	-	-
3	+	+							+
4	+	+							
5	+	+							
6	+	+							+
7	+	+							
8	+	+							+
9		+	-			-		-	

Key:

1 = Verbex
2 = Interstate 4000
3 = Votan
4 = TI
5 = IBM
6 = Intel
7 = Interstate CSRB
8 = ITT
9 = Kurzweil

Table 7. Differences Between Recognizers: Fast Food Restaurant Discrete Speech

[Tukey Test (alpha = .05)]

(+ = row significantly better than column
- = column significantly better than row)

Recognizer	1	2	3	4	5	6	7	8	9
1			-	-	-	-	-	-	
2			-	-	-	-	-	-	-
3	+	+							
4	+	+							
5	+	+							
6	+	+							
7	+	+							
8	+	+							
9		+							

Key:

1 = Verbex
2 = Interstate 4000
3 = Votan
4 = TI
5 = IBM
6 = Intel
7 = Interstate CSRB
8 = ITT
9 = Kurzweil

Results from the Tukey analysis for the recognizer by speaker interaction indicated significantly lower correct recognition performance rates on the Interstate 4000 and Verbex for speakers 2 and 4 (female and male) as compared with all other recognizers tested. For speakers 1 and 3 (female and male) this significantly lower performance rate was found only with the Interstate 4000.

Confusion matrices for the recognizers having more than 10 misrecognitons and error matrices for all recognizers are in Appendix XI.

B. Scenario Two - Connected Speech

The data for the second scenario were initially analyzed using a General Linear Models (GLM) procedure (Goodnight, 1971; Sall, 1978). This statistical analysis was used due to missing data for speaker 1 and 4 for all conditions on the Texas Instruments recognizer and for speaker 3 under the industrial noise condition using the Interstate 4000 and under the Fast Food noise condition for the Texas Instruments recognizer. Though these speakers made several attempts to use these systems, the speaker's templates were either too large for memory (TI) or the recognizer was unable to detect any utterances made by the speaker (Interstate 4000).

The dependent variable for the GLM model was the number of correct sentences, while the independent variables were recognizer, noise condition, and speaker. Interaction effects of recognizer by speaker and recognizer by noise condition were also included. This model accounted for 90% of the total variance. Table 8 indicates that all main effects were significant , as was the interaction effect of recognizer by speaker.

The significant effects were further analyzed using a Tukey Test with an alpha level of .05. Table 9 illustrates the significant differences in performance rates (correct sentences recognized) between recognizers. A plus indicates that the recognizer listed in the row performed significantly (p<=.05) better than the recognizer listed in the column. A minus sign indicates the opposite.

Results of the Tukey test for the noise effect demonstrated significantly higher recognition rates for the no noise and fast food restaurant noise conditions as compared to the industrial noise condition. Results from the analysis of the speaker effect indicated that overall, speaker 2 (female) attained significantly higher recognition rates than did speaker 4 (male) (Appendix XII).

Tables 10 through 13 are matrices indicating significant differences for the speaker by recognizer interaction. The results are limited to those comparisons in which the speaker was held constant. The complete results from this analysis are in Appendix XIII.

Table 8. GLM Results: Scenario 2

Source	df	Type 1 SS	F	p
Recognizer	4	9106.86	35.31	<.001
Noise	2	1009.07	7.82	.002
Speaker	3	858.61	4.44	.013
Recognizer * Noise	8	1018.90	1.98	.094
Recognizer * Speaker	10	1895.80	2.94	.015

Source	df	Type 3 SS	F	p
Recognizer	4	9169.93	35.55	<.001
Noise	2	485.49	3.76	.038
Speaker	3	743.09	3.84	.022
Recognizer * Noise	8	1059.18	2.05	.083
Recognizer * Speaker	10	1895.80	2.94	.015

Table 9. Differences Between Recognizers

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan			-	-	-
TI					-
Interstate	+				-
Verbex	+				-
ITT	+	+	+	+	

Table 10. Differences Between Recognizers: Speaker 1

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan			-		-
TI					
Interstate	+				
Verbex					
ITT	+				

Table 11. Differences Between Recognizers: Speaker 2

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan					
TI					-
Interstate					-
Verbex					
ITT		+	+		

Table 12. Differences Between Recognizers: Speaker 3

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan					-
TI					
Interstate					
Verbex					
ITT	+				

Table 13. Differences Between Recognizers: Speaker 4

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan			-	-	-
TI					
Interstate	+				
Verbex	+				-
ITT	+			+	

Though the interaction effect of recognizer by noise was not significant for this scenario, a Tukey analysis was still completed. The results from this analysis, with the noise conditons held constant, are located in Tables 14 through 16. The complete results of this analysis are located in Appendix XIV.

Matrices for the rejection and misrecognition errors for the sentences are located in Appendix XV. The misrecognition errors were further analyzed using error trees and these are also located in Appendix XVI.

Since the method of data construction for the Tukey test, based on the GLM procedure does not indicate mean recognition scores, these are included in separate tables in Appendix XVII.

Table 14. Differences Between Recognizers: No Noise

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan					-
TI					
Interstate					-
Verbex					
ITT	+		+		

Table 15. Differences Between Recognizers: Industrial Noise

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan					-
TI					
Interstate					-
Verbex					-
ITT	+		+	+	

Table 16. Differences Between Recognizers: Fast Food Noise

[Tukey Test (alpha = .05)]

Recognizer	Votan	T I	Inter	Verbex	ITT
Votan				-	-
TI					-
Interstate					
Verbex	+				
ITT	+	+			

C. Scenario Three - Connected Speech

An Analysis of Variance procedure was used to initially analyze the data for this scenario. The dependent variable was the number of correct sentences, with independent variables consisting of recognizer, noise, speaker, recognizer by noise, and recognizer by speaker. This model accounted for 80 % of the total variance.

As noted in Table 17, only two main effects achieved significance; recognizer and speaker. The data for this scenario are limited by the reduced number of speakers and recognizers tested. Therefore, independent variables that may otherwise have had a significant effect, can only be viewed as having a tendency to affect recognition rates.

Table 17. Analysis of Variance: Scenario 3

Source	df	SS	F	p
Recognizer	2	1658.74	4.92	.03
Noise	2	1185.85	3.52	.06
Speaker	2	3842.30	11.39	<.01
Recognizer * Noise	4	865.48	1.28	.33
Recognizer * Speaker	4	722.37	1.07	.41

The two significant main effects were further analyzed using a Tukey's Studentized Range Test with an alpha of .05. Results from these analyses demonstrate that the ITT recognizer achieved significantly higher recognition rates as compared to the Interstate 4000. Results from the speaker effect indicate that speaker 1 achieved significantly higher recognition rates than speakers 2 and 3. The complete results from these analyses are in Appendix XVIII. No further statistical analysis was performed on the data from scenario three, as no other significant effects were observed.

Matrices for the rejection and misrecognition errors for the sentences are in Appendix XIX. The misrecognition errors were further analyzed using error trees and these are also in Appendix XX.

IX. Discussion

The discussion section is in three parts. First, observations that apply to the entire project in general. Then two sections with remarks mainly pertinent to discrete and connected speech, respectively.

A. General

It is not surprising that recognizer performance was higher, overall, in low noise than in high noise conditions -- after all, people hear better in low noise too. Yet there have been reports of recognizers performing as well or better in very noisy environments than in less noisy ones. The results reported here do not settle this question in either direction. Discrete speech recognition was better under low noise, but connected recognition was better under both low noise and fast food restaurant high noise (i.e., mostly voice noise), as opposed to industrial high noise with its much wider frequency spectrum. These effects, however, were not uniform. Noise had different effects on different recognizers.

The major source of variability in speech recognition is the individual speaker. Both inter and intra speaker variabilities occur, often to a high degree. The effects of inter-speaker variation is minimized when comparing ASRs, providing the same speaker population is used throughout, which it was. While this statistcally minimizes the effects of interspeaker variance, it is still important to recognize that a significant recognizer * speaker interaction term was observed in all but the most tightly constrained case. This implies that it may often be necessary to match speakers and recognizers in many applications.

Intra-speaker variation is affected by many psychological and physiological factors. They impact speech recognition within a single speaker over time periods as short as a few seconds. Day to day variations may be considerable, and may persist for a long time. Thus, if a speaker has an "off" day when a certain recognizer is being tested, those results may prejudice the results against that system (which argues for replication). Significant drift experienced by motivated speakers, however, should be relatively slow changing. Thus, the random deviations should be slowly moving about the target templates. If the sessions are not overly long, the net effect should be tolerable. Having more speakers would have reduced the effect of the variance and, thereby increased the probability of significant findings, leaving these results as somewhat conservative .

With a small speaker pool, the order in which recognition systems are tested may be important, due to experience with speech recognizers. Thus, although order was randomized as far as possible, there is still the possibility that some order effects may still contaminate the results. Without a large speaker pool, which allows full randomization of the order of testing, this factor, if indeed significant, cannot be eliminated.

B. Discrete Speech

All of the above remarks apply in general to the discrete speech part of the experiment. Specific results indicated that the Verbex 4000 and the Interstate S4000 performed significantly poorer than the other systems. Both these machines were designed specifically for connected speech, and were, therefore,operating at a severe disadvantage on an isolated word task. The Kurzweil recognizer, despite its long enrollment process, and its design for large vocabularies, did relatively poorly on the 16 word vocabulary. We do not discount the possibility that the excessively long enrollment procedures (which may be necessary for larger vocabularies) may have "put speakers off" this device, either consciously or subconsciously.

This study found a significant interaction between the various recognizers and various noise conditions. The recognizer * noise term implies that it is not likely that any one of the current recognizers will have the best performance under all noise conditions. This interaction appears to be primarily due to better relative performance of the Kurzweil system with the Fast Food background noise. For this reason, this interaction may not be oberved under different conditions.

The 16 word vocabulary chosen to test the discrete speech recognizers consisted of the ten digits and six control words. This is not a particularly difficult vocabulary, and the fact that all recognizers did not perform at near perfect levels may be a strong, but sad, indication of how much progress remains to be made in ASR design. Even more strongly it indicates how important are the user and user environment. There is little doubt that with carefully chosen, highly trained and motivated users in a controlled environment (not necessarily noise free), near perfect performance can be obtained by all vendors on such small vocabularies. The use of "average" speakers in a loosely controlled environment is, however, justified as being closer to likely application scenarios.

Error analysis consisted of tabulating and analyzing rejection errors and substitution errors (while noting the substituted word). Speakers were not allowed to enter non-vocabulary words or sounds, so that insertion errors did not occur. (An insertion error is when the recognizer interprets some non-vocabulary word, or a sound such as a cough, as a word in the vocabulary.) To measure performance vis-a-vis insertion error avoidance, the recognizer threshold values (or their equivalents) would have to be adjusted, which was deemed impractical at the time.

The generic approach of setting the recognition parameters at manufacturer's recommended levels was taken. Experience has shown, however, that depending on the demands of the application, much benefit may result from an adjustment of these parameters. Indeed, there is evidence that individual speakers may benefit from a fine-tuning of these parameters to their idiosyncratic needs. A close study of how threshold adjustment might affect the relative performance among recognizers was beyond the scope of the project, however, and should be considered for future research.

C. Connected Speech

Table 9 indicates that the ITT ASR did significantly better than the other four, and that the VOTAN did significantly worse than all but one other recognizer. There are several factors that may account for this observation.

In the case of the ITT machine, testing was done in a different room due to overheating problems in the laboratory that housed the other tests. Additionally, the ITT system was a prototype system and required ITT technical assistance. As a prototype, not all the necessary enrollment evaluation routines exist. This forced technical intervention in the enrollment stage and perhaps allowed the system to be tailored more precisely to the scenario than was possible with the development tools available with the other systems. Because of the substantially greater computing power and memory, the ITT machine also permitted the strictest syntaxing. That factor, too, helped to account for the better results.

The use of "syntaxing" is a crucial factor, especially in connected speech, but in discrete speech as well. All the recognizers tested operate by comparing a current template with a series of prestored templates gathered during enrollment. The more prestored templates, the greater the chance for error. Syntaxing allows the comparison to be made on a subset of the prestored templates, hence reducing the possibility for error. The term syntaxing is used because the subsetting operation is generally based on the permissible co-occurrence of utterances. By designing transactions that permit "heavy" syntaxing -- the reduction of candidate templates by an average factor of three or more -- recognition performance can be dramatically improved. As noted above, part of the reason for the success of the ITT recognizer was due to its ability to handle heavy syntaxing.

The VOTAN device, on the other hand, permitted no syntaxing whatsoever, which put it at a disadvantage compared to the other machines. The only ASR the VOTAN did not perform worse than was the TI. The TI machine, however, refused to perform at all for two of the four speakers whose natural speaking rate was slower, apparently due to limits on available memory. In one sense, this lack of data severely compromises any conclusions that could be drawn about the TI machine, or a comparison between it and the VOTAN. In another sense, however, the comparisons are quite valid since the TI was functionally unable to perform the task. Thus, while syntaxing may substantially reduce errors, it may be a two-edged sword; it may increase the task support complexity beyond the capacity of many recognition systems.

Error analysis in connected speech is far more difficult than in discrete speech because a much wider variety of error is possible. In addition to substitution errors, and possibly rejection of all or part of the input, the following errors may occur: Insertion errors: extra words are inserted. Deletion errors: spoken words are omitted. Merge errors: two or more words are recognized as one or more words Split errors: one or more words are recognized as two or more words. These errors may occur in any combination and in any number in a given utterance, sometimes leading to recognizer output that is best described as "word hash."

One way of reporting results for connected speech, is to simply report on the percentage of sentences interpreted without error. This was the initial basis for comparison among the five ASRs used in this connected speech scenario. Additionally, individual tabulations were completed to indicate which of the sentences the recognizers completely rejected (no recognition of any word); and which of the sentences contained any of the previously mentioned errors (Appendix XV). The sentences that contained some type of recogniton error were further analyzed using error trees (Appendix XX).

The error trees compiled for this study were developed in an effort to assess the degree and manner in which an individual recognizer errs in word identification once at least one word in a sentence has been misrecognized, omitted, or inserted. Such an analysis is of interest, because certain types of errors are easier to handle than others (depending on the magnitude of the errors). For example, suppose two recognizers achieve the same percentage correct sentences for a given scenario, but when the first recognizer errs, it returns a sentence that is totally incoherent and unrelated to what was said, whereas the second recognizer returns a sentence with a single substitution error. The second recognizer's performance should be considered superior to the first recognizer since the error of the second recognizer could be corrected more easily than the error of the first recognizer.

In an attempt to quantify errors made in Scenario 2, two values were computed for each recognizer by noise condition, length of transaction (3, 4, or 5 words), and speaker. These values (right margin of Appendix XVI) indicate the number of words correctly identified (R) over the total number of the words (L) in the transaction (R/L) and the number of wrong words (insertions, merges, splits and misrecognitions) (W) over the total number of words (L) in the transaction (W/L). The numbers listed were averages obtained based on analysis of all the errors for the particular recognizer, noise condition, sentence length, and speaker. Since the number of sentences containing errors varied by recognizer (Appendix XIX), the percentages only indicate the type and degree of error that occurs when an error does occur.

The value for the first percentage, R/L, ranged from 0 to 1. Zero indicated that the recognizer did not correctly identify and word spoken, and one meant that the recognizer correctly recognized all the words spoken, but also inserted words that had not been spoken. The value for the second percentage, W/L, is in theory unbounded, with numbers approaching infinity occurring when a recognizer cannot detect the end point of an utterance(as might occur in a high noise environment). However, in this study, the number rarely exceeded a value of 1, except when the recognizer used strict syntaxing and misrecognized the utterance completely.

In theory, with all other factors being equal, the better recognizers, those that would be most amenable to present day error detection and correction strategies, would be those that had high R/L percentages and low W/L percentages. Two caveats must be made. This scoring method judges an omission error to be better than a substitution error (which is debatable). For example, suppose the following sentence was uttered, "Driver move tank out slower". If the recognizer returned the sentence "Driver move tank out faster", the first percentage would be R/L = .8 (4/5 =

.8), while the second percentage would be W/L = .2 (1/5 = .2). (The results of this type of error could also lead to serious problems for the driver of the tank!) However, if the recognizer returned the sentence "Driver move tank out" the R/L still equals .8, but W/L now equals zero. (The Driver of the tank could also ask for the correct speed with which to move the tank!) A second problem with these percentages is that they do not indicate whether the recognizer is consistently making the same mistakes for the same sentences,(in which case retraining the recognizer may solve the problem), or if the errors are inconsistent and randomly distributed (may require redesign of application).

The two percentages calculated could also be supplemented by a third number, not completed for this project. This number would reflect the degree to which the recognizer is able to identify the correct number of word boundaries (correct number of words utterred). This number could be considered the ratio of the difference between the number of words utterred minus the number of words recognized (D) over the total number of words (L) in the sentence (D/L). The better recognizers (more amenable to error detection and correction), would be those whose D/L ratio approached zero.

The R/L and W/L ratios, though initially computed at the lowest level of recognizer * noise * sentence length * speaker are also listed for levels of recognizer * noise * sentence length; recognizer * noise; and recognizer (Appendix XX). Again, as the numbers are based on an average incorrect utterance only, and not on the number of incorrect utterances, the data must be reviewed cautiously. A recognizer that misses one utterance in a thousand but reports no correct words for that sentence will appear the worse than a recognizer that always returns one or two errors in each sentence. Therefore, the error trees must be interpreted with the additional data in Appendix XV. When two recognizers have approximately the same number of misrecognized sentences, the error trees can be used for meaningful comparisons.

The error trees for the third scenario were constructed so as to accurately reflect the type of errors that occurred with the connected digits. Since the scenario employed a highly restricted syntax, for all but the connected digits, the other words in the sentence were correctly recognized (with only a few exceptions).

The three error trees thus reflect the types of errors (S = substitutions, I = insertions, O = omissions), that occurred for each of the recognizers tested (Verbex, ITT, Interstate 4000). The results for each recognizer were further categorized by the noise condition in which the recognizer was tested (N). At the lowest level, the data reflect recognizer error by noise (N), by error type (S,I,O), by speaker (s1, s2, s4).

These results, in contrast to the previous error trees, represent the total number of errors for each speaker by error type, noise and recognizer, and can therefore accurately be compared with each other (between recognizers). However, the results again, do not reflect "consistent" recognizer error where the same mistakes are always made, as opposed to "random" errors. However, consistent recognizer error would be much easier to correct (typically by retraining the particular digit template). The error trees also do not weight the different types of errors in any way when obtaining the average errors per speaker score (SE). The resulting score is somewhat misleading in that the correction for substituted digits would be significantly harder than correction for insertions or omissions.

X. Summary and Conclusions

While there is no conclusive evidence, then, that any of the recognizers tested consistantly excelled beyond the others, some important conclusions can be drawn about the recognition of connected speech in general.

The "care and feeding" of speakers is all important. This point cannot be emphasized too strongly. Performance appears to vary depending on the mood, motivational level, and frustration level experienced by speakers. Systems (and applications) must be designed to minimize these performance moderators. The notion that any worker can use a recognition system with just a few hours of orientation is wrong and may often result in the failure of a project that might otherwise be a success.

At the same time, designer and manufacturers of ASRs must pay attention to the extraction of linguistically significant information from the speech signal. Humans have little trouble understanding other humans when they are angry, sick, or sobbing. The information is present in the speech signal; it remains to be used.

Training time varied somewhat among recognizers. The enrollment procedures for both the Verbex and the Interstate was considerably longer and more tedious than for the ITT, for example. Users of systems are likely to experience a substantial amount of training as long term speech shifts occur. As a result, it is very desireable to have minimal training time to reduce the non-productive time an employee spends on the system.

None of the devices tested could be used in a speaker independent setting, nor did any claim to be speaker independent. Although it is often thought that speaker independence is necessary for application value, the larger vocabularies of the speaker dependent systems make them useful today for many tasks.

Moderate vocabulary size systems are available from a number of vendors that should have the capability of supporting educational as well as performance maintenance roles.

Speaker independent systems are currently limited to small vocabularies and are probably of insufficient robustness to support class or field applications. Watch this group - much work is being done and there will likely be some major progress in the short term.

Speaker dependent discrete systems are currently most noise tolerant. With proper design, they will likely be adequate for most class or field applications.

Speaker dependent connected recognizers are becoming much less noise sensitive. As this evolution proceeds, they will likely be perceived as more appropriate for all applications. There are many assumptions but no current evidence, however, that suggests that humans interact better with a connected speech recognizer.

The wide variation of software support provided by the vendors results in difficult "porting" of applications from system to system. A very useful research and development task would be the development of an "application generator" that not only supported a range of products from different vendors, but also encouraged the voice system integrator to fully consider the many application design issues (eg. prompting, help, editing, error recovery, etc.).

Finally, the technology is certainly mature enough to support both training and maintenance applications. In such a scenario, most of the voiced input would be commands, so even discrete speech recognizers would function well. The addition of an intelligent post-processor to further filter the input would likely reduce the potential impact of most recognizer errors to verification rather than editing or re-entry. Unfortunately, this post-processing function is not available in a generic format and will require application specific development. The basic premises of such a system are, however, known. It is suggested that the next step is to develop an application prototype and, through the prototype, define the requirements for an error detection and correction post processor.

XI. Glossary

Application generator - Software with the capability to automatically generate the necessary programs to support an application baseed on design requirements input.

Coarticulation - The phenomenon observed when pronouncing two words together results in the component sounds being changed.

Connected speech - Speaking words fully and distinctly with no unnatural pauses between them.

Continuous speech - Speech as typified by human to human speech. Words are often run on and sounds are missing.

DbA - Sound pressure measurement (in decibels) using the A weighting scale.

DbC - Sound pressure measurement (in decibels) using the C weighting scale.

Discrete speech - Speech in which each word is fully and distinctly pronounced with short pauses between each word.

Dynamic update - The process of updating speech recognition templates during performance without the need to enter some performance maintenance process.

Enrollment - The process of training the speech recognition system to the user's voice. Templates are extracted from prompted speech to be used for future comparison.

Form factor - Physical attributes of a system. Determines which host systems are compatable, ie. will the board fit?

Front end gain - Amplification applied to the signal provided by the microphone.

Front-end amplifier - Amplifier to provide front end gain.

Intensity - Amplitude of speech or noise usually related in dB.

L_{eq} - Equivalent or perceived loudness.

Loudness - Perceived intensity.

Match score - The degree to which an utterarce matches a stored template.

Multibus - Computer backplane standard.

PC bus - Computer backplane standard.

RS-232 - Communications protocol standard (this one is not always interpreted the same by all vendors).

Speaker dependent - Speech recognition in which the user must have enrolled speech patterns.

Speaker independent - Speech recognition in which utterences are identified using generic information.

Speech onset - The start of an utterance; nominally when the energy level increases above ambient.

Speech template - A pattern derived from speech against which future utterences will be compared.

Speech termination - The end of an utterence; nominally when the energy level returns to ambient.

Syntaxing - The specification of rules which identify the possible (allowed) sequence of words in the vocabulary.

Template - Stored pattern derived from speech during training against which future utterences are compared.

Token - Often used interchangably with template but usually is a template derived from a single utterance.

Transaction generator - Software that automatically generates the necessary programs to suppcrt application transactions. This is a subset of application generators.

Tukey test - A statistical test to isolate sources of significance from pooled information.

XII. References

Baker, J.M. (1982). The performing arts -- how to measure up. In D. S. Pallet (Ed.), Proceedings of the Workshop on Standardization for Speech I/O Technology. Gaithersberg, MD.: National Bureau of Standards.

Berger, E.H., Ward, W.D., Morrill, J.C., & Royster, L.H. (1986). Noise and hearing conservation manual. American Industrial Hygiene Association.

Black, J.W. (1950). The effect of room characteristics upon vocal intensity and rate. The Journal of the Acoustic Society of America, 22 (2), 174-176.

Borden, G.J. (1979). An interpretation of research on feedback interruption in speech. Brain and Language, 7, 307-319.

Doddington, G.R., & Schalk, T.B. (1981). Speech recognition: Turning theory to practice. IEEE Spectrum, 18, 26-32.

Draegert, G.L. (1951). Relationships between voice variables and speech intelligibility in high level noise. Speech Monographs, 18, 272-278.

Dunnett, C.W. (1980). Pairwise multiple comparisons in the homogeneous variance, unequal sample size cases. Journal of the American Statistical Association, 75, 372.

Garber, S.F., Siegel, G.M., Pick, H.L, & Alcorn, S.R. (1976). The influence of selected masking noises on Lombard and sidetone amplification effects. Journal of Speech and Hearing Research, 19, 523-535.

Goodnight, J.H. (1971). The new General Linear Models procedure. Proceedings of the First International SAS Users' Meeting

Howell, K., & Martin, A.M. (1975). An investigation of the effects of hearing protectors on vocal communication in noise. Journal of Sound Vibration, 41 (2), 181-196.

Kryter, K.D. (1946). Effects of ear protective devices on the intelligibility of speech in noise. Journal of the Acoustic Society of America, 18 (2), 413-417.

Lane, H. L. & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. Journal of Speech and Hearing Research, 14, 677-709.

Larson, N., Moody, T., & Joost, M. (1986). The effects of background noise on ASR performance using inertial and headset microphones. North Carolina State University Technical Report No. TR-IE-86-7.

Nusbaum, H.C., Davis, C.N., Pisoni, D.B., & Davis, E. (1986). Testing the performance of isolated uterance speech recognition devices. Proceedings of AVIOS '86, 393-408.

Pallett, D.S. (1985). Performance Assessment of Automatic Speech Recognizers. Journal of Research of the National Bureau of Standards, 90(5), (Sep./Oct.).

Peterson, A.P. (1980). Handbook of noise measurement. Concord, Mass.: GenRad, Inc.

Pisoni, D.B.; Bernacki, R.H., Nusbaum, H.C., & Yuchtman, M. (1985). Some acoustic-phonetic correlates of speech produced in noise. IEEE, 1581-1584.

Plice, G.W. (1983). Choosing a microphone. Speech Technology, 2, (Sept./Oct.), 17.

Rollins, A., & Wiesen, J. (1983). Speech recognition and noise. ICASSP, 523-526.

Sall, J.P. (1978). SAS regression applications. SAS Technical Report A-102, Raleigh, SAS Institute.

Searle, S.R. (1971). Linear Models. New York, John Wiley and Sons.

Sheffe, H. (1959). The Analysis of Variance. New York, John Wiley and Sons.

Siegel, G.M., & Pick, H.L. (1974). Auditory feedback in the regulation of voice. Journal of the Acoustic Society of America, 56(5), 1618-1624.

Tukey, J.W. (1952). Allowances for Various Types of Error Rates. Unpublished IMS address, Chicago, Illinois.

Waller, H.F. (1985). Choosing the right microphone for speech applications. Proceedings of Speech Tech '85. (p.45).

Webster, J.C., & Klumpp, R.G. (1962). Effects of ambient noise and nearby talkers on a face-to-face communication task. The Journal of the Acoustical Society of America, 34 (7), 936-941.

Appendix I. Vendor List

Product name, contact and address:

Dr. H. Mangold
AEG Telefunken Nachrichtentechnik GmBh
Postfach 1120 7150 Bachnang, West Germany

Description of Product Capabilities:

Speaker dependent or independent?
Type of speech:
Method of speech recognition:
Training Method:

Vocabulary Limitations

Number of words in active vocabulary:
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system:
Customer Support:

Product name, contact and address:

AT&T's Conversant 1 Voice System
Dr. Christopher D. Farrar
AT&T
6200 East Broad Street Columbus, Ohio 43213
614-860-3278 or 800-341-2272

Description of Product Capabilities:

Speaker dependent or independent? Both
Type of speech: Isolated and Connected
Method of speech recognition: template with phonetic enhancements
Training Method: 2 to 4 for Dep

Vocabulary Limitations

Number of words in active vocabulary: Dis Ind digits,yes,no,oh; Con Ind digits, yes, no; dep 256 words
Vocabularies in system: n/a
Word length limit: Dis - max 2.01 sec; others application dependent
Built in syntaxing: optional
Response time: 250 ms maximum to next prompt
Minimum time between utterances: for dep is programmable default 195 ms
Templates updated continuously: no

Compatibility of System

System compatibility: Stand alone; Unix operating system; asynchronous, bisynchronous 3270 & SNA/SDLC
Languages Supported: C
Programming required: none required

Microphone / Telephone information

Telephone access: yes
Recommended microphone and/or jack type:
telephone

Testing of ASR

Independent Tests: not available
Tests in noise: no specific testing but meant for telephone lines
Existing Applications: yes stock quotation,

Price and size information

Price: pricing on individual basis; volume discounts available to VAR's
Size of system: 25x22x15 ~100lbs
Customer Support: training and warranty

Product name, contact and address:

SSB-1000 Speech Recognition Board
Mr. Arthur V. Celona
AUDEC
299 Market Street Saddle Brook, New Jersey 07662
201-368-3848

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: discrete
Method of speech recognition: template
Training Method: one pass for enrollment, 2 additional training passes

Vocabulary Limitations

Number of words in active vocabulary: 144
Vocabularies in system:
Word length limit: 2 sec w/ 150 ms gap between words
Built in syntaxing: application dependent
Response time: 250-300 ms
Minimum time between utterances: 150 ms
Templates updated continuously: yes

Compatibility of System

System compatibility: any computer with RS-232 port or 8 bit parallel port. Can stand alone.
Languages Supported: Macro commands, 6502 assemble language, any resident language for host system
Programming required: not required

Microphone / Telephone information

Telephone access: yes with additional design
Recommended microphone and/or jack type:
none recommended

Testing of ASR

Independent Tests: none
Tests in noise: not defined
Existing Applications: Telephone, Remote equipment management, toys

Price and size information

Price: $250 with discounts for multiple purchase
Size of system: 5 in x 5 in; < one pound
Customer Support: yes

Product name, contact and address:

Philip T. Mclaughlin
Audopilot
19 Antoine Court Hunington, New York 11743
516-351-4862

Description of Product Capabilities:

Speaker dependent or independent?
Type of speech:
Method of speech recognition:
Training Method:

Vocabulary Limitations

Number of words in active vocabulary:
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system:
Customer Support:

Product name, contact and address:

Calltalk DVIO Model 100
Mr. J. Levenberg
Calltalk LTD
Hamasger 56 Tel-Aviv, Israel 67214

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: continuous speech
Method of speech recognition: templates
Training Method:

Vocabulary Limitations

Number of words in active vocabulary: 500 words
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time: less than 400 ms
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system: 17x6.5x22.5 55.5lbs
Customer Support: yes

Product name, contact and address:

Mr. Barry Cohen
CE Electronics
481 Eighth Avenue Suite 726 New York, New York 10001

Description of Product Capabilities:

Speaker dependent or independent?
Type of speech:
Method of speech recognition:
Training Method:

Vocabulary Limitations

Number of words in active vocabulary:
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system:
Customer Support:

Product name, contact and address:

Voicescribe 1000
Dr. Janet Baker
Dragon Systems, Inc.
55 Chapel Street Newton, MA. 02158
617-965-5200

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: isolated
Method of speech recognition: template
Training Method: train each word

Vocabulary Limitations

Number of words in active vocabulary: 1000
Vocabularies in system: n/a
Word length limit:
Built in syntaxing:
Response time: near real time
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility: IBM PC/XT or AT
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price: $200 minimum 1,000 units
Size of system:
Customer Support:

Product name, contact and address:

Voicescribe - 20000
Dr. Janet Baker
Dragon Systems, Inc.
55 Chapel Street Newton, MA. 02158
617-965-5200

Description of Product Capabilities:

Speaker dependent or independent? independent
Type of speech: isolated
Method of speech recognition: phonetic
Training Method: 30 minutes

Vocabulary Limitations

Number of words in active vocabulary: 20,000
Vocabularies in system: n/a
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility: IBM PC XT or AT
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price: $500 minimum 1,000 units
Size of system:
Customer Support:

Product name, contact and address:

Mr. Yasuo Sato
Fujitsu, Ltd.
1015 Kami-Odanaka Nakakara-ku, Kawasaki 211 Japan

Description of Product Capabilities:

Speaker dependent or independent?
Type of speech:
Method of speech recognition:
Training Method:

Vocabulary Limitations

Number of words in active vocabulary:
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system:
Customer Support:

Product name, contact and address:

Most activity is IRAD & CRAD in support of Defense Department - no product
Dr. John N. Damoulakis
Gould Electronics
40 Gould Center, Rolling Meadows, Ill. 60008
312-640-4400

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: isolated or connected
Method of speech recognition: template
Training Method: 1 to 5 times inserting individual words

Vocabulary Limitations

Number of words in active vocabulary: 256
Vocabularies in system: 256
Word length limit: minimum word length 0.1 sec.
Built in syntaxing: none
Response time: 200 ms at low noise; 500 ms at 0 dB SNR
Minimum time between utterances: 200 ms
Templates updated continuously: yes, environmentally adaptive

Compatibility of System

System compatibility: Special purpose stand alone; operational on VAX 11/780
Languages Supported: Fortran, C, Pascal
Programming required: none

Microphone / Telephone information

Telephone access: not tested yet
Recommended microphone and/or jack type:
flexible, minimum telephone bandwidth

Testing of ASR

Independent Tests: none
Tests in noise: many test completed in noise
Existing Applications: experimental and evaluation only at this time

Price and size information

Price: Quotation
Size of system: .35 ft cubed
Customer Support: customized products; support negotiated in cont

Product name, contact and address:

Mr. Akira Ichikawa
Hitachi, Ltd.
1-280 Higashi-Koigakubo Kokubunji, Tokyo 185, Japan

Description of Product Capabilities:

Speaker dependent or independent?
Type of speech:
Method of speech recognition:
Training Method:

Vocabulary Limitations

Number of words in active vocabulary:
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system:
Customer Support:

Product name, contact and address:

Voice Communication Adapter
Mr. Fred McNeese
IBM
IBM Entry Systems Division Boca Raton, Florida 33432

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: discrete
Method of speech recognition: template
Training Method: user defined, 4 utterances recommended

Vocabulary Limitations

Number of words in active vocabulary: 64
Vocabularies in system: up to 5
Word length limit: 2 seconds
Built in syntaxing: user defined
Response time: real time
Minimum time between utterances: "brief pause"
Templates updated continuously: no

Compatibility of System

System compatibility: IBM PC
Languages Supported: has transparent keyboard
Programming required: none required

Microphone / Telephone information

Telephone access: yes
Recommended microphone and/or jack type:
high impedance with 2.5mm connector

Testing of ASR

Independent Tests: yes
Tests in noise: yes
Existing Applications:

Price and size information

Price:
Size of system: board
Customer Support: yes

Product name, contact and address:

iSBC 570
Dan Fink
Intel Corp.
3065 Bowers Avenue Santa Clara, CA. 95051
408-987-8080

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: isolated
Method of speech recognition: template
Training Method: three training passes suggested

Vocabulary Limitations

Number of words in active vocabulary: 200
Vocabularies in system: n/a
Word length limit: up to 2 seconds
Built in syntaxing: user defined
Response time: real time
Minimum time between utterances: varied, user defined
Templates updated continuously: yes

Compatibility of System

System compatibility: Multibus channel, serial channel and local channel
Languages Supported: C
Programming required: speech transaction files

Microphone / Telephone information

Telephone access: no
Recommended microphone and/or jack type:
female jack for Shure SM-10 microphone

Testing of ASR

Independent Tests: yes
Tests in noise: yes
Existing Applications: yes

Price and size information

Price: unknown
Size of system: 6.5x17x22 (60 lbs.)
Customer Support: yes

Product name, contact and address:

Vocalink S4000
Mr. Brundage
Interstate Voice Products
1849 West Sequoia Ave Orange, CA. 92668
714-937-9010

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: continuous
Method of speech recognition: template
Training Method: system controlled

Vocabulary Limitations

Number of words in active vocabulary: 100 words
Vocabularies in system: multiple
Word length limit: .1 to 2.0 seconds 15 chars/word
Built in syntaxing: yes
Response time: < 300 ms
Minimum time between utterances: n/a
Templates updated continuously: no

Compatibility of System

System compatibility: PC DOS or MS DOS
Languages Supported: all supported by DOS
Programming required: grammar and translation files defined by user

Microphone / Telephone information

Telephone access: no
Recommended microphone and/or jack type:
headset or wireless

Testing of ASR

Independent Tests: yes
Tests in noise: yes
Existing Applications: yes

Price and size information

Price:
Size of system: 17x4x12 15 lbs.
Customer Support: yes

Product name, contact and address:

Multibus CSR
Richard C. Sadler
ITT Defense Communications Division
492 River Road Nutley, New Jersey 07110-3696
201-284-4234

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: both
Method of speech recognition: template
Training Method: system defined initially, but user defined # of utterances/word

Vocabulary Limitations

Number of words in active vocabulary: 40 active templates & 200 to 300 words
Vocabularies in system: 5 but is expandable to 30
Word length limit: n/a
Built in syntaxing: user programmable syntax 60 nodes 290 words/node
Response time: <.25 sec
Minimum time between utterances: n/a
Templates updated continuously: no

Compatibility of System

System compatibility: Venix/86 OS
Languages Supported: C and assembly
Programming required: system defined for grammar

Microphone Telephone information

Telephone access no
Recommended microphone and/or jack type
application dependent

Testing of ASR

Independent Tests yes
Tests in noise yes
Existing Applications yes

Price and size information

Price $37,000
Size of system 1600 in cubed 45 lbs
Customer Support yes

Product name, contact and address:

Kurzweil VoiceSystems
Mr. Bob Joseph
Kurzweil Applied Intelligence, Inc.
411 Waverly Oaks Road Waltham, Ma. 02154
617-893-5151

Description of Product Capabilities:

Speaker dependent or independent? dependent, limited independent
Type of speech: isolated
Method of speech recognition: template plus other proprietary algorithms
Training Method: one to three times for each utterance

Vocabulary Limitations

Number of words in active vocabulary: 1000
Vocabularies in system: multiple
Word length limit: up to several seconds
Built in syntaxing: optional, user developed
Response time: <.5 sec
Minimum time between utterances: 60-180 ms
Templates updated continuously: no

Compatibility of System

System compatibility: IBM PC compatible, connects to ASCII & 3270 hosts w/o modification to host s/w
Languages Supported: KVS libraries written in C can be linked w/ objects produced by other languages
Programming required: none required

Microphone / Telephone information

Telephone access: limited
Recommended microphone and/or jack type:
5 pin DIN connector, headset & handset available

Testing of ASR

Independent Tests: yes
Tests in noise: reliable in high continuous noise environments
Existing Applications: yes

Price and size information

Price: KVS-AA $6500, volume discounts available
Size of system: 14x6.5x8 18 lbs
Customer Support: yes

Product name, contact and address:

Voice-Macros
Ellen L. Clark
Microphonics
25-37th St. N.E. Suite B   Auburn, Wa. 98002
206-939-2321   800-325-9206

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: discrete
Method of speech recognition: template
Training Method: 1 pass

Vocabulary Limitations

Number of words in active vocabulary: 128
Vocabularies in system:
Word length limit: 2 seconds
Built in syntaxing: no
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility: IBM PC,XT,AT
Languages Supported: DOS compatible
Programming required: DOS compatible

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise: yes
Existing Applications:

Price and size information

Price:
Size of system: board
Customer Support:

Product name, contact and address:

Mr. Jun Oyamada
NEC America Inc.
8 Old Sod Farm Road Melville, New York 11747
516-753-7000

Description of Product Capabilities:

Speaker dependent or independent? dependent, limited independent
Type of speech: isolated and connected
Method of speech recognition: template
Training Method:

Vocabulary Limitations

Number of words in active vocabulary: 500 words
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise: 80 - 85 dB
Existing Applications:

Price and size information

Price: $599 - $9,995
Size of system:
Customer Support: yes

Product name, contact and address:

Austin Bordeaux
RDA Logicon R&D Associates
P.O. Box 9695 4640 Admiralty Way Marina Del Ray, CA. 90295
213-822-1715

Description of Product Capabilities:

Speaker dependent or independent?
Type of speech:
Method of speech recognition:
Training Method:

Vocabulary Limitations

Number of words in active vocabulary:
Vocabularies in system:
Word length limit:
Built in syntaxing:
Response time:
Minimum time between utterances:
Templates updated continuously:

Compatibility of System

System compatibility:
Languages Supported:
Programming required:

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests:
Tests in noise:
Existing Applications:

Price and size information

Price:
Size of system:
Customer Support:

Product name, contact and address:

Coretechs VET 3 Speech Terminal
Wayne Laffitte
Scott Instruments Corp.
1111 Willow Springs Drive Denton, Texas 76201
817-387-9514

Description of Product Capabilities:

Speaker dependent or independent? both
Type of speech: connected and discrete
Method of speech recognition: template (unique representation of spoken word)
Training Method: 1 pass

Vocabulary Limitations

Number of words in active vocabulary: 200 1/2 sec words and 100 recordings
Vocabularies in system: 1
Word length limit: 3 seconds discrete 8 seconds connected
Built in syntaxing: yes
Response time: .25 seconds - software selectable
Minimum time between utterances: .25 seconds - software selectable
Templates updated continuously: no

Compatibility of System

System compatibility: any computer w/ RS232 communications
Languages Supported: any
Programming required: none required

Microphone / Telephone information

Telephone access: yes
Recommended microphone and/or jack type:
Hirose HR10-7R-6

Testing of ASR

Independent Tests: yes, resulted in purchase of Scott VET 3
Tests in noise: up to 110 db noise
Existing Applications: QC/QA data gathering

Price and size information

Price: $8995.00 list , VAR and distributor pricing available
Size of system: desktop 17.6x16.3x4.5 16 lbs; rack mount 19.0x16.3x5.22
Customer Support: yes

Product name, contact and address:

SSI's Phonetic Engine
Leonard L. Backus/Deana J. Murchison (818) 881-0885
Speech Systems Inc.
18356 Oxnard Street Tarzana, California 91356
617-639-2360

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: continuous
Method of speech recognition: phonetic
Training Method: 20 minutes, optional (increases accuracy)

Vocabulary Limitations

Number of words in active vocabulary: 5000
Vocabularies in system: all can be accessed
Word length limit: n/a
Built in syntaxing: yes
Response time: phonetics produced in real time
Minimum time between utterances: n/a
Templates updated continuously: no, templates not used

Compatibility of System

System compatibility: Phonetic process software in C. Currently on VAX and SUN systems
Languages Supported:
Programming required: User inputs to syntax and dictionary utilizing SSI tools

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:
proprietary handset/telephone type

Testing of ASR

Independent Tests:
Tests in noise no
Existing Applications: development applications include command & control, limited dictation, AI

Price and size information

Price: depends on configuration of development and system
Size of system:
Customer Support: yes, cost will be minimal

Product name, contact and address:

TI-Speech Development System
Mr. Doug Palmer
Texas Instruments Inc. M/S 2081
P.O. Box 2909 Austin, Texas 78769
512-250-6005

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: both
Method of speech recognition: template
Training Method: system defined

Vocabulary Limitations

Number of words in active vocabulary: 50 words
Vocabularies in system: 1000 words total
Word length limit: n/a
Built in syntaxing: user defined
Response time: real time
Minimum time between utterances: n/a
Templates updated continuously: no

Compatibility of System

System compatibility: IBM PC and TI
Languages Supported: MS-Basic, MS-Pascal, Lattice C, IQ Lisp, Compiled Basic
Programming required: grammar structures

Microphone / Telephone information

Telephone access: available
Recommended microphone and/or jack type:
1/4 inch jack - hand held mike

Testing of ASR

Independent Tests: yes
Tests in noise: yes
Existing Applications:

Price and size information

Price: $1155
Size of system: board
Customer Support:

Product name, contact and address:

TOSVOICE
Dr. Sadakazu Watanabe
Toshiba Corp.
1,Komukai Toshibacho,Saiwai-Ku,Kawasaki-City,Kanagawa,210,Japan
Kawasaki 044-511-2111

Description of Product Capabilities:

Speaker dependent or independent? independent
Type of speech: discrete
Method of speech recognition: both are used
Training Method: n/a

Vocabulary Limitations

Number of words in active vocabulary: 64
Vocabularies in system: 64
Word length limit: 4 sec
Built in syntaxing: optional
Response time: 200 msec
Minimum time between utterances: i sec
Templates updated continuously: n/a

Compatibility of System

System compatibility: DOS; PL-40
Languages Supported: PL-40, Fortran
Programming required: none

Microphone / Telephone information

Telephone access: yes
Recommended microphone and/or jack type:
telephone; Shure SM12; canon connector

Testing of ASR

Independent Tests: yes
Tests in noise: 75 dB(A)
Existing Applications: yes

Price and size information

Price: unknown
Size of system: 500x900x900 mm; 50 Kg
Customer Support: no

Product name, contact and address:

Verbex Series 4000
Mr. Chris Seelbach
Verbex/Voice Industries Corp.
10 Madison Ave. , Morristown, New Jersey 07960
201-267-7507

Description of Product Capabilities:
Speaker dependent or independent? dependent
Type of speech: continuous
Method of speech recognition: template
Training Method: system defined

Vocabulary Limitations

Number of words in active vocabulary: 100 words
Vocabularies in system: multiple
Word length limit: 15 characters
Built in syntaxing: yes
Response time: <300 ms
Minimum time between utterances: n/a
Templates updated continuously: no

Compatibility of System

System compatibility: IBM PC compatible
Languages Supported: all supported by DOS
Programming required: grammars and translation tables

Microphone / Telephone information

Telephone access: no
Recommended microphone and/or jack type:
Shure VR- 230

Testing of ASR

Independent Tests: yes
Tests in noise: yes
Existing Applications: yes

Price and size information

Price: unknown
Size of system: 17x4x12 15 lbs
Customer Support: yes

Product name, contact and address:

VCS Technology
Dr. R.E. Helms
Voice Control Systems
16610 Dallas Parkway, Dallas, Texas 75248
214-248-8244

Description of Product Capabilities:

Speaker dependent or independent? independent
Type of speech: discrete
Method of speech recognition: phonetic
Training Method: n/a

Vocabulary Limitations

Number of words in active vocabulary: 20
Vocabularies in system: 1 kbyte/vocabulary words
Word length limit: 1.5 seconds
Built in syntaxing: optional
Response time: 250 msec
Minimum time between utterances: n/a
Templates updated continuously: n/a

Compatibility of System

System compatibility: stand alone
Languages Supported: application specific
Programming required: application specific

Microphone / Telephone information

Telephone access: yes
Recommended microphone and/or jack type:
application-specific

Testing of ASR

Independent Tests: unknown
Tests in noise: yes- specific versions have been developed
Existing Applications: yes

Price and size information

Price: cost to produce is app. $100
Size of system: 35 square inches
Customer Support: yes

Product name, contact and address:

VPC 2100
Mr. Bruce Ryon
Votan
4487 Technology Drive, Freemont, CA. 94538-6343
415-490-7600

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: both
Method of speech recognition: template
Training Method: discrete words; 2 utterances recommended; can extract continuous phrases

Vocabulary Limitations

Number of words in active vocabulary: 80 (more with fewer training passes or optional expantion vocabulary
Vocabularies in system: n/a
Word length limit: n/a
Built in syntaxing: no, available through vocabulary subsets
Response time: real time
Minimum time between utterances: n/a
Templates updated continuously: no

Compatibility of System

System compatibility: IBM PC's and compatibles
Languages Supported: C 86
Programming required: none

Microphone / Telephone information

Telephone access: yes
Recommended microphone and/or jack type: gooseneck and handheld

Testing of ASR

Independent Tests: yes
Tests in noise: yes
Existing Applications: yes

Price and size information

Price:
Size of system: board
Customer Support: yes

Product name, contact and address:

Series 100 Voice Data Collection System
W.A. Hardister
Westinghouse Electric Corporation
Southern Regional Office One Knollwood Place, Asheville, N.C. 28804
704-645-4321

Description of Product Capabilities:

Speaker dependent or independent? dependent
Type of speech: continuous
Method of speech recognition: template
Training Method:

Vocabulary Limitations

Number of words in active vocabulary: 200
Vocabularies in system:
Word length limit: n/a
Built in syntaxing: yes
Response time: real time
Minimum time between utterances: n/a
Templates updated continuously: no

Compatibility of System

System compatibility
Languages Supported: n/a
Programming required: system defined

Microphone / Telephone information

Telephone access:
Recommended microphone and/or jack type:

Testing of ASR

Independent Tests
Testing in noise: up to [illegible] dB
Existing Applications: yes

Price and size information

Price:
Size of system: [illegible] lbs
Customer Support: yes

Product name, contact and address:

Seraphine
Mr. Herve Couturier
XCOM
BP 29 Montbonnot Saint Martin, Saint-Ismier, France 38330
76-52-00-46

Description of Product Capabilities:

Speaker dependent or independent? both
Type of speech: discrete and connected
Method of speech recognition: templates for each speaker
Training Method: single pass for individual words

Vocabulary Limitations

Number of words in active vocabulary: 100 words
Vocabularies in system: 100->200 words
Word length limit: 6 seconds; up to 7 words
Built in syntaxing: yes
Response time: size and syntax dependent; 1 sec for 0 to 999 recognition
Minimum time between utterances: 1 sec
Templates updated continuously: no

Compatibility of System

System compatibility: stand alone system - Multibus and RS232C
Languages Supported: all
Programming required: no programming required for test

Microphone / Telephone information

Telephone access: under study
Recommended microphone and/or jack type:
Shure SM10

Testing of ASR

Independent Tests: no
Tests in noise: no
Existing Applications: yes

Price and size information

Price: $3,000
Size of system: SBC board or 445x300x70 mm cabinet
Customer Support: yes, free in France

Appendix II. Discreet Speech Vocabulary

Vocabulary Words - Scenario 1

1. Zero
2. One
3. Two
4. Three
5. Four
6. Five
7. Six
8. Seven
9. Eight
10. Nine
11. Yes
12. No
13. Up
14. Down
15. Right
16. Left

Appendix III. Connected Speech Vocabularies

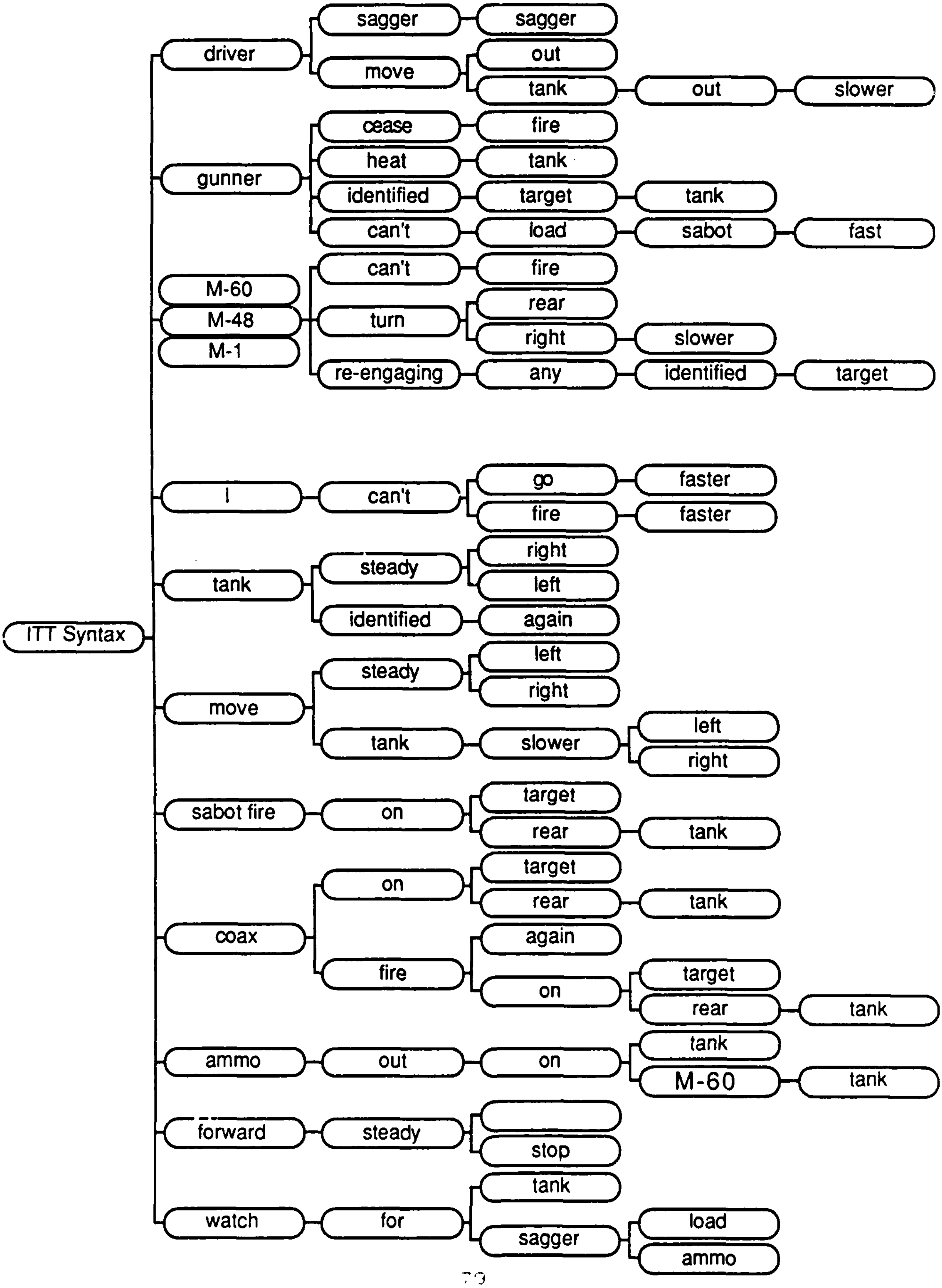
Vocabulary Words - Scenario 2

1. Driver
2. Move
3. Out
4. Sagger
5. Gunner
6. Cease
7. Fire
8. Heat
9. Tank
10. Steady
11. Right
12. Left
13. Coax
14. Can't
15. Go
16. Faster
17. For
18. Re-engaging
19. Fast
20. Again
21. M-48
22. M-60
23. Turn
24. Rear
25. Identified
26. Sabot
27. On
28. Target
29. M-1
30. Slower
31. I
32. Ammo
33. Forward
34. Stop
35. Watch
36. Load
37. Any

Vocabulary Words - Scenario 3

1. Part
2. Number
3. Has
4. Failed
5. How
6. Many
7. Of
8. Are
9. In
10. Stock
11. Which
12. Replaces
13. What
14. Repair
15. Procedures
16. Tool
17. Is
18. Required
19. To
20. Install
21. Zero
22. One
23. Two
24. Three
25. Four
26. Five
27. Six
28. Seven
29. Eight
30. Nine

Appendix IV. ITT Syntax for Scenario 2



Appendix V. Other Connected Syntax (Verbex, Interstate, and TI)

Syntax Structure for Scenario 2 (except ITT)

First	Second	Third
Driver	Move	Out
Gunner	Sagger	Sagger
Tank	Cease	Fire
Move	Heat	Tank
Coax	Steady	Right
Can't	Fire	Left
M-48	Go	Again
M-60	Can't	Faster
Sabot	Turn	Rear
M-1	Identified	Target
I	On	Slower
Fire	Tank	On
Ammo	Out	Steady
Forward	For	Any
Watch	Re-engaging	Load

Fourth Word	Fifth Word
Slower	Tank
Right	Slower
Faster	Stop
Target	Ammo
Tank	Target
M-60	Fast
Out	
Steady	
Load	
Identified	
Sabot	

1. First Word -> Second Word -> Third Word
2. First Word -> Second Word -> Third Word -> Fourth Word
3. First Word -> Second Word -> Third Word -> Fourth Word -> Fifth Word

Appendix VI. Test Sentences

Sentence List - Scenario 2

Three Words

1. Driver move out
2. Driver sagger sagger
3. Gunner cease fire
4. Gunner heat tank
5. Tank steady right
6. Move steady left
7. Coax fire again
8. Can't go faster
9. M-48 can't fire
10. M-60 turn rear
11. Tank identified again
12. Coax on target

Four Words

13. M-1 turn right slower
14. Move tank slower right
15. I can't fire faster
16. Coax fire on target
17. Fire on rear tank
18. Gunner identified target tank

Five Words

19. Ammo out on M-60 tank
20. Driver move tank out slower
21. Forward steady steady steady stop
22. Watch for sagger load ammo
23. M-1 re-engaging any identified target
24. Gunner can't load sabot fast

Sentence List - Senario 3

Basic Sentences

1. Part #_______________ has failed.
2. How many of part #_______________ are in stock.
3. Which part replaces Part # ___________________.
4. Part #_________________ has failed what are repair procedures.
5. What tool is required to install part # _______________.

Increasingly longer sequences of numbers were used.

Scenario 3 - Actual Sentences Tested

1. Part number six has failed.
2. How many of part number nine are in stock
3. Which part replaces part number two.
4. Part number four has failed what are repair procedures
5. What tool is required to install part number seven.
6. Part number two eight has failed.
7. How many of part number three nine are in stock.
8. Which part replaces part number seven four.
9. Part number one six has failed what are repair procedures.
10. What tool is required to install part number zero five.
11. Part number seven six one has failed.
12. How many of part number zero two four are in stock.
13. Which part replaces part number three five eight.
14. Part number nine two two has failed what are repair procedures.
15. What tool is required to install part number nine nine one.
16. Part number six three two one has failed.
17. How many of part number four four six six are in stock.
18. Which part replaces part number eight seven eight three.
19. Part number six six one one has failed what are repair procedures.
20. What tool is required to install part number two two two eight.
21. Part number seven eight three three seven has failed.
22. How many of part number nine four zero zero nine are in stock.
23. Which part replaces part number nine seven seven three three.
24. Part number one two six six two has failed what are repair procedures.
25. What tool is required to install part number zero one one nine four.

Appendix VII. Rejection / Misrecognition Matricies: Discrete Task

Rejections - Discrete Scenario

Recognizer	Noise	0	1	2	3	4	5	6	7	8	9	YES	NO	UP	DOWN	RIGHT	LEFT	TOTAL	$\bar{X}$
Verbex	None	2	2	1	6	1	1	1		5	2	4	2	2	1	12	11	53	.08
	Industrial	14	21	4	19	15	4	23	16	13	10	9	15	25	8	25	11	232	.36
	Fast Food Restaurant	8	14	2	10	6	11	23	16	14	5	11	5	20	6	21	21	193	.30
IN4000	None	5	4		8	1	5	1	4	4		1	2	6	1	16	10	68	.11
	Industrial	15	23	10	15	12	18	12	17	20	14	19	11	26	9	29	23	273	.43
	Fast Food Restaurant	16	20	2	15	16	4	25	10	13	10	14	12	24	16	30	18	245	.38
VOTAN	None																	0	0.00
	Industrial																	0	0.00
	Fast Food Restaurant																	0	0.00
TI	None	3		2				2		1								8	.01
	Industrial			3			1	1	4	1							3	13	.02
	Fast Food Restaurant	1	1	3					1				2		3			11	.02
IBM	None																	0	0.00
	Industrial	3	4	3	3	3	5	2	4	5	2	1	3	3	3	3	4	51	.08
	Fast Food Restaurant																	0	0.00
INTEL	None					3		1		1	1							6	.01
	Industrial								1	1						1		3	.00
	Fast Food Restaurant		3	1		1		1	1								1	8	.01
INCSRB	None	1	1						3					6	1	6	3	21	.03
	Industrial		2	1	1	1	3	1	1		7	1		5	3	6	4	36	.06
	Fast Food Restaurant		6		1	1	8	2	2	1	5		2	8		3	3	42	.07
KVS	None	1					2			1	1		2	1			2	10	.02
	Industrial	6	4	7	1	4	3	21	9	8	3	9	1	19	2	2	8	107	.17
	Fast Food Restaurant		6	7	2	11		17	13	5	3	4		11	1		5	85	.13
ITT	None																	0	0.00
	Industrial																	0	0.00
	Fast Food Restaurant																	0	0.00

Misrecognition - Discrete Scenario

Recognizer	Noise	0	1	2	3	4	5	6	7	8	9	YES	NO	UP	DOWN	RIGHT	LEFT	TOTAL	X̄
Verbex	None													1				1	.001
	Industrial			1									1					2	.0
	Fast Food Restaurant					1												1	.001
IN4000	None																	0	0.0
	Industrial					4												4	.0
	Fast Food Restaurant										1							1	.001
VOTAN	None													2				2	.0
	Industrial																4	4	.0
	Fast Food Restaurant		1											1	1	1		4	.0
TI	None									1								1	.001
	Industrial																	0	0.0
	Fast Food Restaurant												1					1	.001
IBM	None		4	1		6								7		1		19	.0
	Industrial	1				3			1		3	1			1			10	.0
	Fast Food Restaurant	1				5		1			6							13	.0
INTEL	None												1					1	.001
	Industrial												1					1	.001
	Fast Food Restaurant																	0	0.0
INCSRB	None						1				1			1				4	.0
	Industrial	3			4		2		2	3		2	1	4		4		26	.0
	Fast Food Restaurant	1	1					1	1					3		5		12	.0
KVS	None																	1	.001
	Industrial		3	1	1		2	2	1		1	1	1	2	1			16	.0
	Fast Food Restaurant		1	4			3	4	2		5	1	1	2	1	1		30	.0
ITT	None															1		1	.001
	Industrial			1														2	.0
	Fast Food Restaurant			1										1				2	.01

Appendix VIII. Tukey Analysis of Means: Discrete Task

Recognizer	Grouping A	 B	 C	Mean	N	Mean # of Correct Utterances
ITT	*			159.58	12	.997
VOTRAN	*			159.17	12	.995
INTEL	*			158.42	12	.990
TI	*			157.17	12	.982
IBM	*	*		152.25	12	.952
INTERCSRB	*	*		148.25	12	.927
KURZWEIL		*		139.17	12	.870
VERBEX			*	119.67	12	.748
INTERSTATE4000			*	110.75	12	.692

Appendix IX. Tukey Analysis of Noise Effects: Discrete Task

	GROUPING				# CORRECT UTTERANCES
NOISE	A	B	MEAN	N	#TOTAL UTTERANCES
NONE	*		154.53	36	.966
	*				
FAST FOOD		*			
RESTAURANT		*	142.00	36	.888
		*			
INDUSTRIAL		*	138.28	36	.864

Appendix X. Tukey Analysis of Recognizer * Noise: Discrete Task

		GROUPING							
RECOGNIZER	NOISE	A	B	C	D	E	F	MEAN	N
ITT	NONE	*						159.75	4
ITT	RESTAURANT	*	*					159.50	4
ITT	INDUSTRIAL	*	*					159.50	4
VOTAN	NONE	*	*					159.50	4
VOTAN	RESTAURANT	*	*					159.00	4
INTEL	NONE	*	*					159.00	4
VOTAN	INDUSTRIAL	*	*					159.00	4
INTEL	INDUSTRIAL	*	*					158.25	4
INTEL	RESTAURANT	*	*					158.00	4
TI	NONE	*	*					157.75	4
TI	RESTAURANT	*	*	*				157.00	4
KURZWEIL	NONE	*	*	*				157.00	4
TI	INDUSTRIAL	*	*	*				156.75	4
IBM	RESTAURANT	*	*	*				156.75	4
IBM	NONE	*	*	*				155.25	4
INTERSTATE CSRB	NONE	*	*	*				153.75	4
VERBEX	NONE	*	*	*				146.50	4
INTERSTATE CSRB	RESTAURANT	*	*	*				146.50	4
IBM	INDUSTRIAL	*	*	*				144.75	4
INTERSTATE CSRB	INDUSTRIAL	*	*	*				144.50	4
INTERSTATE 4000	NONE	*	*	*				143.00	4
KURZWEIL	RESTAURANT		*	*	*			131.25	4
KURZWEIL	INDUSTRIAL			*	*	*		129.25	4
VERBEX	RESTAURANT				*	*	*	111.50	4
VERBEX	INDUSTRIAL					*	*	101.00	4
INTERSTATE 4000	RESTAURANT						*	98.50	4
INTERSTATE 4000	INDUSTRIAL						*	90.75	4

Appendix XI. Confusion and Error Matricies: Discrete Task

Utterance	Word Recognized															
	0	1	2	3	4	5	6	7	8	9	YES	NO	UP	DOWN	RIGHT	LEFT	[illegible]
Zero							3					1					1
One					1												4
Two																	1
Three			1				1		1						1		2
Four																	2
Five													1		1		11
Six											1						3
Seven			1				1				1						5
Eight			1				3										1
Nine																	12
Yes							2										1
No													2			1	2
Up				2		3										2	19
Down																	4
Right							1		1		1		1			5	15
Left							1						1				10

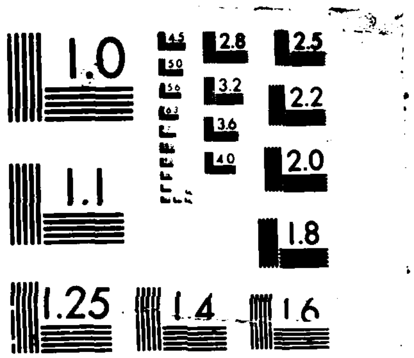
Discrete Utterance Confusion Matrix: Interstate CSRB

Utterance	Word Recognized																
	0	1	2	3	4	5	6	7	8	9	YES	NO	UP	DOWN	RIGHT	LEFT	REJ
Zero																	7
One					2	1					1						10
Two							4				1						14
Three											1						3
Four																	15
Five										3			2				5
Six				5					1								38
Seven												2		1			22
Eight																	14
Nine						1						1	2	1	1		7
Yes																2	13
No										1			1				3
Up		1								2		1					31
Down												1	1				3
Right					1												2
Left								1			1		3	1			15

Discrete Utterance Confusion Matrix: Kurzweil

AD-A179 762 AN EVALUATION OF SPEECH RECOGNITION TECHNOLOGY(U) 2/2
BATTELLE MEMORIAL INST RESEARCH TRIANGLE PARK NC
M G JOOST ET AL. 05 DEC 86 DAAG29-81-D-0100

UNCLASSIFIED NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

Utterance	Word Recognized																
	0	1	2	3	4	5	6	7	8	9	YES	NO	UP	DOWN	RIGHT	LEFT	REJ
Zero												2					3
One										1				1		2	4
Two				1													3
Three																	3
Four		3										9	1			1	3
Five																	5
Six			1														2
Seven																1	4
Eight																	5
Nine		3				5						1					2
Yes							1										1
No																	3
Up												4			1	2	3
Down										1							3
Right																1	3
Left																	4

Discrete Utterance Confusion Matrix: IBM

Appendix XII. Tukey Analysis of Speaker Effects: Scenario 2

```
TUKEY'S STUDENTIZED RANGE  (HSD)  TEST  FOR  VARIABLE:    CORRECT
RECOGNITION (COR)

ALPHA=0.05 CONFIDENCE=0.95 DF=24 MSE-64.4803
CRITICAL VALUE OF STUDENTIZED RANGE=4.166

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY  '***'


                SIMULTANEOUS                      SIMULTANEOUS
                    LOWER        DIFFERENCES          UPPER
  SUBJECT        CONFIDENCE        BETWEEN         CONFIDENCE
COMPARISON          LIMIT           MEANS             LIMIT

2   -   1          -4.529           4.050           12.629
2   -   3          -2.953           5.441           13.835
2   -   4           0.054           8.633           17.213   ***

1   -   2         -12.629          -4.050            4.529
1   -   3          -7.477           1.391           10.259
1   -   4          -4.460           4.583           13.627

3   -   2         -13.835          -5.441            2.953
3   -   1         -10.259          -1.391            7.477
3   -   4          -5.675           3.192           12.060

4   -   2         -17.213          -8.633           -0.054   ***
4   -   1         -13.627          -4.583            4.460
4   -   3         -12.060          -3.192            5.675
```

Appendix XIII. Tukey Analysis of Speaker * Recognizer: Scenario 2

RECOGNIZER * SUBJECT COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
ITT #2 - ITT #1	-25.469	0.000	25.469	
ITT #2 - ITT #4	-25.469	0.000	25.469	
ITT #2 - ITT #3	-23.469	2.000	27.469	
ITT #2 - VER #2	-13.136	12.333	37.802	
ITT #2 - INT #1	-8.469	17.000	42.469	
ITT #2 - VOT #2	-8.136	17.333	42.802	
ITT #2 - INT #4	-4.802	20.667	46.136	
ITT #2 - VER #1	-4.136	21.333	46.802	
ITT #2 - VER #3	-3.802	21.667	47.136	
ITT #2 - INT #3	-5.975	22.500	50.975	
ITT #2 - INT #2	0.198	25.667	51.136	***
ITT #2 - VER #4	2.198	27.667	53.136	***
ITT #2 - TI #2	3.531	29.000	54.469	***
ITT #2 - TI #3	1.525	30.000	58.475	***
ITT #2 - VOT #3	12.531	38.000	63.469	***
ITT #2 - VOT #1	19.864	45.333	70.802	***
ITT #2 - VOT #4	28.198	53.667	79.136	***
ITT #1 - ITT #2	-25.469	0.000	25.469	
ITT #1 - ITT #4	-25.469	0.000	25.469	
ITT #1 - ITT #3	-23.469	2.000	27.469	
ITT #1 - VER #2	-13.136	12.333	37.802	
ITT #1 - INT #1	-8.469	17.000	42.469	
ITT #1 - VOT #2	-8.136	17.333	42.802	
ITT #1 - INT #4	-4.802	20.667	46.136	
ITT #1 - VER #1	-4.136	22.500	50.975	
ITT #1 - INT #3	-3.802	21.667	47.136	
ITT #1 - VER #3	-5.975	22.500	50.975	
ITT #1 - INT #2	0.198	25.667	51.136	***
ITT #1 - VER #4	2.198	27.667	53.136	***
ITT #1 - TI #2	3.531	29.000	54.469	***
ITT #1 - TI #3	1.525	30.000	58.475	***
ITT #1 - VOT #3	12.531	38.000	63.469	***
ITT #1 - VOT #1	19.864	45.333	70.802	***
ITT #1 - VOT #4	28.198	53.667	79.136	***
ITT #4 - ITT #2	-25.469	0.000	25.469	
ITT #4 - ITT #1	-25.469	0.000	25.469	
ITT #4 0 ITT #3	-23.469	2.000	27.469	
ITT #4 - VER #2	-13.136	12.333	37.802	

RECOGNIZER * SUBJECT COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
ITT #4 - INT #1	-8.469	17.000	42.469
ITT #4 - VOT #2	-8.136	17.333	42.802
ITT #4 - INT #4	-4.802	20.667	46.136
ITT #4 - VER #1	-4.136	21.333	46.802
ITT #4 - VER #3	-3.802	21.667	47.136
ITT #4 - INT #3	-5.975	22.500	50.975
ITT #4 - INT #2	0.198	25.667	51.136 ***
ITT #4 - VER #4	2.198	27.667	53.136 ***
ITT #4 - TI #2	3.531	29.000	54.469 ***
ITT #4 - TI #3	1.525	30.000	58.475 ***
ITT #4 - VOT #3	12.531	38.000	63.469 ***
ITT #4 - VOT #1	19.864	45.333	70.802 ***
ITT #4 - VOT #4	28.198	53.667	79.136 ***
ITT #3 - ITT #2	-27.469	-2.000	23.469
ITT #3 - ITT #1	-27.469	-2.000	23.469
ITT #3 - ITT #4	-27.469	-2.000	23.469
ITT #3 - VER #2	-13.136	10.333	35.802
ITT #3 - INT #1	-10.469	15.000	40.469
ITT #3 - VOT #2	-10.136	15.333	40.802
ITT #3 - INT #4	-6.802	18.667	44.136
ITT #3 - VER #1	-6.136	19.333	44.802
ITT #3 - VER #3	-5.802	19.667	45.136
ITT #3 - INT #3	-7.975	20.500	48.975
ITT #3 - INT #2	-1.802	23.667	49.136
ITT #3 - VER #4	0.198	25.667	51.136 ***
ITT #3 - TI #2	1.531	27.000	52.469 ***
ITT #3 - TI #3	-0.475	28.000	56.475
ITT #3 - VOT #3	10.531	36.000	61.469 ***
ITT #3 - VOT #1	17.864	43.333	68.802 ***
ITT #3 - VOT #4	26.198	51.667	77.136 ***
VER #2 - ITT #2	-37.802	-12.333	13.136
VER #2 - ITT #1	-37.802	-12.333	13.136
VER #2 - ITT #4	-37.802	-12.333	13.136
VER #2 - ITT #3	-35.802	-10.333	15.136
VER #2 - INT #1	-20.802	4.667	30.136
VER #2 - VOT #2	-20.469	5.000	30.469
VER #2 - INT #4	-17.136	8.333	33.802
VER #2 - VER #1	-16.469	9.000	34.469
VER #2 - VER #3	-16.136	9.333	34.802
VER #2 - INT #3	-18.309	10.167	38.642
VER #2 - INT #2	-12.136	13.333	38.802
VER #2 - VER #4	-10.136	15.333	40.802
VER #2 - TI #2	-8.802	16.667	42.136

	SIMULTANEOUS		SIMULTANEOUS	
RECOGNIZER	LOWER	DIFFERENCES	UPPER	
* SUBJECT	CONFIDENCE	BETWEEN	CONFIDENCE	
COMPARISON	LIMIT	MEANS	LIMIT	
VER #2 - TI #3	-10.809	17.667	46.142	
VER #2 - VOT #3	0.198	25.667	51.136	***
VER #2 - VOT #1	7.531	33.000	58.469	***
VER #2 - VOT #4	15.864	41.333	66.802	***
INT #1 - ITT #2	-42.469	-17.000	8.469	
INT #1 - ITT #1	-42.469	-17.000	8.469	
INT #1 - ITT #4	-42.469	-17.000	8.469	
INT #1 - ITT #3	-40.469	-15.000	8.469	
INT #1 - VER #2	-30.136	-4.667	10.469	
INT #1 - VOT #2	-25.136	0.333	25.802	
INT #1 - INT #4	-21.802	3.667	29.136	
INT #1 - VER #1	-21.136	4.333	29.802	
INT #1 - VER #3	-20.802	4.667	30.136	
INT #1 - INT #3	-22.975	5.500	33.975	
INT #1 - INT #2	-16.802	8.667	34.136	
INT #1 - VER #4	-14.802	10.667	36.136	
INT #1 - TI #2	-13.469	12.000	37.469	
INT #1 - TI #3	-15.475	13.000	41.475	
INT #1 - VOT #3	-4.469	21.000	46.469	
INT #1 - VOT #1	2.864	28.333	53.802	***
INT #1 - VOT #4	11.198	36.667	62.136	***
VOT #2 - ITT #2	-42.802	-17.333	8.136	
VOT #2 - ITT #1	-42.802	-17.333	8.136	
VOT #2 - ITT #4	-42.802	-17.333	8.136	
VOT #2 - ITT #3	-40.802	-15.333	10.136	
VOT #2 - VER #2	-30.469	-5.000	20.469	
VOT #2 - INT #1	-25.802	-0.333	25.136	
VOT #2 - INT #4	-22.136	3.333	28.802	
VOT #2 - VER #1	-21.469	4.000	29.469	
VOT #2 - VER #3	-21.136	4.333	29.802	
VOT #2 - INT #3	-23.309	5.167	33.642	
VOT #2 - INT #2	-17.136	8.333	33.802	
VOT #2 - VER #4	-15.136	10.333	35.802	
VOT #2 - TI #2	-13.802	11.667	37.136	
VOT #2 - TI #3	-15.809	12.667	41.142	
VOT #2 - VOT #3	-4.802	20.667	46.136	
VOT #2 - VOT #1	2.531	28.000	53.469	***
VOT #2 - VOT #4	10.864	36.333	61.802	***
INT #4 - ITT #2	-46.136	-20.667	4.802	
INT #4 - ITT #1	-46.136	-20.667	4.802	
INT #4 - ITT #4	-46.136	-20.667	4.802	
INT #4 - ITT #3	-44.136	-18.667	6.802	
INT #4 - VER #2	-33.802	-8.333	17.136	

RECOGNIZER * SUBJECT COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
INT #4 - INT #1	-29.136	-3.667	21.802	
INT #4 - VOT #2	-28.802	-3.333	22.136	
INT #4 - VER #1	-24.802	0.667	26.136	
INT #4 - VER #3	-24.469	1.000	26.469	
INT #4 - INT #3	-26.642	1.833	30.309	
INT #4 - INT #2	-20.469	5.000	30.469	
INT #4 - VER #4	-18.469	7.000	32.469	
INT #2 - TI #2	-17.136	8.333	33.802	
INT #2 - TI #3	-19.142	9.333	37.809	
INT #2 - VOT #3	-8.136	17.333	42.802	
INT #2 - VOT #1	-0.802	24.667	50.136	
INT #2 - VOT #4	7.531	33.000	58.469	***
VER #1 - ITT #2	-46.802	-21.333	4.136	
VER #1 - ITT #1	-46.802	-21.333	4.136	
VER #1 - ITT #4	-46.802	-21.333	4.136	
VER #1 - ITT #3	-44.802	-19.333	6.136	
VER #1 - VER #2	-34.469	-9.000	16.469	
VER #1 - INT #1	-29.802	-4.333	21.136	
VER #1 - VOT #2	-29.469	-4.000	21.469	
VER #1 - INT #4	-26.136	-0.667	24.802	
VER #1 - VER #3	-25.136	0.333	25.802	
VER #1 - INT #3	-27.309	1.167	29.642	
VER #1 - INT #2	-21.136	4.333	29.802	
VER #1 - VER #4	-19.136	6.333	31.802	
VER #1 - TI #2	-17.802	7.667	33.136	
VER #1 - TI #3	-19.809	8.667	37.142	
VER #1 - VOT #3	-8.802	16.667	42.136	
VER #1 - VOT #1	-1.469	24.000	49.469	
VER #1 - VOT #4	6.864	32.333	57.802	***
VER #3 - ITT #2	-47.136	-21.667	3.802	
VER #3 - ITT #1	-47.136	-21.667	3.802	
VER #3 - ITT #4	-47.136	-21.667	3.802	
VER #3 - ITT #3	-45.136	-19.667	5.802	
VER #3 - VER #2	-34.802	-9.333	16.136	
VER #3 - INT #1	-30.136	-4.667	20.802	
VER #3 - VOT #2	-29.802	-4.333	21.136	
VER #3 - INT #4	-26.469	-1.000	24.469	
VER #3 - VER #1	-25.802	-0.333	25.136	
VER #3 - INT #3	-27.642	0.833	29.309	
VER #3 - INT #2	-21.469	4.000	29.469	
VER #3 - VER #4	-19.469	6.000	31.469	
VER #3 - TI #2	-18.136	7.333	32.802	

	SIMULTANEOUS		SIMULTANEOUS	
RECOGNIZER	LOWER	DIFFERENCES	UPPER	
* SUBJECT	CONFIDENCE	BETWEEN	CONFIDENCE	
COMPARISON	LIMIT	MEANS	LIMIT	
VER #3 - TI #3	-20.142	8.333	36.809	
VER #3 - VOT #3	-9.136	16.333	41.802	
VER #3 - VOT #1	-1.802	23.667	49.136	
VER #3 - VOT #4	6.531	32.000	54.469	***
INT #3 - ITT #2	-50.975	-22.500	5.975	
INT #3 - ITT #1	-50.975	-22.500	5.975	
INT #3 - ITT #4	-50.975	-22.500	5.975	
INT #3 - ITT #3	-48.975	-20.500	7.975	
INT #3 - VER #2	-38.642	-10.167	18.309	
INT #3 - INT #1	-33.975	-5.500	22.975	
INT #3 - VOT #2	-33.642	-5.167	23.309	
INT #3 - INT #4	-30.309	-1.833	26.462	
INT #3 - VER #1	-29.642	-1.167	27.309	
INT #3 - VER #3	-29.309	-0.833	27.642	
INT #3 - INT #2	-25.309	3.167	31.642	
INT #3 - VER #4	-23.309	5.167	33.672	
INT #3 - TI #2	-21.975	6.500	34.975	
INT #3 - TI #3	-23.693	7.500	38.693	
INT #3 - VOT #3	-12.975	15.500	48.975	
INT #3 - VOT #1	-5.642	22.833	51.309	
INT #3 - VOT #4	2.691	31.167	59.642	***
INT #2 - ITT #2	-51.136	-25.667	-0.198	***
INT #2 - ITT #1	-51.136	-25.667	-0.198	***
INT #2 - ITT #4	-51.136	-25.667	-0.198	***
INT #2 - ITT #3	-49.136	-23.667	1.802	
INT #2 - VER #2	-38.802	-13.333	12.136	
INT #2 - INT #1	-34.136	-8.667	16.802	
INT #2 - VOT #2	-33.802	-8.333	17.136	
INT #2 - INT #4	-30.469	-5.000	20.469	
INT #2 - VER #1	-29.802	-4.333	21.136	
INT #2 - VER #3	-29.469	-4.000	21.469	
INT #2 - INT #3	-31.642	-3.167	25.309	
INT #2 - VER #4	-23.469	2.000	27.469	
INT #2 - TI #2	-22.136	3.333	28.802	
INT #2 - TI #3	-24.142	4.333	32.809	
INT #2 - VOT #3	-13.136	12.333	37.802	
INT #2 - VOT #1	-5.802	19.667	45.136	
INT #2 - VOT #4	2.531	28.000	53.469	***
VER #4 - ITT #2	-53.136	-27.667	-2.198	***
VER #4 - ITT #1	-53.136	-27.667	-2.198	***
VER #4 - ITT #4	-53.136	-27.667	-2.198	***
VER #4 - ITT #3	-51.136	-25.667	-0.198	***

RECOGNIZER * SUBJECT COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
VER #4 - VER #2	-40.802	-15.333	10.136
VER #4 - INT #1	-36.136	-10.667	14.802
VER #4 - VOT #2	-35.802	-10.333	15.136
VER #4 - INT #4	-32.469	-7.000	18.469
VER #4 - VER #1	-31.802	-6.333	19.136
VER #4 - VER #3	-31.469	-6.000	19.469
VER #4 - INT #3	-33.642	-5.167	23.309
VER #4 - INT #2	-27.469	-2.000	23.469
VER #4 - TI #2	-24.136	1.333	26.802
VER #4 - TI #3	-26.142	2.333	30.809
VER #4 - VOT #3	-15.136	10.333	35.802
VER #4 - VOT #1	-7.802	17.667	43.136
VER #4 - VOT #4	0.531	26.000	51.469 ***
TI #2 - ITT #2	-54.469	-29.000	-3.531 ***
TI #2 - ITT #1	-54.469	-29.000	-3.531 **
TI #2 - ITT #4	-54.469	-29.000	-3.531 ***
TI #2 - ITT #3	-52.469	-27.000	-1.531 ***
TI #2 - VER #2	-42.136	-16.667	8.802
TI #2 - INT #1	-37.469	-12.000	13.469
TI #2 - VOT #2	-37.136	-11.667	13.802
TI #2 - INT #4	-33.802	-8.333	17.136
TI #2 - VER #1	-33.136	-7.667	17.802
TI #2 - VER #3	-32.802	-7.333	18.136
TI #2 - INT #3	-34.975	-6.500	21.975
TI #2 - INT #2	-28.802	-3.333	24.136
TI #2 - VER #4	-26.802	-1.333	24.136
TI #2 - TI #3	-27.475	1.000	29.475
TI #2 - VOT #3	-16.469	9.000	34.469
TI #2 - VOT #1	-9.136	16.333	41.802
TI #2 - VOT #4	-0.802	24.667	50.136
TI #3 - ITT #2	-58.475	-30.000	-1.525 ***
TI #3 - ITT #1	-58.475	-30.000	-1.525 ***
TI #3 - ITT #4	-58.475	-30.000	-1.525 ***
TI #3 - ITT #3	-56.475	-28.000	0.475
TI #3 - VER #2	-46.142	-17.667	10.809
TI #3 - INT #1	-41.475	-13.000	15.475
TI #3 - VOT #2	-41.142	-12.667	15.809
TI #3 - INT #4	-37.809	-9.333	10.142
TI #3 - VER #1	-37.142	-8.667	19.809
TI #3 - VER #3	-36.809	-8.333	20.142
TI #3 - INT #3	-38.693	-7.500	23.693
TI #3 - INT #2	-32.809	-4.333	24.142
TI #3 - VER #4	-30.809	-2.333	26.142

RECOGNIZER * SUBJECT COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
TI #3 - TI #2	-29.475	-1.000	27.475
TI #3 - VOT #3	-20.475	8.000	36.475
TI #3 - VOT #1	-13.142	15.333	43.809
TI #3 - VOT #4	-4.809	23.667	52.142
VOT #3 - ITT #2	-63.469	-38.000	-12.531 ***
VOT #3 - ITT #1	-63.469	-38.000	-12.531 ***
VOT #3 - ITT #4	-63.469	-38.000	-12.531 ***
VOT #3 - ITT #3	-61.469	-36.000	-10.531 ***
VOT #3 - VER #2	-51.136	-25.667	-0.198 ***
VOT #3 - INT #1	-46.469	-21.000	4.469
VOT #3 - VOT #2	-46.136	-20.667	4.802
VOT #3 - INT #4	-42.802	-17.333	8.136
VOT #3 - VER #1	-42.136	-16.667	8.802
VOT #3 - VER #3	-41.802	-16.333	9.136
VOT #3 - INT #3	-43.975	-15.500	12.975
VOT #3 - INT #2	-37.802	-12.333	13.136
VOT #3 - VER #4	-35.802	-10.333	15.136
VOT #3 - TI #2	-34.469	-9.000	16.469
VOT #3 - TI #3	-36.475	-8.000	20.475
VOT #3 - VOT #1	-18.136	7.333	32.802
VOT #3 - VOT #4	-9.802	15.667	41.136
VOT #1 - ITT #2	-70.802	-45.333	-19.864 ***
VOT #1 - ITT #1	-70.802	-45.333	-19.864 ***
VOT #1 - ITT #4	-70.802	-45.333	-19.864 ***
VOT #1 - ITT #3	-68.802	-43.333	-17.864 ***
VOT #1 - VER #2	-58.469	-33.000	-7.531 ***
VOT #1 - INT #1	-53.802	-28.333	-2.864 ***
VOT #1 - VOT #2	-53.469	-28.000	-2.531 ***
VOT #1 - INT #4	-50.136	-24.667	0.802
VOT #1 - VER #1	-49.469	-24.000	1.469
VOT #1 - VER #3	-49.136	-23.667	1.802
VOT #1 - INT #3	-51.309	-22.833	5.642
VOT #1 - INT #2	-45.136	-19.667	5.802
VOT #1 - VER #4	-43.136	-17.667	7.802
VOT #1 - TI #2	-41.802	-16.333	9.136
VOT #1 - TI #3	-43.809	-15.333	13.142
VOT #1 - VOT #3	-32.802	-7.333	18.136
VOT #1 - VOT #4	-17.136	8.333	33.802
VOT #4 - ITT #2	-79.136	-53.667	-28.198 ***
VOT #4 - ITT #1	-79.136	-53.667	-18.198 ***
VOT #4 - ITT #4	-17.136	-53.667	-28.198 ***
VOT #4 - ITT #3	-77.136	-51.667	-26.198 ***

RECOGNIZER * SUBJECT COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
VOT #4 - VER #2	-66.802	-41.333	-15.846 ***
VOT #4 - INT #1	-62.106	-36.667	-11.198 ***
VOT #4 - VOT #2	-61.802	-36.333	-10.864 ***
VOT #4 - INT #4	-58.469	-33.000	-7.531 ***
VOT #4 - VER #1	-57.802	-32.333	-6.864 ***
VOT #4 - VER #3	-57.469	-32.000	-6.531 ***
VOT #4 - INT #3	-59.642	-31.167	-2.691 ***
VOT #4 - INT #2	-53.469	-28.000	-2.531 ***
VOT #4 - VER #4	-51.469	-26.000	-0.531 ***
VOT #4 - TI #2	-50.136	-24.667	0.802
VOT #4 - TI #3	-52.142	-23.667	4.809
VOT #4 - VOT #3	-41.136	-15.667	9.802
VOT #4 - VOT #1	-33.802	-8.333	17.136

Appendix XIV. Tukey Analysis of Recognizer * Noise: Scenario 2

```
TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE:  CORRECT
ALPHA=0.05 CONFIDENCE=0.95 DF=24 MSE-64.4803
CRITICAL VALUE OF STUDENTIZED RANGE=5.319

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY  '***'


                          SIMULTANEOUS                          SIMULTANEOUS
RECOGNIZER                   LOWER                                 UPPER
 * NOISE                  CONFIDENCE          BETWEEN           CONFIDENCE
COMPARISON                   LIMIT             MEANS               LIMIT

ITT #2 - ITT #0          -21.354             0.000             21.354
ITT #2 - ITT #1          -19.854             1.500             22.854
ITT #2 - VER #2           -9.354            12.000             33.354
ITT #2 - VER #0           -7.104            14.250             35.604
ITT #2 - INT #2           -5.854            15.500             36.854
ITT #2 - INT #0            1.646            23.000             44.354 ***
ITT #2 - TI  #0           -2.153            24.000             50.153
ITT #2 - TI  #1           -7.764            26.000             59.764
ITT #2 - INT #1            3.985            27.000             50.065 ***
ITT #2 - VOT #0           18.146            34.500             55.854 ***
ITT #2 - VOT #2           14.646            36.000             57.354 ***
ITT #2 - VER #1           14.646            36.000             57.354 ***
ITT #2 - TI  #2           10.347            36.500             62.653 ***
ITT #2 - VOT #1           23.896            45.250             66.604 ***

ITT #0   - ITT #2        -21.354             0.000             21.354
ITT #0   - ITT #1        -19.854             1.500             22.854
ITT #0   - VER #2         -9.354            12.000             33.354
ITT #0   - VER #0         -7.104            14.250             35.604
ITT #0   - INT #2         -5.854            15.500             36.854
ITT #0   - INT #0          1.646            23.000             44.354 ***
ITT #0   - TI  #0         -2.153            24.000             50.153
ITT #0   - TI  #1         -7.764            26.000             59.764
ITT #0   - INT #1          3.935            27.000             50.065 ***
ITT #0   - VOT #0         13.146            34.500             55.854 ***
ITT #0   - VOT #2         14.646            36.000             57.354 ***
ITT #0   - VER #1         14.646            36.000             57.354 ***
ITT #0   - TI  #2         10.347            36.500             62.653 ***
ITT #0   - VOT #1         23.896            45.250             66.604 ***

ITT #1 - ITT #2          -22.854            -1.500             19.854
```

RECOGNIZER * NOISE COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
ITT #1 - ITT #0	-22.854	-1.500	19.854
ITT #1 - VER #2	-10.854	10.500	31.854
ITT #1 - VER #0	-8.604	12.750	34.104
ITT #1 - INT #2	-7.354	14.000	35.354
ITT #1 - INT #0	0.146	21.500	42.854 ***
ITT #1 - TI #0	-3.653	22.500	48.653
ITT #1 - TI #1	-9.264	24.500	58.264
ITT #1 - INT #1	2.435	25.500	48.565 ***
ITT #1 - VOT #0	11.646	33.000	54.354 ***
ITT #1 - VOT #2	13.146	34.500	55.854 ***
ITT #1 - VER #1	13.146	34.500	55.854 ***
ITT #1 - TI #2	8.847	35.000	61.153 ***
ITT #1 - VOT #1	22.396	43.750	65.104 ***
VER #2 - ITT #2	-33.354	-12.000	9.354
VER #2 - ITT #0	-33.354	-12.000	9.354
VER #2 - ITT #1	-31.854	-10.500	10.854
VER #2 - VER #0	-19.104	2.250	23.604
VER #2 - INT #2	-17.854	3.500	24.854
VER #2 - INT #0	-10.354	11.000	32.354
VER #2 - TI #0	-14.153	12.000	38.153
VER #2 - TI #1	-19.764	14.000	47.764
VER #2 - INT #1	-8.065	15.000	38.065
VER #2 - VOT #0	1.146	22.500	43.854
VER #2 - VOT #2	2.646	24.000	45.354 ***
VER #2 - VER #1	2.646	24.000	45.354 ***
VER #2 - TI #2	-1.653	24.500	50.653
VER #2 - VOT #1	11.896	33.250	54.604 ***
VER #0 - ITT #2	-35.604	-14.250	7.104
VER #0 - ITT #0	-35.604	-14.250	7.104
VER #0 - ITT #1	-34.104	-12.750	8.604
VER #0 - VER #2	-23.604	-2.250	19.104
VER #0 - INT #2	-20.104	1.250	22.604
VER #0 - INT #0	-12.604	8.750	30.104
VER #0 - ITT #0	-16.403	9.750	35.903
VER #0 - TI #1	-22.014	11.750	45.514
VER #0 - INT #1	-10.315	12.750	35.815
VER #0 - VOT #0	-1.104	20.250	41.604
VER #0 - VOT #2	0.396	21.750	43.104 ***
VER #0 - VER #1	0.396	21.750	43.104 ***
VER #0 - TI #2	-3.903	22.250	48.403
VER #0 - VOT #1	9.646	31.000	52.354 ***

RECOGNIZER * NOISE COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
INT #2 - ITT #2	-36.854	-15.500	5.854	
INT #2 - ITT #0	-36.854	-15.500	5.854	
INT #2 - ITT #1	-35.354	-14.000	7.354	
INT #2 - VER #2	-24.854	-3.500	17.854	
INT #2 - VER #0	-22.604	-1.250	20.104	
INT #2 - INT #0	-13.854	7.500	28.854	
INT #2 - TI #0	-17.653	8.500	34.653	
INT #2 - TI #1	-23.264	10.500	44.264	
INT #2 - INT #1	-11.565	11.500	34.656	
INT #2 - VOT #0	-2.354	19.000	40.354	
INT #2 - VOT #2	-0.854	20.500	41.854	
INT #2 - VER #1	-0.854	20.500	41.854	
INT #2 - TI #2	-5.153	21.000	47.153	
INT #2 - VOT #1	8.396	29.750	51.104	***
INT #0 - ITT #2	-44.354	-23.000	-1.646	***
INT #0 - ITT #0	-44.354	-23.000	-1.646	***
INT #0 - ITT #1	-42.854	-21.500	-0.146	***
INT #0 - VER #2	-32.354	-11.000	10.354	
INT #0 - VER #0	-30.104	-8.750	12.604	
INT #0 - INT #2	-28.854	-7.500	13.854	
INT #0 - TI #0	-25.153	1.000	27.153	
INT #0 - TI #1	-30.764	3.000	36.764	
INT #0 - INT #1	-19.065	4.000	27.065	
INT #0 - VOT #0	-9.854	11.500	32.854	
INT #0 - VOT #2	-8.354	13.000	34.354	
INT #0 - VER #1	-8.354	13.000	34.354	
INT #0 - TI #2	-12.653	13.500	39.653	
INT #0 - VOT #1	0.896	22.250	43.604	***
TI #0 - ITT #2	-50.153	-24.000	2.153	
TI #0 - ITT #0	-50.153	-24.000	2.153	
TI #0 - ITT #1	-48.653	-22.500	3.653	
TI #0 - VER #2	-38.153	-12.000	14.153	
TI #0 - VER #0	-35.903	-9.750	16.403	
TI #0 - INT #2	-34.653	-8.500	17.653	
TI #0 - INT #0	-27.153	-1.000	25.153	
TI #0 - TI #1	-34.986	2.000	38.986	
TI #0 - INT #1	-24.568	3.000	30.568	
TI #0 - VOT #0	-15.653	10.500	36.653	
TI #0 - VOT #2	-14.153	12.000	38.153	
TI #0 - VOT #1	-14.153	12.000	38.153	
TI #0 - TI #2	-17.699	12.500	42.699	
TI #0 - VER #1	-4.903	21.250	47.403	
TI #1 - ITT #2	-59.764	-26.000	7.764	

RECOGNIZER * NOISE COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
TI #1 - ITT #0	-59.764	-26.000	7.764
TI #1 - ITT #1	-58.264	-24.500	9.264
TI #1 - VER #2	-47.764	-14.000	19.764
TI #1 - VER #0	-45.514	-11.750	22.014
TI #1 - INT #2	-44.264	-10.500	23.264
TI #1 - INT #0	-36.764	-3.000	30.764
TI #1 - TI #0	-38.986	-2.000	34.986
TI #1 - INT #1	-33.871	1.000	35.871
TI #1 - VOT #0	-25.264	8.500	42.264
TI #1 - VOT #2	-23.764	10.000	43.764
TI #1 - VER #1	-23.764	10.000	43.764
TI #1 - TI #2	-26.486	10.500	47.486
TI #1 - VOT #1	-14.514	19.250	53.014
INT #1 - ITT #2	-50.065	-27.000	-3.935 ***
INT #1 - ITT #0	-50.065	-27.000	-3.935 ***
INT #1 - ITT #1	-48.565	-25.500	-2.435 ***
INT #1 - VER #2	-38.065	-15.000	8.065
INT #1 - VER #0	-35.815	-12.750	10.315
INT #1 - INT #2	-34.565	-11.300	11.565
INT #1 - INT #0	-27.065	-4.000	19.065
INT #1 - TI #0	-30.568	-3.000	24.568
INT #1 - TI #1	-35.871	-1.000	33.871
INT #1 - VOT #0	-15.565	7.500	30.565
INT #1 - VOT #2	-14.065	9.000	32.065
INT #1 - VER #1	-14.065	9.000	32.065
INT #1 - TI #2	-18.068	9.500	37.068
INT #1 - VOT #1	-4.815	18.250	41.315
VOT #0 - ITT #2	-55.854	-34.500	-13.146 ***
VOT #0 - ITT #0	-55.854	-34.500	-13.146 ***
VOT #0 - ITT #1	-54.354	-33.000	-11.646 ***
VOT #0 - VER #2	-43.854	-22.500	-1.146 ***
VOT #0 - VER #0	-41.604	-20.250	1.104
VOT #0 - INT #2	-40.354	-19.000	2.354
VOT #0 - INT #0	-32.854	-11.500	9.854
VOT #0 - TI #0	-36.653	-10.500	15.653
VOT #0 - TI #1	-42.264	-8.500	25.264
VOT #0 - INT #1	-30.565	-7.500	15.565
VOT #0 - VOT #2	-19.854	1.500	22.854
VOT #0 - VER #1	-19.854	1.500	22.854
VOT #0 - TI #2	-24.153	2.000	28.153
VOT #0 - VOT #1	-10.604	10.750	32.104

RECOGNIZER * NOISE COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT
VOT #2 - ITT #2	-57.354	-36.000	-14.646 ***
VOT #2 - ITT #0	-57.354	-36.000	-14.646 ***
VOT #2 - ITT #1	-55.854	-34.500	-13.146 ***
VOT #2 - VER #2	-45.354	-24.000	-2.646 ***
VOT #2 - VER #0	-43.104	-21.750	-0.396 ***
VOT #2 - INT #2	-41.854	-20.500	0.854
VOT #2 - INT #0	-34.354	-13.000	8.354
VOT #2 - TI #0	-38.153	-12.000	14.153
VOT #2 - TI #1	-43.764	-10.000	23.764
VOT #2 - INT #1	-32.065	-9.000	14.065
VOT #2 - VOT #0	-22.854	-1.500	19.854
VOT #2 - VER #1	-21.354	0.000	21.354
VOT #2 - TI #2	-25.653	0.500	26.653
VOT #2 - VOT #1	-12.104	9.250	30.604
VER #1 - ITT #2	-57.354	-36.000	-14.646 ***
VER #1 - ITT #0	-57.354	-36.000	-14.646 ***
VER #1 - ITT #1	-55.854	-34.500	-13.146 ***
VER #1 - VER #2	-45.354	-24.000	-2.646 ***
VER #1 - VER #0	-43.104	-21.750	-0.396 ***
VER #1 - INT #2	-41.854	-20.500	0.854
VER #1 - INT #0	-34.354	-13.000	8.354
VER #1 - TI #0	-38.153	-12.000	14.153
VER #1 - TI #1	-43.764	-10.000	23.764
VER #1 - INT #1	-32.065	-9.000	14.065
VER #1 - VOT #0	-22.854	-1.500	19.854
VER #1 - VOT #2	-21.354	0.000	21.354
VER #1 - TI #2	-25.653	0.500	26.653
VER #1 - VOT #1	-12.104	9.250	30.604
TI #2 - ITT #2	-62.653	-36.500	-10.347 ***
TI #2 - ITT #0	-62.653	-36.500	-10.347 ***
TI #2 - ITT #1	-61.153	-35.000	-8.847 ***
TI #2 - VER #2	-50.653	-24.500	1.653
TI #2 - VER #0	-48.403	-22.250	3.903
TI #2 - INT #2	-47.153	-21.000	5.153
TI #2 - INT #0	-39.653	-13.500	12.653
TI #2 - TI #0	-42.699	-12.500	17.699
TI #2 - TI #1	-47.486	-10.500	26.486
TI #2 - INT #1	-37.068	-9.500	18.068
TI #2 - VOT #0	-28.153	-2.000	24.153
TI #2 - VOT #2	-26.653	-0.500	25.653
TI #2 - VER #1	-26.653	-0.500	25.653
TI #2 - VOT #1	-17.403	8.750	34.903

RECOGNIZER * NOISE COMPARISON	SIMULTANEOUS LOWER CONFIDENCE LIMIT	DIFFERENCES BETWEEN MEANS	SIMULTANEOUS UPPER CONFIDENCE LIMIT	
VOT #1 - ITT #2	-66.604	-45.250	-23.896	***
VOT #1 - ITT #0	-66.604	-45.250	-23.896	***
VOT #1 - ITT #1	-65.104	-43.750	-22.396	***
VOT #1 - VER #2	-54.604	-33.250	-11.896	***
VOT #1 - VER #0	-52.354	-31.000	-9.646	***
VOT #1 - INT #2	-51.104	-29.750	-8.396	***
VOT #1 - INT #0	-43.604	-22.250	-0.896	***
VOT #1 - TI #0	-47.403	-21.250	4.903	
VOT #1 - TI #1	-53.014	-19.250	14.514	
VOT #1 - INT #1	-41.315	-18.250	4.815	
VOT #1 - VOT #0	-32.104	-10.750	10.604	
VOT #1 - VOT #2	-30.604	-9.250	12.104	
VOT #1 - VER #1	-30.604	-9.250	12.104	
VOT #1 - TI #2	-34.903	-8.750	17.403	

Appendix XV. Rejection/Misrecognition Matricies: Scenario 2

NUMBER OF WORDS REJECTED
(Across all Speakers)

SENTENCE LENGTH

RECOGNIZER	NOISE	3 WORD														4 WORD								5 WORD								
		1	2	3	4	5	6	7	8	9	10	11	12	TOTAL	%	13	14	15	16	17	18	TOTAL	%	19	20	21	22	23	24	TOTAL	%	TOT REJ.
VOTAN	#0													0	0							0	0							0	0	0
	#1													0	0							0	0							0	0	0
	#2													0	0					1		1	.01							0	0	0
TI	#0	1	1			1	1	2	1			3		10	.10		2					2	.04	1			2	1		4	.08	16
	#1	1					3	2			2		2	10	.21							0	0							0	0	10
	#2	4			1		6	6	1			4	4	26	.27	1	5		3		1	10	.21	1			2	1	1	5	.10	41
INT	#0	4	4		4	4	4	1	3	1	7	3	2	37	.19	2	7	1	2	4	4	20	.21	13		1	7		1	22	.23	79
	#1		2		1	3	5		3	1	4	4		23	.16	3	2	1	2	5		13	.18	7	1	2	2	1	1	14	.19	50
	#2	3	2	1			1				1	2		10	.05		3	2	1	4	3	13	.14	6	1	2	4			13	.14	36
VER	#0	1	1		1	3	3				4			13	.07		1		1	6	1	9	.09	9		4	4			17	.18	39
	#1	4	4	4	5	9	7	5	2	1	4	3		48	.25	5	5	1	4	6	5	26	.27	11	4	3	11	1		30	.31	104
	#2	1					1			1		1		4	.02			2		1	1	4	.04	5		2	2		1	10	.10	18
ITT	#0													0	0							0	0							0	0	0
	#1													0	0							0	0							0	0	0
	#2													0	0							0	0							0	0	0

#0 NO NOISE
#1 INDUSTRIAL
#2 FAST FOOD RESTAURANT

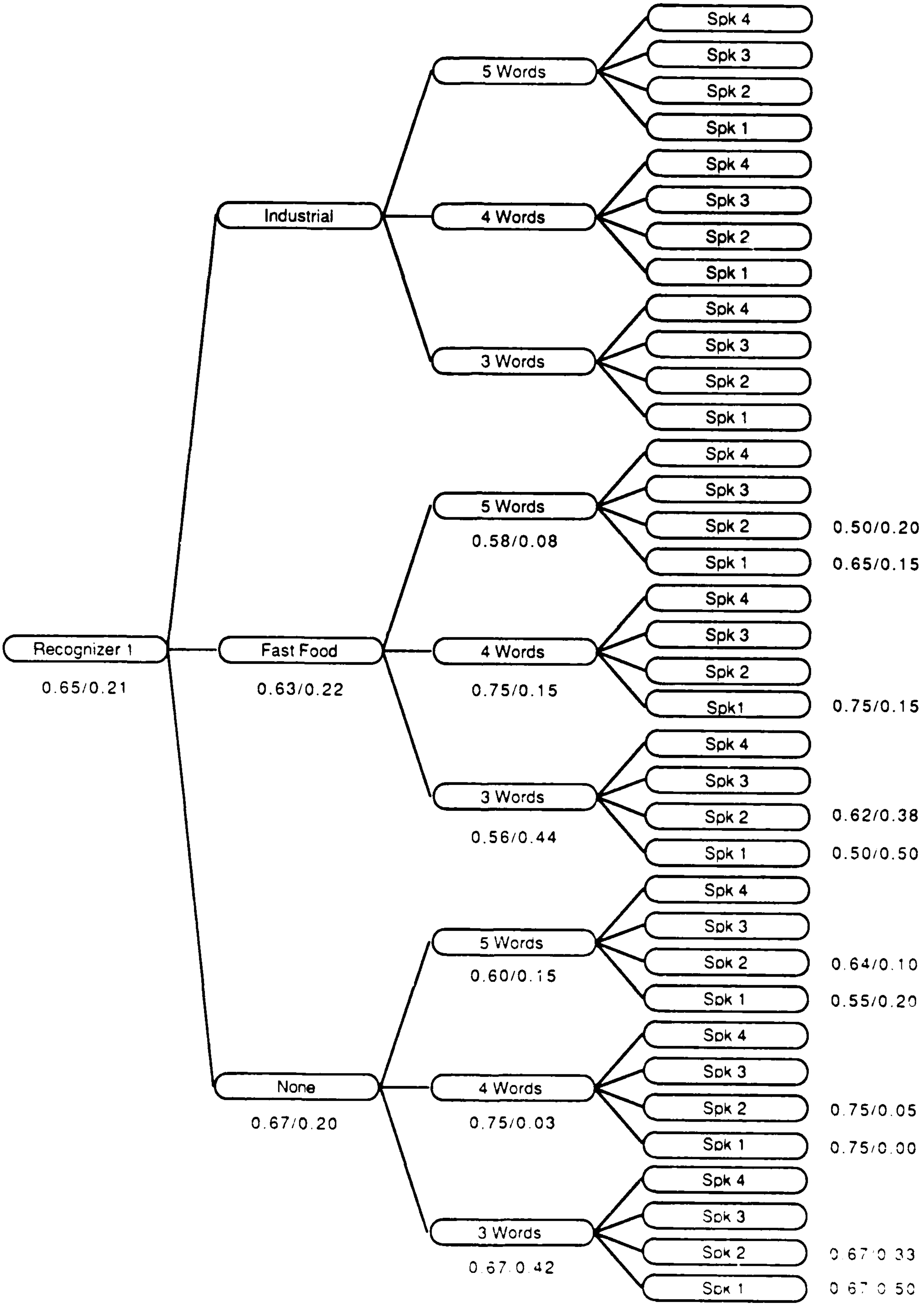
NUMBER OF WORDS MISRECOGNIZED
(Across all Speakers)

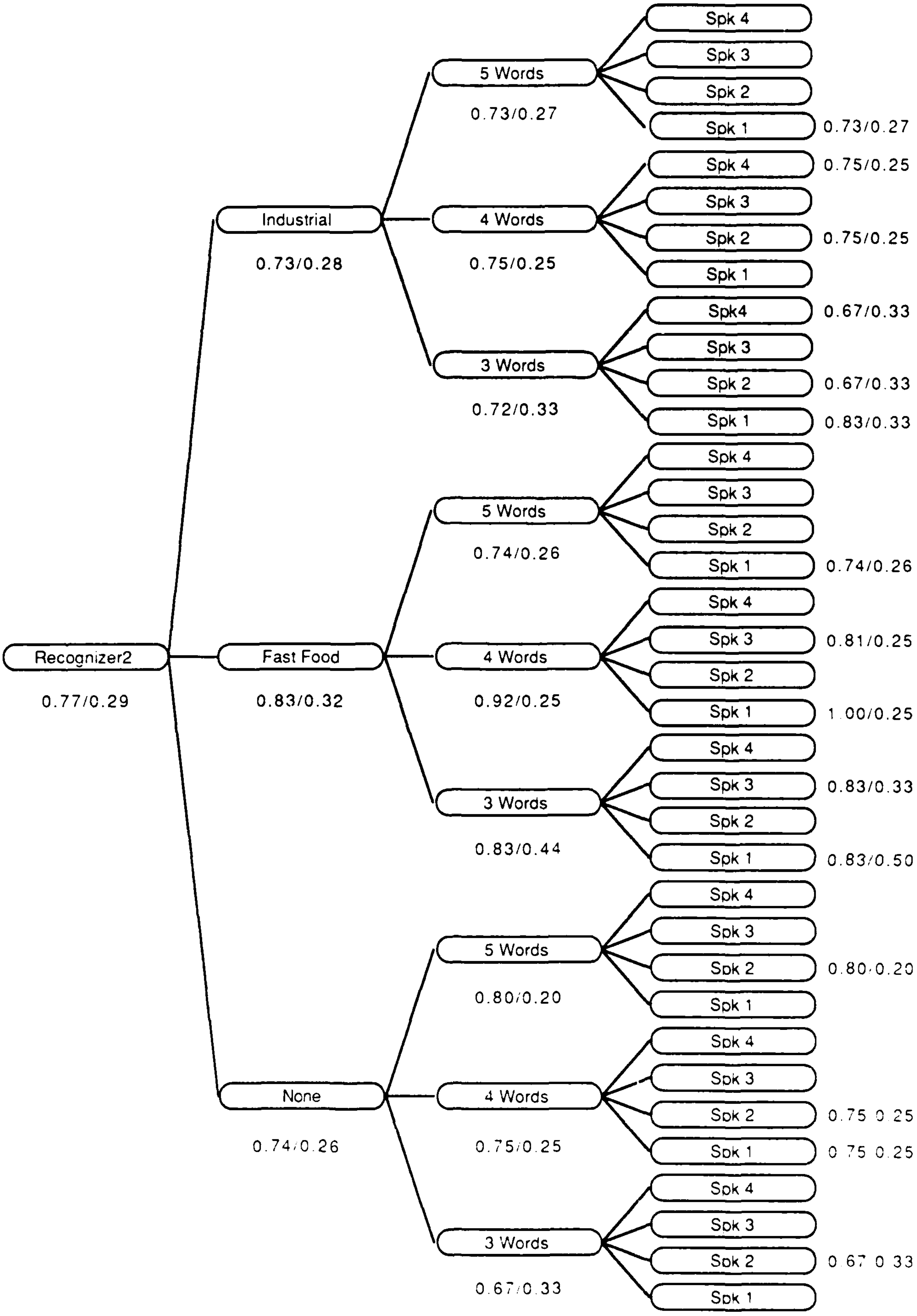
SENTENCE LENGTH

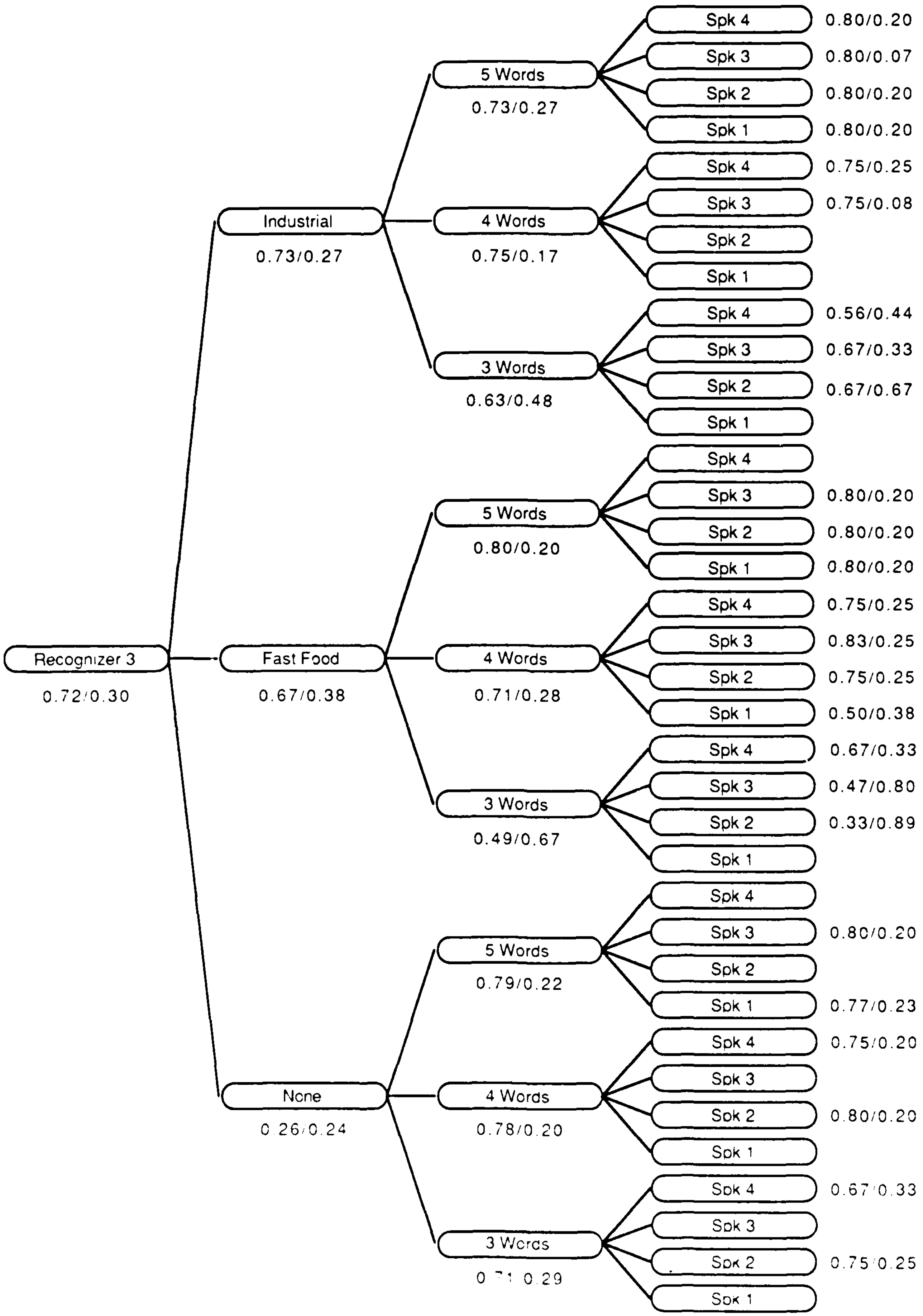
RECOGNIZER	NOISE	3 WORD														4 WORD								5 WORD								
		1	2	3	4	5	6	7	8	9	10	11	12	TOTAL	$\bar{X}$	13	14	15	16	17	18	TOTAL	$\bar{X}$	19	20	21	22	23	24	TOTAL	$\bar{X}$	TOT MISRECOG.
VOTAN	#0	7	3	9	1	2	1	9	7		5	5	5	54	.28	1	6	9	12	11	4	43	.45	12	7	2	14	4	2	41	.43	138
	#1	9	2	5	3	7	5	10	8	1	7	7	6	70	.36	7	6	8	12	13	5	51	.53	15	8	5	15	8	8	59	.61	180
	#2	11	2	7	2	3	6	9	6	1	6	9	4	66	.34	2	5	3	11	10	5	36	.38	9	8	2	12	7	2	40	.42	131
TI	#0	1				2		1						4	.04		1	3	1			5	.10	2	2	1	1	4		10	.21	18
	#1		2									2		4	.08				1	3		4	.17	2	3			2	1	8	.33	16
	#2		1		2	3		2	1			2	2	13	.14				1	1	3	5	.10	3	4		1	5	1	14	.30	32
INTER	#0								2					2	.01	1			3	1		5	.05				5			5	.05	12
	#1					1			5				5	11	.08		1	1	2	4	2	10	.14	3	1		4			8	.11	29
	#2								1			1	3	5	.03	1		1	3	4		9	.09	5			4			9	.09	23
VERBEX	#0								1					1	.01			6	1			7	.07	4			6			10	.10	18
	#1	2	3	1	1	4		1	2		1	1	2	18	.09	1	2	5			3	11	.11	1			3		1	5	.05	34
	#2	3				2	1		1				1	8	.04			4	2	1		7	.07	4	3		8			15	.16	30
ITT	#0													0	.0							0	.0							0	.0	0
	#1				1			1					1	3	.02			1				1	.01	1		1				2	.02	6
	#2													0	.0							0	.0							0	.0	0

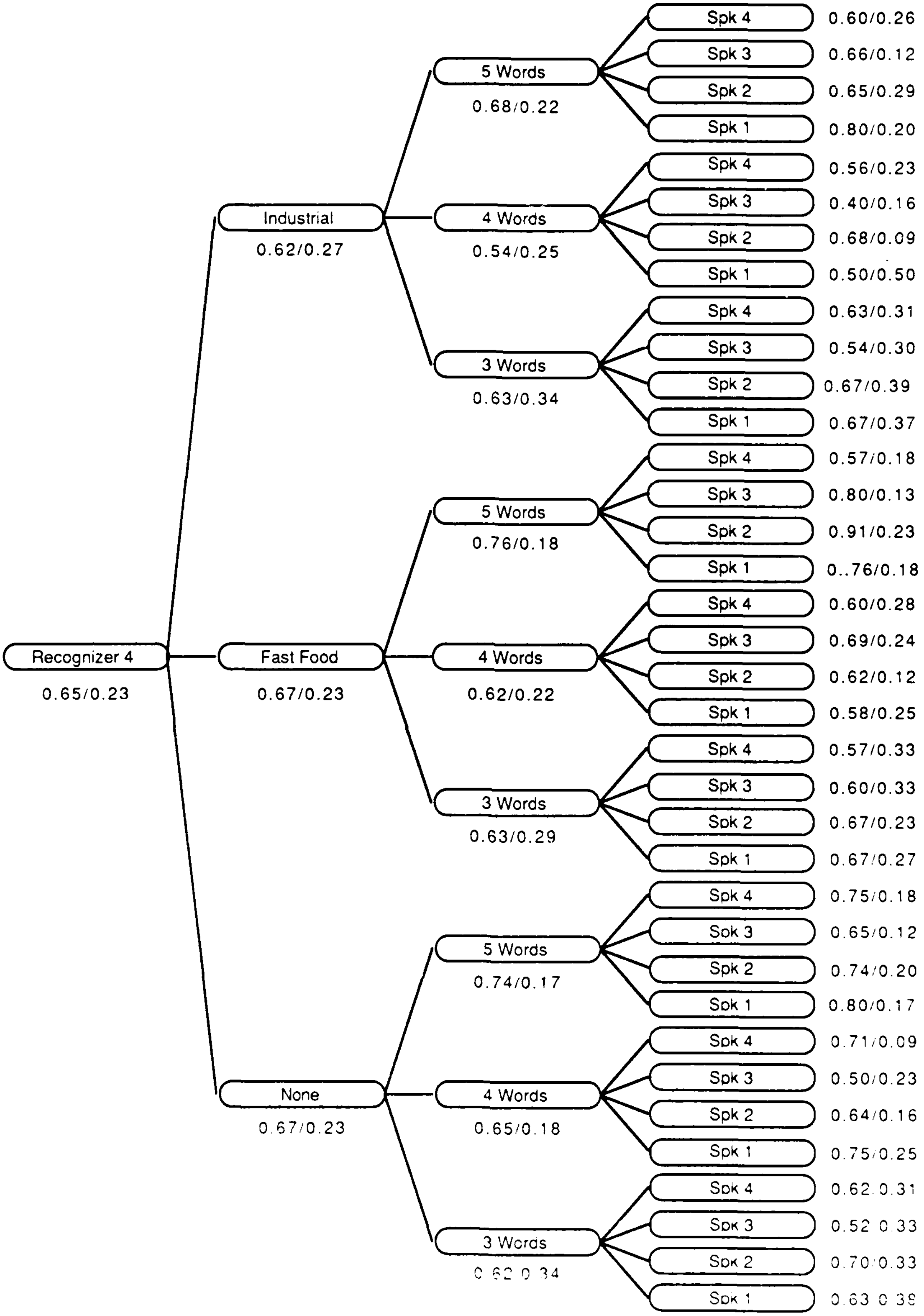
#0 NO NOISE
#1 INDUSTRIAL
#2 FAST FOOD RESTAURANT

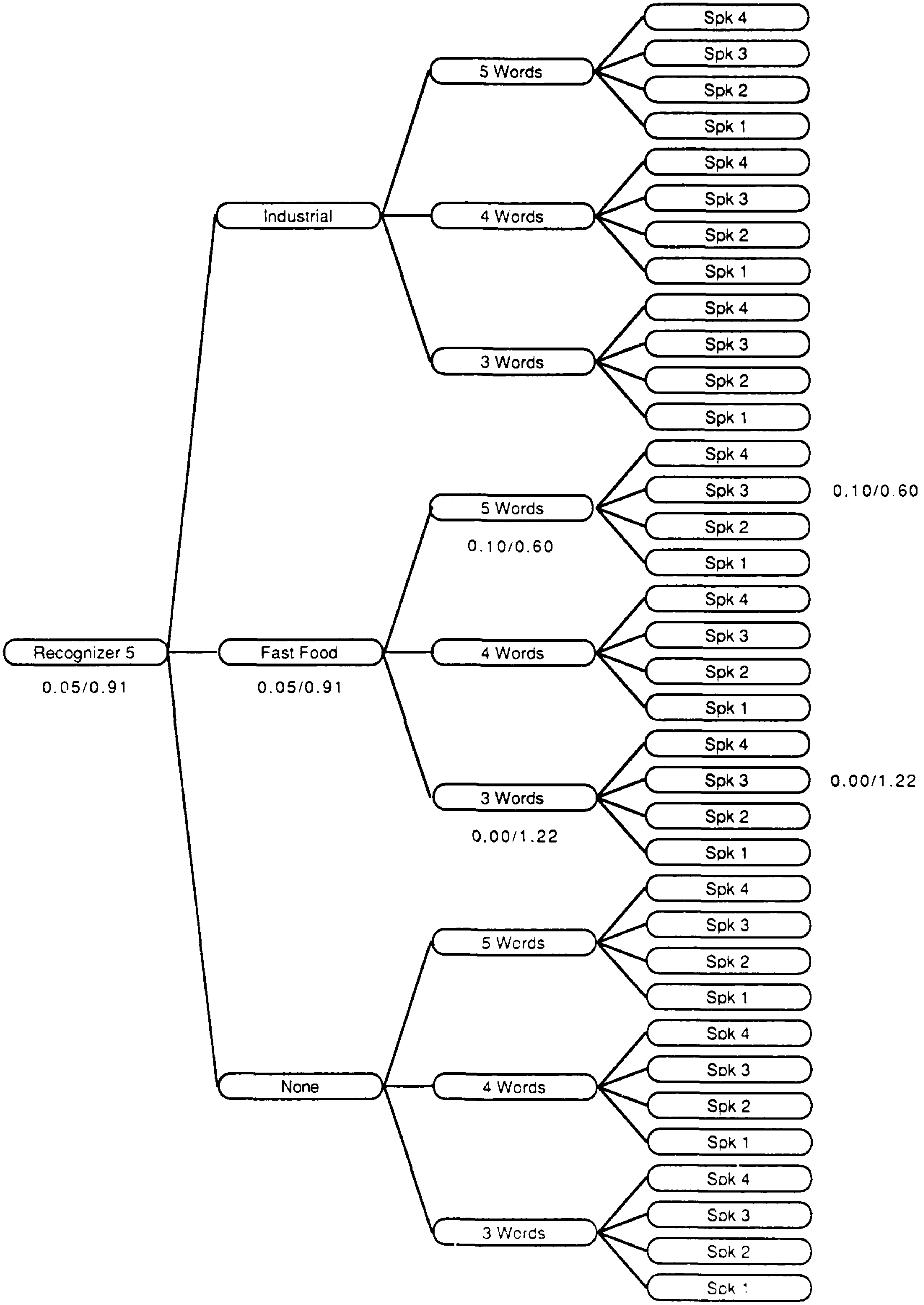
Appendix XVI. Misrecognition Error Trees: Scenario 2




Recognizer2
0.77/0.29
Industrial
0.73/0.28
5 Words
0.73/0.27
Spk 4
Spk 3
Spk 2
Spk 1
0.73/0.27
4 Words
0.75/0.25
Spk 4
0.75/0.25
Spk 3
Spk 2
0.75/0.25
Spk 1
3 Words
0.72/0.33
Spk4
0.67/0.33
Spk 3
Spk 2
0.67/0.33
Spk 1
0.83/0.33
Fast Food
0.83/0.32
5 Words
0.74/0.26
Spk 4
Spk 3
Spk 2
Spk 1
0.74/0.26
4 Words
0.92/0.25
Spk 4
Spk 3
0.81/0.25
Spk 2
Spk 1
1.00/0.25
3 Words
0.83/0.44
Spk 4
Spk 3
0.83/0.33
Spk 2
Spk 1
0.83/0.50
None
0.74/0.26
5 Words
0.80/0.20
Spk 4
Spk 3
Spk 2
0.80/0.20
Spk 1
4 Words
0.75/0.25
Spk 4
Spk 3
Spk 2
0.75/0.25
Spk 1
0.75/0.25
3 Words
0.67/0.33
Spk 4
Spk 3
Spk 2
0.67/0.33
Spk 1


Recognizer 3
0.72/0.30
Industrial
0.73/0.27
5 Words
0.73/0.27
Spk 4
0.80/0.20
Spk 3
0.80/0.07
Spk 2
0.80/0.20
Spk 1
0.80/0.20
4 Words
0.75/0.17
Spk 4
0.75/0.25
Spk 3
0.75/0.08
Spk 2
Spk 1
3 Words
0.63/0.48
Spk 4
0.56/0.44
Spk 3
0.67/0.33
Spk 2
0.67/0.67
Spk 1
Fast Food
0.67/0.38
5 Words
0.80/0.20
Spk 4
Spk 3
0.80/0.20
Spk 2
0.80/0.20
Spk 1
0.80/0.20
4 Words
0.71/0.28
Spk 4
0.75/0.25
Spk 3
0.83/0.25
Spk 2
0.75/0.25
Spk 1
0.50/0.38
3 Words
0.49/0.67
Spk 4
0.67/0.33
Spk 3
0.47/0.80
Spk 2
0.33/0.89
Spk 1
None
0.26/0.24
5 Words
0.79/0.22
Spk 4
Spk 3
0.80/0.20
Spk 2
Spk 1
0.77/0.23
4 Words
0.78/0.20
Spk 4
0.75/0.20
Spk 3
Spk 2
0.80/0.20
Spk 1
3 Words
0.71/0.29
Spk 4
0.67/0.33
Spk 3
Spk 2
0.75/0.25
Spk 1

Recognizer 4
0.65/0.23
Industrial
0.62/0.27
5 Words
0.68/0.22
Spk 4
0.60/0.26
Spk 3
0.66/0.12
Spk 2
0.65/0.29
Spk 1
0.80/0.20
4 Words
0.54/0.25
Spk 4
0.56/0.23
Spk 3
0.40/0.16
Spk 2
0.68/0.09
Spk 1
0.50/0.50
3 Words
0.63/0.34
Spk 4
0.63/0.31
Spk 3
0.54/0.30
Spk 2
0.67/0.39
Spk 1
0.67/0.37
Fast Food
0.67/0.23
5 Words
0.76/0.18
Spk 4
0.57/0.18
Spk 3
0.80/0.13
Spk 2
0.91/0.23
Spk 1
0..76/0.18
4 Words
0.62/0.22
Spk 4
0.60/0.28
Spk 3
0.69/0.24
Spk 2
0.62/0.12
Spk 1
0.58/0.25
3 Words
0.63/0.29
Spk 4
0.57/0.33
Spk 3
0.60/0.33
Spk 2
0.67/0.23
Spk 1
0.67/0.27
None
0.67/0.23
5 Words
0.74/0.17
Spk 4
0.75/0.18
Spk 3
0.65/0.12
Spk 2
0.74/0.20
Spk 1
0.80/0.17
4 Words
0.65/0.18
Spk 4
0.71/0.09
Spk 3
0.50/0.23
Spk 2
0.64/0.16
Spk 1
0.75/0.25
3 Words
0.62/0.34
Spk 4
0.62/0.31
Spk 3
0.52/0.33
Spk 2
0.70/0.33
Spk 1
0.63/0.39


Recognizer 5
0.05/0.91
Industrial
Fast Food
0.05/0.91
None
5 Words
4 Words
3 Words
5 Words
0.10/0.60
4 Words
3 Words
0.00/1.22
5 Words
4 Words
3 Words
Spk 4
Spk 3
Spk 2
Spk 1
Spk 4
Spk 3
Spk 2
Spk 1
Spk 4
Spk 3
Spk 2
Spk 1
Spk 4
Spk 3
0.10/0.60
Spk 2
Spk 1
Spk 4
Spk 3
Spk 2
Spk 1
Spk 4
Spk 3
0.00/1.22
Spk 2
Spk 1
Spk 4
Spk 3
Spk 2
Spk 1
Spk 4
Spk 3
Spk 2
Spk 1
Spk 4
Spk 3
Spk 2
Spk 1

Appendix XVII. Performance Means: Scenario 2

Scenario 2   CORRECT SENTENCES

RECOGNIZER	NOISE	SUM	MEAN	TOTAL CORRECT ---------- TOTAL UTTERANCES
VOT	0	246.00	61.50	.64
	1	203.00	50.75	.53
	2	240.00	60.00	.63
TI	0	144.00	72.00	.75
	1	70.00	70.00	.73
	2	119.00	59.50	.62
INT	0	292.00	73.00	.76
	1	207.00	69.00	.72
	2	322.00	80.50	.84
VER	0	327.00	81.75	.85
	1	240.00	60.00	.63
	2	336.00	84.00	.88
ITT	0	384.00	96.00	1.00
	1	378.00	94.50	.98
	2	384.00	96.00	1.00
ALL		3892.00	74.85	.78

Scenerio 2

SENTENCES REJECTED

RECOGNIZER	NOISE	SUM	MEAN	TOTAL REJECTED ---------- TOTAL UTTERANCES
VOT	0	0.00	0.00	.00
	1	0.00	0.00	.00
	2	1.00	0.25	.003
TI	0	16.00	8.00	.08
	1	10.00	10.00	.10
	2	41.00	20.50	.21
INT	0	80.00	20.00	.21
	1	51.00	17.00	.18
	2	36.00	9.00	.09
VER	0	39.00	9.75	.10
	1	109.00	27.25	.28
	2	18.00	4.50	.05
ITT	0	0.00	0.00	.00
	1	0.00	0.00	.00
	2	0.00	0.00	.00
ALL		401.00	7.71	

Scenario 2 SENTENCES MISRECOGNIZED

				TOTAL CORRECT ---------- TOTAL UTTERANCES
		SUM	MEAN	
RECOGNIZER	NOISE			
VOT	0	138.00	34.50	.36
	1	181.00	45.25	.47
	2	143.00	35.75	.37
TI	0	32.00	16.00	.17
	1	16.00	16.00	.17
	2	32.00	16.00	.17
INT	0	12.00	3.00	.03
	1	30.00	10.00	.10
	2	26.00	6.50	.07
VER	0	18.00	4.50	.05
	1	35.00	8.75	.09
	2	30.00	7.50	.08
ITT	0	0.00	0.00	.00
	1	6.00	1.50	.02
	2	0.00	0.00	.00
ALL		699.00	13.44	

```
TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE:  CORRECT
ALPHA=0.05 CONFIDENCE=0.95 DF=24 MSE-64.4803
CRITICAL VALUE OF STUDENTIZED RANGE=4.166

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY  '***'

                         SIMULTANEOUS                          SIMULTANEOUS
                            LOWER         DIFFERENCES             UPPER
RECOGNIZER               CONFIDENCE         BETWEEN            CONFIDENCE
COMPARISON                  LIMIT            MEANS                LIMIT

ITT         VER            10.592           20.250              29.908   ***
ITT         INT            10.989           20.864              30.739   ***
ITT         TI             16.307           28.900              41.493   ***
ITT         VOT            28.425           38.083              47.741   ***


VER         ITT           -29.908          -20.250             -10.592   ***
VER         INT            -9.261            0.614              10.489
VER         TI             -3.943            8.650              21.243
VER         VOT             8.175           17.833              27.491   ***

INT         ITT           -30.739          -20.864             -10.989   ***
INT         VER           -10.489           -0.614               9.261
INT         TI             -4.723            8.036              20.796
INT         VOT             7.345           17.220              27.095   ***

TI           ITT           -41.493          -28.900             -16.307   ***
TI          INT           -21.243           -8.650               3.943
TI          INT           -20.796           -8.036               4.723
TI          VOT            -3.409            9.183              21.776

VOT         ITT           -47.741          -38.083             -28.425   ***
VOT         VER           -27.491          -17.833              -8.175   ***
VOT         INT           -27.095          -17.220              -7.345   ***
VOT         TI            -21.776           -9.183               3.409
```

```
TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR VARIABLE:  CORRECT
ALPHA=0.05 CONFIDENCE=0.95 DF=24 MSE-64.4803
CRITICAL VALUE OF STUDENTIZED RANGE=3.532

COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY  '***'


                         SIMULTANEOUS                              SIMULTANEOUS
                            LOWER        DIFFERENCES                  UPPER
     NOISE               CONFIDENCE        BETWEEN                 CONFIDENCE
   COMPARISON               LIMIT           MEANS                     LIMIT


   2          - 0          -6.240          0.444                  7.129
   2          - 1           2.318          9.208                 16.098   ***

   0          - 2          -7.129         -0.444                  6.240
   0          - 1           1.874          8.764                 15.654   ***

   1          - 2         -16.098         -9.208                 -2.318   ***
   1          - 0         -15.654         -8.654                 -1.874   ***


   0 = NO NOISE
   1 = INDUSTRIAL NOISE
   2 = FAST-FOOD RESTAURANT NOISE
```

Appendix XVIII. Tukey Analyses: Scenario 3

	GROUPING			CORRECT UTTERANCES
RECOGNIZER	A	B	MEAN	TOTAL # UTTERANCES
ITT	*		87.444	.87
VERBEX	*	*	74.222	.74
INTERSTATE		*	68.778	.69

Grouping by Connected Recognizer Performance

	GROUPING			CORRECT UTTERANCES
SPEAKER	A	B	MEAN	TOTAL # UTTERANCES
1	*		92.667	.93
2		*	73.889	.74
3		*	63.889	.64

Grouping by Speaker Performance - Connected Speech

	GROUPING		CORRECT UTTERANCES
NOISE	A	MEAN	TOTAL # UTTERANCES
FAST FOOD RESTAURANT	*	85.444	.85
NONE	*	75.667	.76
INDUSTRIAL	*	69.333	.69

Grouping by Noise Source - Connected Speech

Appendix XIX. Rejection/Misrecognition Matricies: Scenario 3

	NOISE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	TOTAL
VERBEX	NONE	1	4		4	3	1	2		3	4		3	2	3	3		2		3	7		3	2	5	3	58
	INDUSTRIAL	2	7	1	5	3		7	1	7	3	3	4	3	6	4	2	5	5	3	4	4	6	4	5	5	99
	FAST FOOD RESTAURANT	1	4	1	4	2		2		2	2		2	1	1	3					5					1	31
INTERSTATE	NONE		7	1	1	9	3	8	2	5	8	3	4	3	2	7	4	4	5	3	9	4	7	3	6	10	118
	INDUSTRIAL	3	6	1	4	7		8	2	5	6		8	2	1	6	1	6	5	1	8	3	5	3	3	7	101
	FAST FOOD RESTAURANT		1			3		3			1		3	1		1		4		1	1		1	1	1	1	23
ITT	NONE																										
	INDUSTRIAL								1									1						1			3
	FAST FOOD RESTAURANT																										

Rejections - Connected Speech Scenario 3

	VERBEX					INTERSTATE					ITT					
	0	1	2	TOTAL	$\bar{X}$	0	1	2	TOTAL	$\bar{X}$	0	1	2	TOTAL	$\bar{X}$	TOTAL
1 DIGIT	12	18	12	42	.23	18	21	4	43	.24	0	0	0	0	0	85
2 DIGIT	10	18	6	34	.19	26	21	4	51	.28	0	1	0	1	.01	86
3 DIGIT	11	20	7	38	.21	19	17	5	41	.23	0	0	0	0	0	79
4 DIGIT	12	19	5	36	.20	25	21	6	52	.29	0	1	0	1	.01	89
5 DIGIT	13	24	1	38	.21	30	21	4	55	.21	0	1	0	1	.01	94
SENT. 1	2	11	1	14	.08	14	7	0	21	.12	0	0	0	0	0	35
SENT. 2	14	29	8	51	.28	30	33	12	75	.42	0	1	0	1	.01	127
SENT. 3	4	14	2	20	.11	14	13	2	29	.16	0	2	0	2	.02	51
SENT. 4	18	26	7	51	.28	17	14	2	33	.18	0	0	0	0	0	84
SENT. 5	20	19	13	52	.29	43	34	7	84	.47	0	0	0	0	0	136

Rejection by Sentence - Scenario 3

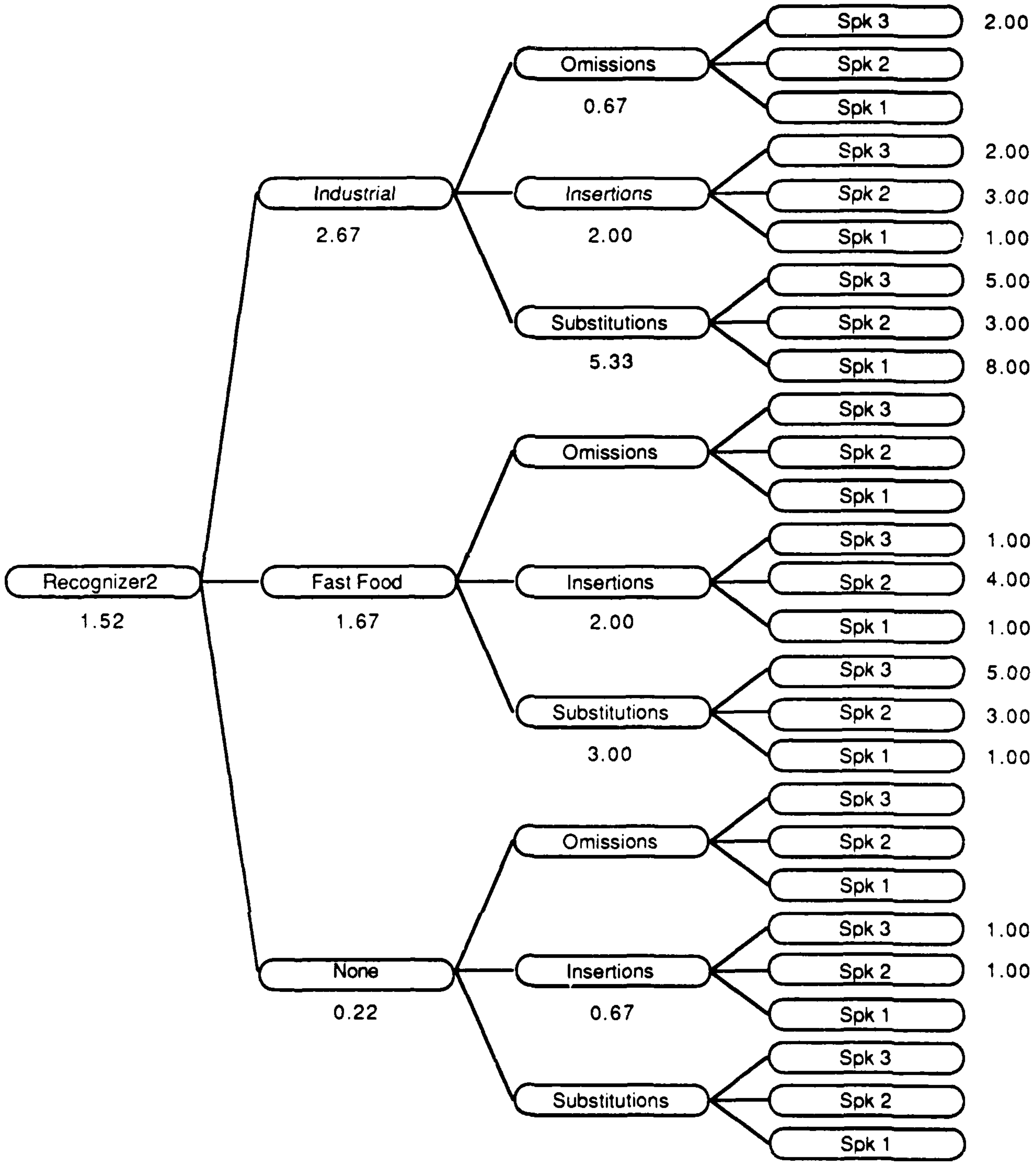
	VERBEX					INTERSTATE					ITT					
	0	1	2	TOTAL	$\bar{X}$	0	1	2	TOTAL	$\bar{X}$	0	1	2	TOTAL	$\bar{X}$	TOTAL
1 DIGIT	2	1	0	3	.02	0	3	2	5	.03	3	5	5	13	.07	21
2 DIGIT	3	10	3	16	.09	0	2	4	6	.03	7	9	9	25	.14	47
3 DIGIT	4	2	6	12	.07	2	4	5	11	.06	5	10	9	24	.13	47
4 DIGIT	2	5	4	11	.06	0	4	8	12	.07	10	10	9	29	.16	52
5 DIGIT	0	1	1	2	.01	0	0	6	6	.03	5	8	7	20	.11	28
SENT. 1	2	6	5	13	.07	0	4	3	7	.04	4	5	7	16	.09	36
SENT. 2	4	3	2	9	.05	1	1	8	10	.06	4	18	14	36	.20	55
SENT. 3	0	4	1	5	.03	1	3	1	5	.03	7	7	9	23	.13	33
SENT. 4	5	4	6	15	.08	0	5	5	10	.06	5	4	6	15	.08	40
SENT. 5	0	2	0	2	.01	0	0	8	8	.04	10	8	3	21	.12	31

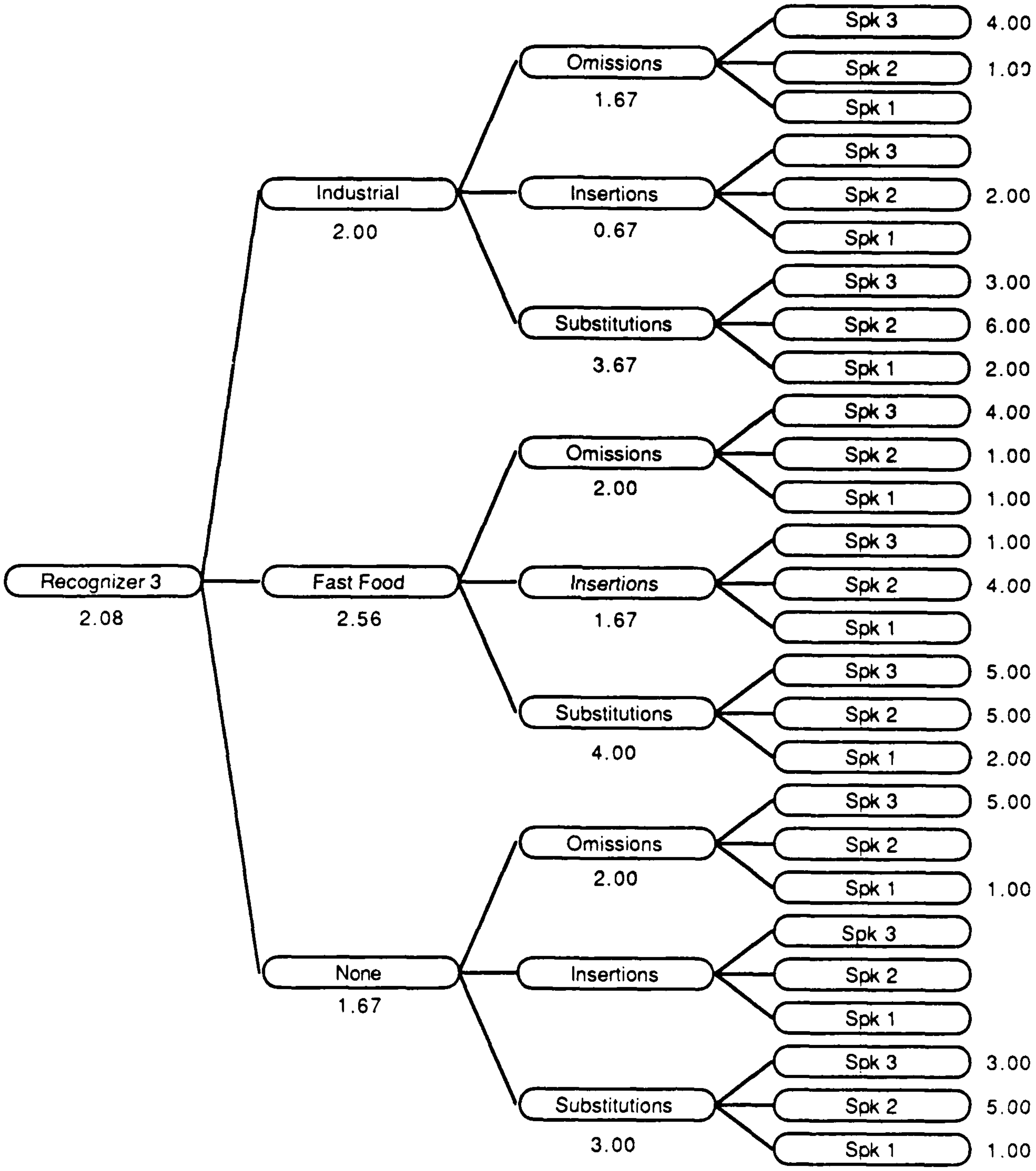
Misrecognition by Sentence - Scenario 3

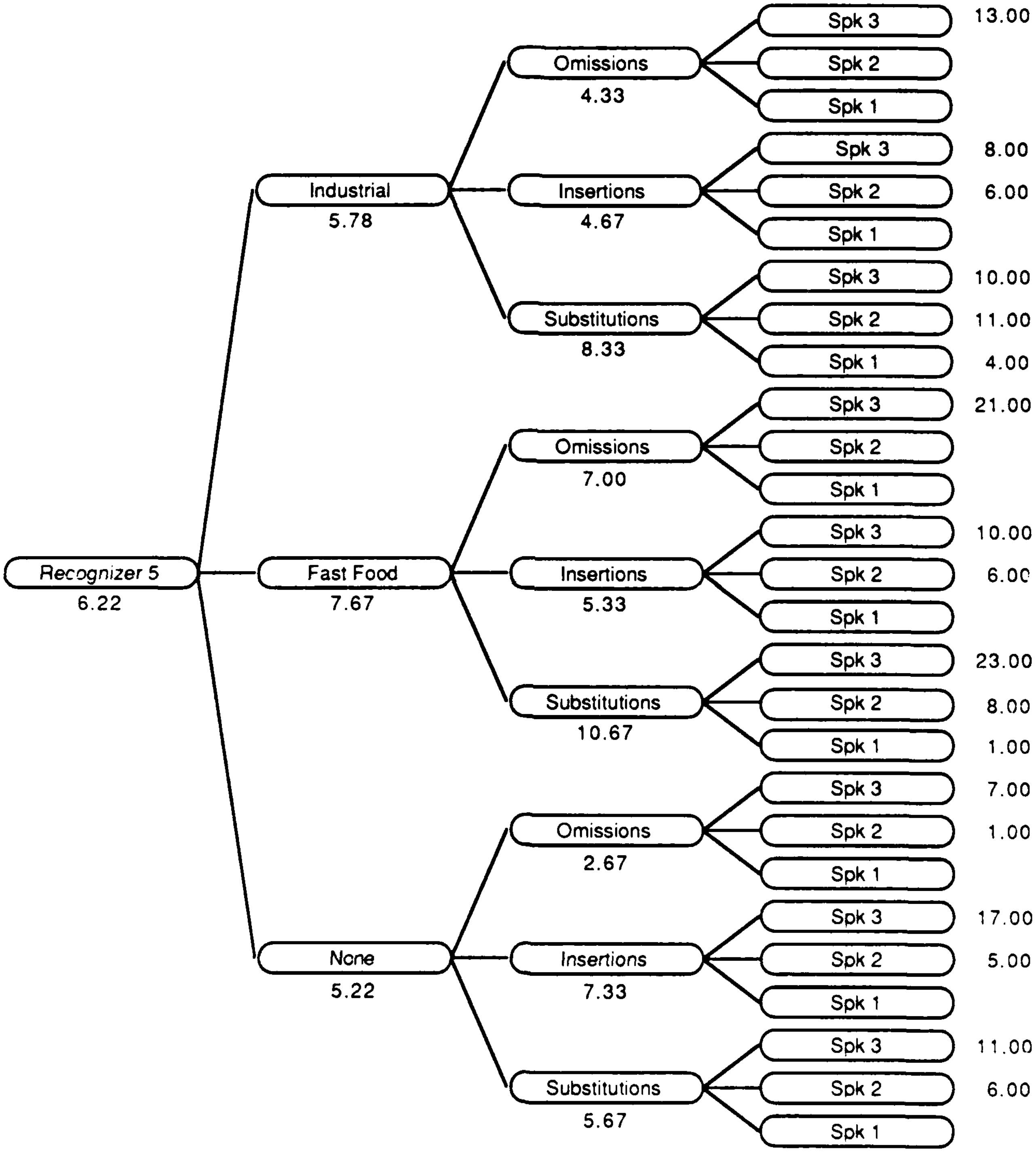
	NOISE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	TOTAL
VERBEX	NONE	1	1				1	2							4			1		1							11
	INDUSTRIAL			1			6	1	2	1			1		1			1	1	1	2				1		19
	FAST FOOD RESTAURANT						3					1	1		4		1		1	2			1				14
INTERSTATE	NONE												1	1													2
	INDUSTRIAL		1	1	1		2					1		2	1		1			3							13
	FAST FOOD RESTAURANT		2				2	1			1					3	1	1	1	2	3		2		3	1	25
ITT	NONE			3			2		1		4	1				4		1	2	5	2	1	3	1			30
	INDUSTRIAL		4	1			1	4	2		2		6	1		3	2		1	4	3	2	4	2			42
	FAST FOOD RESTAURANT		2	3			2	3	4			2	6		1		1	1	2	2	3	2	2		3		39

Misrecognitions - Connected Speech Scenario 3

Appendix XX. Misrecognition Error Trees: Scenario 3




Recognizer 3
2.08
Industrial
2.00
Omissions
1.67
Spk 3
4.00
Spk 2
1.00
Spk 1
Insertions
0.67
Spk 3
Spk 2
2.00
Spk 1
Substitutions
3.67
Spk 3
3.00
Spk 2
6.00
Spk 1
2.00
Fast Food
2.56
Omissions
2.00
Spk 3
4.00
Spk 2
1.00
Spk 1
1.00
Insertions
1.67
Spk 3
1.00
Spk 2
4.00
Spk 1
Substitutions
4.00
Spk 3
5.00
Spk 2
5.00
Spk 1
2.00
None
1.67
Omissions
2.00
Spk 3
5.00
Spk 2
Spk 1
1.00
Insertions
Spk 3
Spk 2
Spk 1
Substitutions
3.00
Spk 3
3.00
Spk 2
5.00
Spk 1
1.00


Recognizer 5
6.22
Industrial
5.78
Omissions
4.33
Spk 3
13.00
Spk 2
Spk 1
Insertions
4.67
Spk 3
8.00
Spk 2
6.00
Spk 1
Substitutions
8.33
Spk 3
10.00
Spk 2
11.00
Spk 1
4.00
Fast Food
7.67
Omissions
7.00
Spk 3
21.00
Spk 2
Spk 1
Insertions
5.33
Spk 3
10.00
Spk 2
6.00
Spk 1
Substitutions
10.67
Spk 3
23.00
Spk 2
8.00
Spk 1
1.00
None
5.22
Omissions
2.67
Spk 3
7.00
Spk 2
1.00
Spk 1
Insertions
7.33
Spk 3
17.00
Spk 2
5.00
Spk 1
Substitutions
5.67
Spk 3
11.00
Spk 2
6.00
Spk 1

END

6-87

DTIC