Department of Veterans Affairs

Journal of Rehabilitation Research and Development Vol. 28 No. 3, 1991
Pages 59–66

Experiments in the application of isolated-word recognition to secondary driving controls for the disabled

Eric C. Quintin, MS; Shivram K. Halan, BS; Kadry A. Abdelhamied, PhD
SYM Systems, Sunnyville, CA 94086; Department of Biomedical Engineering, Louisiana Tech University, Ruston, LA 71272

Abstract—Adaptive driving controls such as mechanical hand controls or electromechanical contact switches are now available which allow even the most severely impaired to drive. The residual functions, however, are all directed toward primary controls such as steering, braking, and accelerating, limiting the ability to operate secondary controls such as the horn, turn signals, ignition, and headlights, etc. In this paper, we discuss the application of speech recognition technology when operating these secondary controls. The performance of a speech recognition system inside a vehicle is studied, and the types of noise that degrade the recognition accuracy are also identified. Results are presented on the degradation in recognition performance caused by engine noise, fan noise, and interfering speech.

Key words: *driving conditions, microphone, primary and secondary controls, severely impaired, speech recognition system, vehicle adaptations.*

INTRODUCTION

With the aid of modern technology, a person with severe physical impairments can function well in a given environment. It is the environment that turns an impairment into a disability (16). If the environment can be changed in such a way that the impairment no longer places limitations on the person, then the disability can be overcome. Environmental barriers limit functioning, thereby reducing independence, which is a key issue. Transportation is one of the most important contributing factors in achieving independence. Of the two options, public transportation or independent transportation, most disabled persons prefer the latter, since it offers greater flexibility and a sense of independence (3).

Depending on the extent of the impairment, adaptive devices may be needed to operate some or all of the driving controls (which can be divided into primary controls and secondary controls). Primary controls, which affect the direction and speed of the vehicle, include steering, braking, and accelerating. Secondary controls, which are essential to the coordination, regulation, and safe operation of the vehicle, include the ignition switch, headlights, horn, turn-signal indicators, windshield wipers, sun visor, and air conditioner, etc. Currently, adaptive driving controls are usually mechanical aids such as a joystick, or electrical and electromechanical aids activated by various types of contact switches. For most disabled drivers, these switches are adequate to operate the vehicle with minimal difficulty. However, drivers with severe motor impairments, such as quadriplegia or double amputations of the upper extremities, may need to use all of their residual functioning to operate the primary controls. These individuals will not be able to activate the secondary controls that must be used simultaneously with the primary controls for the safe operation of the vehicle. Adaptive devices, such as head or chin switches, either affect the visual field of the driver, or are a source of diversion. In addition, it may be awkward trying to activate more than one switch with the head or chin while driving. Since most people with disabilities often have the ability to speak, a more convenient approach for operating the secondary controls is the use of speech. A speech recognition system could be

Address all correspondence and requests for reprints to: Kadry Abdelhamied, PhD, PO Box 10348 TS, Department of Biomedical Engineering, Louisiana Tech University, Ruston, LA 71272.

DOI: 10.1682/JRRD.1991.07.0059

Journal of Rehabilitation Research and Development Vol. 28 No. 3 Summer 1991

used to activate as many functions as needed without interfering with the driver's visual field.

The main factors that determine the complexity of a speech recognition system are the mode of speech presented as its input, the vocabulary size, and the number of speakers (15). Isolated word recognition systems require the speaker to pause between words, so that the system can determine the beginning and ending of a word, which is crucial for accurate recognition. Recognition is performed by comparing the "test" word with the reference templates previously recorded by the subject. The template that is closest to it is determined to be the word spoken by the subject (12,14,15). Vocabulary size affects the computation time, and mismatches occur when the number of words having similar templates increases. In speaker-dependent systems, the system is trained to recognize the voice of the persons who will be using the system. Speaker-dependent, limited-vocabulary, isolated-word recognition systems are adequate to operate secondary controls for adaptive driving. In such an application, the system is only required to recognize a few dozen of the short commands produced by a few speakers.

However, conditions inside a vehicle under which the system operates are less than ideal. There is background noise present that will affect the performance of the system. Vehicle interior noise levels of up to 75 dBa were reported for vehicles running at 55 mph on a smooth road. Sources of noise inside a vehicle include the engine, fan, talking passengers, and traffic (2). It is generally agreed that speech recognition systems exhibit degraded performance under noisy conditions (4,10); that the environment in which the system is to be used is the main determinant of the kind of additive noise that may be present (11); and that the nature of the interfering noise will dictate what methods can (or cannot) be used to reduce the signal degradation caused by the noise. Different methods include close-speaking noise-canceling microphones (6,8,9,19), signal processing (1,5,7,18), and training the system in the same noisy environment in which the system is to be tested (14,15,17).

The literature pertaining to practical application of speech recognition and improving speech recognition in noisy conditions does not report on speech recognition inside a vehicle. Consequently, it is not possible to rely on previous work to determine how various types of noise encountered inside a vehicle (such as engine noise, fan/blower noise, and talking passengers) will degrade recognition accuracy. It is the purpose of this study to determine how different sources of noise inside a vehicle contribute to degraded recognition performance. Methods that can be used to eliminate the effect of each, and/or combinations of these sources of noise, will be investigated. Also, design criteria for the development of voice-operated secondary controls for adaptive driving for the disabled will be outlined. A set of experiments was conducted to address these issues.

SPEECH RECOGNITION EXPERIMENTS

System configuration

Random noise represented by interfering speech and continuous noise including engine noise and fan/blower noise were chosen as a representative sample of noise for the experiments. Driving conditions were simulated by conducting the training and testing in an actual vehicle (parked during testing). The vehicle used was a modified Collins (Omni Baron) window van with a Ford E-350 chassis. The speech recognition system chosen was the IBM-compatible add-on board, the Introvoice VI by Voice Connection. Introvoice VI is an isolated-word, speaker-dependent system capable of recognizing up to 500 words (20). The microphone used was an Audio-Technica AT9100, close-speaking unidirectional microphone with a sensitivity of -60 dBm. The vocabulary consists of 34 typical driving control words taken directly from a control panel of a contact switch-activated adaptive driving control system. A list of these words is given in **Table 1**. Two male graduate students, both of whom spoke English as their native language, participated in the experiments.

Training and testing

The experiments were set up to obtain baseline performance results for the recognition system under nearly ideal conditions, and then under various training and testing noise conditions inside the vehicle. The words in the vocabulary were read one after another into the microphone until the list was exhausted. The procedure was repeated for a total of four passes to constitute the training process.

The system was tested under various conditions with the microphone distance at 2 inches and 12 inches. The distance of 2 inches is recommended by the Introvoice VI user's manual, while 12 inches is the distance between the center of the steering wheel and the driver's mouth. In testing the system, the subject was prompted to read the word that appeared on the screen, and to verify whether the word was correctly recognized. Three types of errors were encountered. They are: *deaf*—spoken word is not heard

Table 1.
Vocabulary used in the experiments.

Left turn	Radio on	Fold
Right turn	Radio off	Unfold
Horn	Radio volume	Up
Dimmer	Scan stations	Down
Wiper	Temperature mode	door lock
Shift	Temperature control	Lights
Cruise on	Fan	Release parking brake
Cruise off	Interior lights	Parking brake
Cruise set	Wash	Ignition off
Cruise coast	Lift mode	Start
Left window	door open	
Right window	Door close	

at all; *substitution*—spoken word is recognized as a different word in the vocabulary; and, *no match*—spoken word is heard but not recognized as any of the words in the vocabulary (speaker beeps).

Experiment #1

Experiment #1 was carried out in a quiet room measuring 9 × 12 ft. Training was done by reciting the vocabulary into the microphone four times by each subject, with the microphone distance being 2 inches away from the speaker. The system was tested under the same conditions in which it was trained. For the second phase of the testing, the speech templates that were trained with the microphone at 12 inches were used as references. The procedure was repeated, but with the microphone at 12 inches, to match the training distance. The frequency of occurrence of each type of error, and the word substituted in the case of substitution error, was recorded for both of the conditions (microphone at 2 inches, and at 12 inches).

Experiment #2

Experiment #2 was carried out in a quiet van. The recognition system was moved to the van with the background condition of no noise. The procedure discussed in Experiment #1 was repeated. Testing was performed using the reference template derived from training in the quiet van, and with the reference template of the first experiment. This was necessary to see if any correlation existed between the environment the system is trained and tested in, and the performance of the recognition system.

Experiment #3

Experiment #3 was carried out in the van with the engine running and the overhead fan on. The continuous noise level next to the microphone was measured with a Realistic Sound-Level Meter (Radio Shack catalog number 33-2050), and was found to be 64 dBa. This condition is referred to as a "noisy" condition. The windows were rolled up to eliminate white noise. The training and testing was the same as before. The number of conditions was increased as we included the reference templates from both the previous experiments.

Experiment #4

This experiment was conducted in the van with the engine running, overhead fan on, and the front fan on. Reference templates from the previous training were used. The front fan was used because the overhead fan was not a realistic source of noise while driving. The front fan was kept at high speed and the noise level was about 72 dBa.

Experiment #5

All the conditions of the other experiments were used, along with the addition of an interfering speaker. Templates from earlier experiments were used. The voice of a talking passenger is random and varies for different passengers. Recognition performance in the presence of an interfering speaker was tested. The first subject, T1, was the interfering speaker for the second subject, T2, and vice versa. The interfering speech was maintained at a constant sound level, and was made up of the same words used in the test

Journal of Rehabilitation Research and Development Vol. 28 No. 3 Summer 1991

Table 2.
Results of the experiments under various conditions.

Sub	Training Env and Mic. dist in Inches	Recog. Env. and Mic. dist in Inches	Recog. accuracy in % (n=136)	Types of Errors: Deaf	Recognition Errors: Substitution	Recognition Errors: No match
T1 EXP 1	QR 2	QR 2	97	1	3	0
	QR 12	QR 12	90	6	8	0
T2 EXP 1	QR 2	QR 2	100	0	0	0
	QR 12	QR 12	84	17	5	0
T1 EXP 2	QR 2	QV 2	96	0	5	0
	QV 2	QV 2	97	1	3	0
	QR 12	QV 12	86	2	17	0
	QV 12	QV 12	98	0	2	0
T2 EXP 2	QR 2	QV 2	93	1	8	0
	QV 2	QV 2	98	3	0	0
	QR 12	QV 12	69	8	34	0
	QV 12	QV 12	90	7	6	0
T1 EXP 3	QR 2	NV 2	90	0	13	0
	QV 2	NV 2	98	0	3	0
	NV 2	NV 2	100	0	0	0
	QR 12	NV 12	71	11	29	0
	QV 12	NV 12	90	1	12	0
	NV 12	NV 12	97	0	4	0
T2 EXP 3	QR 2	NV 2	99	0	2	0
	QV 2	NV 2	100	0	0	0
	NV 2	NV 2	100	0	0	0
	QR 12	NV 12	57	18	40	0
	QV 12	NV 12	97	2	2	0
	NV 12	NV 12	98	2	1	0
T1 EXP 3	NV 2	QV 2	96	3	2	0
	NV 12	QV 12	92	0	11	0
T2 EXP 3	NV 2	QV 2	96	2	4	0
	NV 12	QV 12	86	7	12	0
T1 EXP 4	QV 2	NVF 2	83	0	23	0
	QV 12	NVF 12	97	0	4	0
T2 EXP 4	QV 2	NVF 2	96	0	5	0
	QV 12	NVF 12	93	0	9	0
T1 EXP 5	QV 2	same + IS	71	0	38	1
	QV 12	same + IS	35	0	16	0
	NV 2	same + IS	88	0	16	0
	NV 12	same + IS	35	0	18	0
T2 EXP 5	QV 2	same + IS	88	0	16	0
	QV 12	same + IS	56	2	14	0
	NV 2	same + IS	86	8	11	0
	NV 12	same + IS	53	0	17	0

QR—Quiet Room; QV—Quiet Van; NV—Noisy Van (Engine Running and Overhead Fan On); NVF—Noisy Van With Front Fan On; IS—Interfering Speaker.

vocabulary, to determine whether the system would be confused by the same words being spoken by a different speaker. The cassette player was placed vertically in the passenger's seat. The volume was set at conversational level, and the noise level was between 65 dBa and 85 dBa.

RESULTS AND DISCUSSION

Recognition accuracy results from the above experiments are summarized in **Table 2**. Included in the table are the different types of errors that occurred. Examining

the table, it is shown that the Introvoice VI recognition system was capable of achieving accuracies of up to 100 percent when tested under nearly ideal conditions of training, and testing the system in a quiet room with the microphone at 2 inches. Lower recognition accuracies were obtained for different training and testing environments (see **Figure 1**), for the microphone distance at 12 inches (see **Figure 2**), and for noisy environments (see **Figure 3**).

Training/testing conditions

As shown in **Figure 1**, the recognition performance was better for identical training and testing conditions; it degraded when the two environments were different. This degradation was higher at the 12-inch position than at the 2-inch position. For identical training and testing conditions at the 2-inch position, the system could perform as well in noisy conditions as it did in the quiet room. Apparently the microphone was effective in canceling the noise (72 dBa) at this close speaking distance. The results also indicate that the performance was better when training was done in a quiet van, and tested in a noisy van, than in the reverse case. For example, the accuracy for T2 dropped noticeably from 97 percent to 86 percent when the training and testing conditions were reversed.

Microphone distance

Generally, recognition rates were lower at the 12-inch position than at the 2-inch position, as shown in **Figure 2**. At 12 inches, and for training and testing done in a quiet room, these rates were 90.0 percent for T1, and 84.0 percent for T2, as compared to 97.0 percent and 100.0 percent, respectively, at the 2-inch position. **Figure 2** also shows that the system could perform in noisy conditions, as well as in a quiet room at the 2-inch position. At the 12-inch position, the recognition performance was extremely poor, particularly when the training and testing conditions were different. For example, the recognition rate obtained for T2 was 57 percent when the system was trained in a quiet room and tested in a noisy van. The predominant type of error in this case was substitution error. These probably occurred due to the differences in the acoustic properties of the two environments.

Noise condition

The type of noise also affects the performance. Continuous noise did affect the system performance, but it was not a major factor. This does not mean that we can disregard it, since the major errors occurring were substitution errors, and such errors cannot be accepted. When the interfering speaker was included in the testing environment,

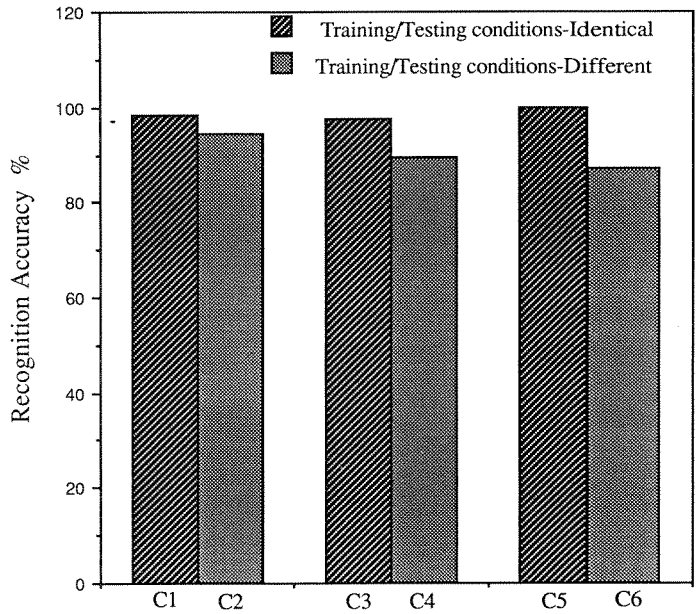

Figure 1.
Effect of training and testing conditions on recognition accuracy.
C1—Training/Testing in quiet room.
C2—Training in quiet room; Testing in noisy van.
C3—Training/Testing in quiet van.
C4—Training in quiet van; Testing in noisy van with front fan on.
C5—Training/Testing in noisy van.
C6—Training in noisy van; Testing in noisy van with interfering speaker.

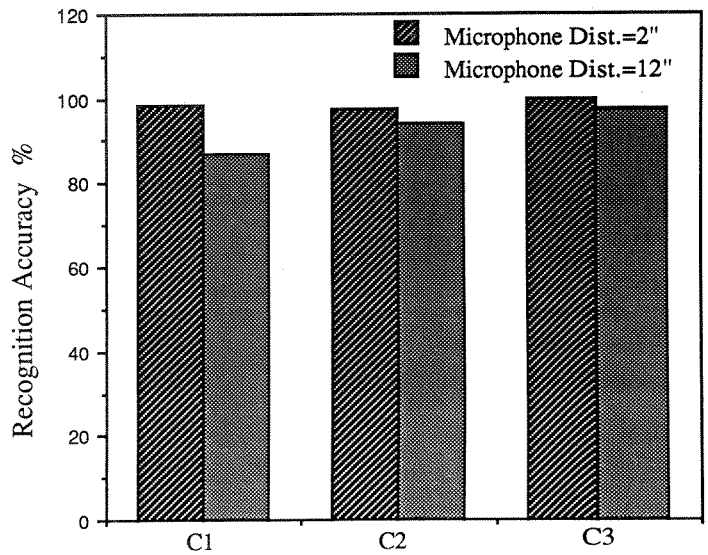

Figure 2.
Effect of microphone distance on recognition accuracy.
C1—Training/Testing in quiet room.
C2—Training/Testing in quiet van.
C3—Training/Testing in noisy van.

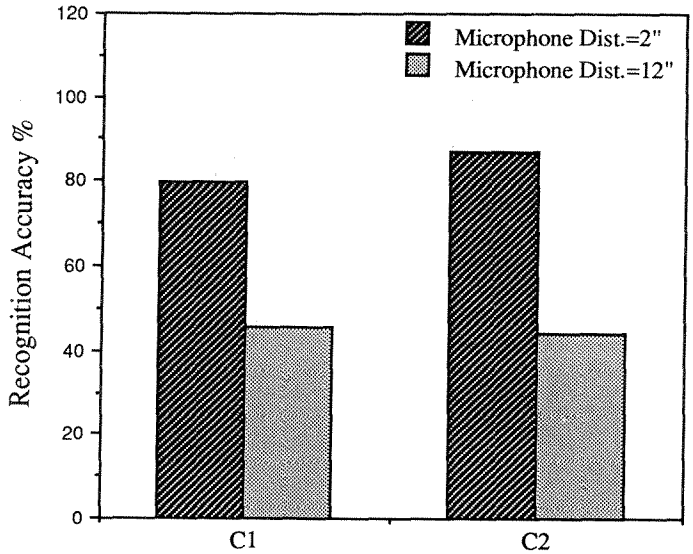

Figure 3.
Effect of interfering speech on recognition accuracy.
C1—Training in quiet van; Testing in quiet van with interfering speaker.
C2—Training in noisy van; Testing in noisy van with interfering speaker.

the performance was unacceptable. **Figure 3** shows that the performance accuracy dropped to 83.25 percent at the 2-inch position, and drastically to 44.75 percent at the 12-inch position when the interfering speech was included. These results indicate that interfering speech is the main difficulty in our application. Radio noise, which is similar in nature to interfering speech, is likely to cause similar problems. The noise level of interfering speech was 20 dB higher than the noise caused by the engine and overhead fan together, and about 12 dB higher than the noise caused by the front fan at high speed. The difference in noise levels explains why the degradation caused by interfering speech was higher than that caused by other noise conditions.

Analysis of variance

One-way analysis of variance (ANOVA) of recognition accuracy was used to test the significance of each factor. The results are summarized in **Table 3**. The results indicated that the two subjects were not significantly different ($p < 0.7758$, $F=0.085$). The results also indicated that changes in training and recognition conditions, with respect to noise level and physical environment, were correlated with recognition accuracy ($p < 0.0035$, $F=3.659$). This means that variations in training and recognition conditions had a significant effect on accuracy. To test whether recognition accuracy is higher when training and testing conditions were the same, conditions were grouped according to two possible relationships between training and testing—same or different. The results also show that the mean recognition accuracies for both groups are significantly different, and that performance is better when the conditions are the same ($p < 0.0187$, $F=6.038$). In addition,

Table 3.
Results of one-way ANOVA analysis of recognition accuracy.

Source of Variation (Factors)	Groups	F-ratio	Between Group Significance Level
Subject	T1,T2	.085	.7758
Training/Recognition Conditions (Conditions)	r/r, r/nv, qv/qv, qv/nv, qv/qvp, r/qv nv/nv,nv/qv, nv/nvp, qv/nvf	3.659	.0035
Training and Recognition Conditions (TR)	Same, Different	6.038	.0187
Microphone Distance (micdist)	2" ,12"	9.068	.0046
Engine and Overhead Fan (Engine)	On, Off	.008	.9291
Front Fan	On (high), Off	.553	.4694
Interfering Speaker	Yes,no	28.176	.0000

nv—noisy van; qv—quiet van; r—room; qvp—quiet van + interfering speaker; nvp—noisy van + interfering speaker; nvf—noisy van + front fan on.

it is indicated that the performance is significantly different for the two microphone positions ($p < 0.0046$, $F=9.068$). Engine and overhead fan noise was not correlated with recognition accuracy ($p < 0.9291$, $F=0.008$). Further, there is no strong support for a correlation between the front fan and recognition accuracy ($p < 0.4694$, $F=0.553$). Finally, the noise produced by the interfering speaker was highly correlated to the recognition accuracy, especially at a microphone distance of 12 inches ($p < 0.00001$, $F=28.176$).

The results from a two-factor ANOVA analysis indicated the factors that showed strong correlation between microphone distance and talking passenger ($p < 0.0002$, $F=16.68$), as well as between microphone distance and training/recognition conditions ($p < 0.0013$, $F=4.98$). From this we see that the effect the interfering speech has on recognition accuracy depends on the microphone distance. It can also be concluded that all three main factors (microphone distance, talking passenger, and training testing conditions) have significant effects on recognition accuracy.

CONCLUSIONS

It has been determined that certain conditions inside a vehicle contribute to degradation in recognition accuracy. However, we have not yet defined how much degradation can be allowed before the system becomes unacceptable. One way is to determine the error rate for existing secondary control systems, and use that as the reference for the speech-driven system. The intended application is a critical factor in deciding the minimum level of accuracy. It is essential that very few errors occur if they are likely to cause accidents. The microphone distance, talking passengers, and similarity of training and testing environments were seen to be significantly correlated with recognition accuracy.

The performance of the system with the microphone at 2 inches was highly acceptable. The decision to use either configuration depends on the individual's needs with no compromise on safety. Close-speaking noise-canceling microphones (8,19) are suggested here to remove engine and fan noise (of continuous nature and below 90 dBa). Training the system in the environment in which it is to be used is also important to achieve high recognition accuracy, as has been suggested (14,15,17). These methods do not take changes in noise condition into consideration. They deal with continuous background noise and not random noise such as interfering speech. Removal of interfering speech is the most difficult problem of all (13).

This is the focus of our current work. It should be noted that in addition to noise causing recognition errors, the vocabulary used has an effect on recognition accuracy (18). In this study, some words were slightly modified. Examples include: "temperature mode control" to "temperature mode"; and, "park-brake engage" to "parking brake." No effort was made to optimize the vocabulary to minimize similarity between words. Modifying the vocabulary so that word similarity will not be a source of errors can also be used to achieve the accuracy required for this application.

REFERENCES

1. **Ahmed MS:** Comparison of noisy speech enhancement algorithms in terms of LPC perturbation inches. *IEEE Trans Acoust Speech Signal Process* 37(1):121-125, 1989.
2. **Bagga KS, Repick EP:** Development of an interior sound level measurement procedure for light vehicles—SAE J1477. In *Surface Vehicle Noise and Vibration Conference Proceedings,* 293-302, 1985.
3. **Hale P, Shipp M:** Consumer program for persons with disabilities. Report prepared for General Motors by Louisiana Tech University, Ruston, LA, 1989.
4. **Hanson B, Hisashi W:** Spectral slope distance measures with linear prediction analysis for word recognition in noise. *IEEE Trans Acoust Speech Signal Process* 35(7):968-973, 1987.
5. **Harrison W, Lim J, Singer E:** A new application of adaptive noise cancellation. *IEEE Trans Acoust Speech Signal Process* 34(1):21-27, 1986.
6. **Herrold R:** Microphones—Start with the basics. *Audio* 46-54, 1985.
7. **Lim J, Oppenheim A:** Enhancement and bandwidth compression of noisy speech. *Proc IEEE (USA)* 7(12):1586-1604, 1979.
8. **Martin TB:** Practical applications of voice input to machines. *Proc IEEE (USA)* 64(4):487-501, 1976.
9. **Michael W, Hill P:** Performance of a real time TMS32010-based adaptive noise canceller (ANC). *IEEE Trans Acoust Speech Signal Process* 36(3):411-412, 1988.
10. **Nadas A, Nahamoo D, Picheny M:** Speech recognition using noise-adaptive prototypes. *IEEE Trans Acoust Speech Signal Process* 37(10):1495-1503, 1989.
11. **O'Shaughnessy D:** Enhancing speech degraded by additive noise or interfering speakers. *IEEE Comm Mag,* 46-52, Feb. 1989.
12. **Parsons T:** *Voice and Speech Processing,* 358-360. New York: McGraw-Hill Book Company, 1987.
13. **Parsons T:** Separation of speech from interfering speech by means of harmonic selection. *J Acoust Soc Am* 60(4):911-918, 1976.

14. **Rabiner LR, Levinson S:** Isolated and connected word recognition-theory and applications. *IEEE Trans Commun* 29(5):621-659, 1981.
15. **Reddy D:** Speech recognition by machine: A review. *Proc IEEE (USA)* 64(4):501-531, 1976.
16. **Stolov WC (Ed.):** *Handbook of Severe Disability.* Washington, DC: U.S. Government Printing Office, 1981.
17. **Van Summers W, et al.:** Effects of noise on speech production: Acoustic and perceptual analyses. *J Acoust Soc Am* 84(3):917-928, 1988.
18. **Weiss M:** Use of an adaptive noise canceler as an input preprocessor for a hearing aid. *J Rehabil Res Dev* 24(4):93-102, 1987.
19. **White GM, Neely RB:** Speech recognition experiments with linear prediction, bandpass filtering and dynamic programming. *IEEE Trans Acoust Speech Signal Process* 24:183-188, 1976.
20. *IntroVoice VI User's Guide.* Irvine, CA: The Voice Connection, Inc., 1987.