

UNCLASSIFIED

Defense Technical Information Center
Compilation Part Notice

ADP014017

TITLE: Evaluation of ASR Sensors

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Multi-modal Speech Recognition Workshop 2002

To order the complete compilation report, use: ADA415344

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:

ADP014015 thru ADP014027

UNCLASSIFIED

Evaluation of ASR Sensors

Justin Taylor¹, Jason Heinrich², Jung H. Kim¹, Sung H. Yoon²

¹Department of Electrical and Computer Engineering

²Department of Computer Science

North Carolina A&T State University

Greensboro, NC 27411

Tel: (336) 334-7760 x 219 Fax: (336)334-7244

E-mail: jt981532@ncat.edu or jh015778@ncat.edu

Abstract

This paper addresses the testing and analyzing of various microphones versus the Physiological Microphone (provided by Pete Fisher of the Army Research Laboratory) in different working conditions [1,2]. We explore different techniques and environments in which a user interfaces a selected ASR program. The testing of multiple microphones provided us with varied results based on environment. The software of choice for our research was Dragon Naturally Speaking 5.0.

1. Introduction

Automatic Speech Recognition systems enable users to operate their computer through the use of their voice. This advancement has benefited casual consumers, professionals and handicapped individuals alike. The development of a microphone allowing the user to move about freely and eliminate background noise has become necessary for practical use by professionals and consumers alike. Although significant progress has been made in ASR there are still limitations that must be taken into consideration. The technology that is on the market for consumers today, operates efficiently only under controlled conditions and through dictation, not conversation.

Factors to be considered in recognition accuracy:

- Environment (background noise, room size)
- Computer Hardware (CPU speed, RAM, soundcard)
- Amount of training with software
- Position of microphone
- Speaking style and clarity

- Microphone type
- Variability in the consumers speech (e.g., stress, colds)

These factors are considered to determine the most effective speech recognition procedure for each microphone based on environment.

2. System Descriptions

Our research was recorded based on the results provided by two test machines. The machines were both using Intel based processors.

System A

- Pentium III 0.5 GHz
- 256 Mb pc133 RAM
- Yamaha DS-XG Sound Card

System B

- Pentium IV 1.4 GHz
- 256 Mb RDRAM
- Sound Blaster Live! 5.1

System C

- Pentium IV 1.4 GHz
- 256 Mb RDRAM
- Sound Blaster Live! 5.1

System D

- Pentium IV 1.8 GHz
- 256 Mb RDRAM
- SoundBlaster Live! 5.1

The testing phase of the research continued through the use of four styles of microphones.

Microphone types:

- Telex H-551 Headset Microphone (Reference Mic.) (System B)
 - USB digital stereo headset
- Physiological Microphone (P-Mic)
 - Throat Microphone that detects vibration through skin and bone (System A)
- Telex M-60
 - Super-directional linear array microphone (System C)
- Telex M-40
 - Standard desktop microphone (System D)

Our findings were based on the aforementioned hardware combined with a predetermined method of testing. All computers exceeded the hardware requirements of Dragon Naturally Speaking v5.0. Through preliminary testing, we found all recognizer engines operated at the same speed when dictating. Therefore, microphones were arbitrarily assigned to each computer.

3. P-Mic Description

The Physiological Microphone is optimized for hands-free use. The microphone is designed to eliminate most background noise. It has its own power source, which is a 7.5-volt silver-oxide battery. Two of the microphones we used were a stationary desktop microphone (Telex M-40) and a super-directional linear array based microphone (Telex M-60). The P-Mic has a power switch allowing the user to pause in dictation without having to remove the microphone or stop the program. The Telex M-40 is lacking a power switch, which is inconvenient in ASR. Physically, the P-Mic does not resemble a typical microphone. The P-Mic is worn like a collar, and has a silicon contact sensor which is placed slightly to the left or right of the throat, due to the symmetrical nature of the throat. The P-Mic is small and lightweight. The width of the collar and diameter of the sensor is about 1 inch. With the P-Mic the user can move about freely and have both hands available. Traditional microphones used in ASR require that the user remain stationary, thus limiting productivity in the workplace. The P-Mic plugs into the "Line-In" jack on the sound card via a phono plug, whereas traditional microphones use the microphone jack.

4. Procedure for Microphone Testing

Testing was performed in a typical, quiet research laboratory environment. Our research lab's dimensions are 22' x 17'. The room is prone to little outside noise interference. A radio playing a recorded talk radio conversation at variable volumes was used to produce background noise. The recorded talk radio show was selected for consistency, allowing each microphone to be subject to the same interference. The simulated conversation source was emitted 10' behind the speaker.

Before testing we positioned four computers such that they could be tested simultaneously by one user. Each of the four microphones was assigned arbitrarily to a computer. We then performed the basic training required according to the Dragon Naturally Speaking documentation. Next a 400-word passage was dictated once while correcting and training all errors that occurred. The 400-word passage contained general vocabulary. After training, the Telex M-40 and Telex M-60 were attached to a microphone stand and positioned directly in front of the speaker. The user then attached the H-551 and the P-mic enabling all four microphones to be tested at the same time. The speaker tested each microphone with background noise set at; no additional noise, 60dB, 70dB, and 80dB respectively. The environment where we tested had an average of 50 dB of background noise. The quiet conditions were to facilitate the peak performance of each of the four microphones.

The speaker then started Dragon Naturally Speaking on all four computers. The speaker read the passage speaking at an average volume of 80dB. With the speaker speaking at 80 dB and noise at 50 dB, the difference of 30 dB provides an ideal speech-to-noise ratio for ASR. The speaker's volume was chosen to keep him from resisting the urge to compete with the added background noise, especially at the highest level of noise (80dB). This allowed the experiment to be performed at speech-to-noise ratios varying from excellent to very poor for speech recognition purposes. Each test was performed three times per sound level and the results were averaged. The dictated passages were printed and saved for analysis of mistakes made during dictation.

5. Results

The results for the four microphones tested are documented in the plot below. Results per microphone in each environment are the average of three test sessions, recording the accuracy rate. The equation we used was $[(\text{Errors} / \text{Total Words}) * 100 = \text{Percent Error}]$; then, $[100 - \text{Percent Error} = \text{Accuracy Rate}]$. Each capitalization error, period, paragraph indentation, etc. was counted as an error, and a wrong word or a skipped word was counted as one error. Therefore, Type I and Type II errors were counted as one error. Multiple word phrases recorded in error in the place of one word were counted as one error (example: user says, "comma" and program records, "come on", = one error).

Table 1 contains the results for the microphones tested at each level of background noise. The last column depicts the total percentage change from quiet conditions to 80dB background noise.

Table 1. (Performance in %)

Mic. Type	No Noise	60dB	70dB	80dB	Total Chg.
H551	99.0	98.5	96.5	89.75	9.25%
M-60	98.75	97.25	92.5	85.25	13.5%
M-40	95.5	94.25	87.5	81.75	13.75%
P-Mic	97.5	96.0	93.75	92.0	5.5%

The graph below illustrates that microphone performance was above 94% accuracy when speech-to-noise ratios were ideal. Notice that the steepest drop for the acoustic microphones occurred between 70 and 80 dB, whereas the slope of the P-Mic continues along a fairly straight line. The P-Mic never dropped more than 3% between increased levels of background noise.

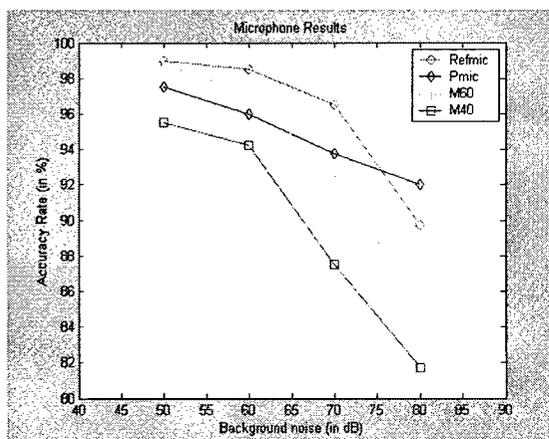


Figure 1 (Combined Results)

Table 2 breaks down the percent change in increased background noise. The acoustic microphones' performance all dropped in parallel as the levels of background noise were increased. The P-Mic's performance, on the other hand, did not decrease at a higher percentage with the addition of background noise. (Specifically from 60 to 70dB versus 70 to 80dB.

Table 2. (Percent Change)

Mic. Type	No Noise to 60dB	60 to 70dB	70 to 80dB
H551	0.5%	2.0%	5.25%
M-60	1.5%	4.75%	7.25%
M-40	1.25%	2.75%	5.75%
P-Mic	1.5%	2.25%	1.75%

5. Conclusions

It is concluded that the Physiological Microphone out performed its competition the most at the most stressful speech-to-noise ratios. The physiological microphone's performance was relatively unhampered by very poor speech-to-noise ratios. Our acoustic microphones' largest drop in recognition accuracy occurred at 80dB. The acoustic microphones dropped at least 5% at this level, whereas the P-Mic dropped only 1.75%. The P-Mic's total percent change of errors was about to half that of the reference microphone. Although the P-Mic performed above the rest, the 99% accuracy at quiet conditions still eluded it. Our data leads us to believe that the P-Mic has great potential when used in high background noise areas. We feel that the addition of an acoustic sensor used in tandem with the Physiological Microphone will boost recognition accuracy.

6. Future Endeavors

In the near future, we plan on acquiring a more accurate sound level meter, with a low range of 30dB. We would also like to acquire an electronic mouth to aid in our normalization process. Plans to create and implement a throat/neck simulator are also being arranged. This simulator, used with the electronic mouth will allow for a minimum of user errors and a near complete normalization of the test environment when using a pre-recorded file. We are also interested in acquiring other throat

sensors and testing their performance versus the Physiological Microphone.

7. References

[1] Pete Fisher: "Physiological Sensor For Speech Recognition", Proc. MultiModal Speech Recognition Worksop, Greensboro, NC (2002).

[2] Pete Fisher: "Alternative Speech Sensors For Military Applications", Proc. MultiModal Speech Recognition", Greensboro, NC (2002).