

8

Speech Recognition and Synthesis

Douglas W. Beeks

Rockwell Collins

[8.1 Introduction](#)

[8.2 How Speech Recognition Works:
A Simplistic View](#)

Types of Speech Recognizers • Vocabularies • Modes of
Operation for Speech Recognizers • Methods of Error
Reduction

[8.3 Recent Applications](#)

[8.4 Flight Deck Applications](#)

Navigation Functions • Communication
Functions • Checklist

[Defining Terms](#)

[References](#)

[Bibliography](#)

[Further Information](#)

8.1 Introduction

The application of speech recognition (SR) in aviation is rapidly evolving and moving toward more common use on future flightdecks. The concept of using SR in aviation is not new. The use of speech recognition and voice control (VC) has been researched for more than 20 years, and many of the proposed benefits have been demonstrated in varied applications. Continuing advances in computer hardware and software are making the use of voice control applications on the flightdeck more practical, flexible, and reliable. There is little argument that the easiest and most natural and ideal way for a human to interact with a computer is by direct voice input (DVI).

While speech recognition has improved over the past several years, speech recognition has not reached the level of capability and reliability of one person talking to another. Using SR and DVI in a flightdeck atmosphere likely brings to mind thoughts of the computer on board the starship Enterprise from the science fiction classic *Star Trek*, or possibly of the HAL9000 computer from the movie *2001: A Space Odyssey*. The expectation of a voice control system like the computer on the Enterprise and the HAL9000 computer, is that it be highly reliable, work in adverse and stressful conditions, be transparent to the user, and understand its users accurately without having to tailor their individual speech and vocabulary to suit the system. Current speech recognition and voice control systems are not able to achieve this level of performance expectations, although the ability and flexibility of speech recognition and its application to voice control has increased over the past few years. Whether or not a speech recognition system will ever be able to function to the level of one person speaking to another remains to be seen.

The current accuracy rate of speech recognition is in the lower to mid 90% range. Some speaker-dependent systems, and generally those with small vocabularies, have shown accuracy rates into the upper 90% range. While at first glance that might sound good, consider that with a 90% accuracy rate, 1 in 10 words will be incorrectly recognized. Also consider that this 90% and greater accuracy may be under ideal conditions; many times this high accuracy rate is achieved in a controlled and sterile lab environment. Under actual operating conditions, including cockpit noise, random noises, bumps and thumps, multiple people talking at once, etc. the accuracy rate of speech recognition systems can erode significantly.

Currently, several military applications are planning on using SR to provide additional methods to support the Man-Machine Interface (MMI) to reduce the workload on the pilot in advanced aircraft. Boeing is incorporating SR into the new Joint Strike Fighter, and the Eurofighter Typhoon is also adding SR capabilities to its aircraft. Numerous aviation companies worldwide are conducting research and studies into how the available SR technology can be incorporated into current equipment designs and designs of the future for both the civilian and military marketplace. Speech recognition technology will likely be first used in military applications, with the technology working its way into civil aviation by the year 2005.

8.2 How Speech Recognition Works: A Simplistic View

Speech recognition is based on statistical pattern matching. One of the more common methods of speech recognition based on pattern matching uses Hidden Markov Modeling (HMM) comprising two types of pattern models, the acoustical model and the language model. Which of the two models will be used, and in some cases both will be required, depends on the complexity of the application. Complex speech recognition applications, such as those supporting continuous or connected speech recognition, will use a combination of the acoustical and language models.

In a simple application using only the acoustical model, the application will process the uttered word into phonemes, which are the fundamental part of speech. These phonemes are converted to a digital format. This digital format, or pattern, is then matched against stored patterns by the speech processor in search of a match from a stored database of word patterns. From the match, the phoneme, and word can be identified.

In a more complex method, the speech processor will convert the utterance to a digital signal by sampling the voice input at some rate, commonly 16 kHz. The required acoustical signal processing can be accomplished using several techniques. Some commonly used techniques are Linear Predictive Coding (LPC) cochlea modeling, Mel Frequency Cepstral Coefficients (MFCC), and others. For this example, the sampled data is converted to the frequency domain using a fast-Fourier transformation. The transformation will analyze the stored data at $1/30^{\text{th}}$ to $1/100^{\text{th}}$ of a second (3.3 ms to 100 ms) intervals, and convert the value into the frequency domain. The resulting graph from the converted digital input will be compared against a database of known sounds. From these comparisons, a value known as a feature number will be determined.

The feature numbers will be used to reference a phoneme found using that feature number. This, ideally, would be all that is required to identify a particular phoneme, however, this will not work for a number of reasons. Background noises, the user not pronouncing a word the same way every time, and the sound of a phoneme will vary, depending on the surrounding phonemes that may add variance to the sound being processed. To overcome problems of variability of the different phonemes, the phonemes are assigned to more than one feature number. Since the speech input was analyzed at an interval of $1/30^{\text{th}}$ to $1/100^{\text{th}}$ of a second and a phoneme or sound may last from 500 ms to 2 s, many feature numbers may be assigned to a particular sound. By using statistical analysis of these feature numbers and the probability that any one sound may contain those feature numbers, the probability of that sound being a particular phoneme can be determined.

To be able to recognize words and complete utterances, the speech recognizer must also be able to determine the beginning and the end of a phoneme. The most common method to determine the beginning and endpoint is by using the Hidden Markov Models (HMM) technique. The HMM is a state transition model and will use probabilities of feature numbers to determine the likelihood of transitioning from

one state to another. Each phoneme is represented by a HMM. The English language is made up of 45 to 50 phonemes. A sequence of HMM will represent a word. This would be repeated for each word in the vocabulary. While the system can now recognize phonemes, phonemes do not always sound the same, depending on the phoneme preceding and following it. To address this problem, phonemes are placed in groups of three, called tri-phones, and as an aid in searching, similar sounding tri-phones are grouped together.

From the information obtained from the HMM state transitions, the recognizer is able to hypothesize and determine which phoneme likely was spoken, and then by referring this to a lexicon, the recognizer is able to determine the word that likely was spoken.

This is an overly simplified definition of the speech recognition process. There are numerous adaptations of the HMM technique and other modeling techniques. Some of these techniques are neural networks (NNs), dynamic time warping (DTW), and combinations of techniques.

8.2.1 Types of Speech Recognizers

There are two types of speech recognizers, speaker-dependent and speaker-independent.

8.2.1.1 Speaker-Dependent Systems

Speaker-dependent recognition is exactly that, speaker dependent. The system is designed to be used by one person. To operate accurately, the system will need to be “trained” to the user’s individual speech patterns. This is sometimes referred to as “enrollment” of the speaker with the system. The speech patterns for the user will be recorded and patterned from which a template will be created for use by the speech recognizer. Because of the required training and storage of specific speech templates, the performance and accuracy of the speaker-dependent speech recognition engine will be tied to the voice patterns of a specific registered user. Speaker-dependent recognition, while being the most restrictive, is the most accurate, with accuracy rates in the mid to upper 90% range. For this reason, past research and applications for cockpit applications have opted to use speaker-dependent recognition.

The major drawback of this system is that it is dedicated to a single user, and that it must be trained prior to its use. Many applications will allow the speech template to be created elsewhere prior to use on the hosting system. This can be done at separate training stations prior to using the target system by transferring the created user voice template to the target system. If more than one user is anticipated, or if the training of the system is not desirable, a speaker-independent system might be an option.

8.2.1.2 Speaker-Independent Recognizers

Speaker-independent recognition systems are independent of the user. This type of system is intended to allow multiple users to access a system using voice input. Examples of speaker-independent systems are directory assist programs and an airline reservation system with a voice input driven menu system. Major drawbacks with a speaker-independent system, in addition to increased complexity and difficult implementation, are its lower overall accuracy rate, higher system overhead, and slower response time. The impact of these drawbacks continues to lessen with increased processor speeds, faster hardware, and increased data storage capabilities.

A variation of the speaker-independent system is the speaker-adaptive system. The speaker-adaptive system will adapt to the speech pattern, vocabulary, and style of the user. Over time, as the system adapts to the users’ speech characteristics, the error rate of the system will improve, exceeding that of the independent recognizer.

8.2.2 Vocabularies

A vocabulary is a list of words that are valid for the recognizer. The size of a vocabulary for a given speech recognition system affects the complexity, processing requirements, and the accuracy of that system. There are no established definitions for how large a vocabulary should be, but systems using smaller vocabularies can result in better recognizer accuracy. As a general rule, a small vocabulary may contain up to 100 words, a medium vocabulary may contain up to 1000 words, a large vocabulary may contain up to 10,000 words,

and a very large vocabulary may contain up to 64,000 words, and above that the vocabulary is considered unlimited. Again, this is a general rule and may not be true in all cases.

The size of a vocabulary will be dependent upon the purpose and intended function of the application. A very specific application may require only a few words and make use of a small vocabulary, while an application that would allow dictation or setting up airline reservations would require a very large vocabulary.

How can the size and contents of a vocabulary be determined? The words used by pilots are generally specific enough to require a small to medium vocabulary. Words that can or should be in the vocabulary could be determined in a number of ways. Drawing from the knowledge of how pilots would engage a desired function or task is one way. This could be done using a questionnaire or some similar survey method.

Another way to gather words for the vocabulary is to set up a lab situation and use the “Wizard of Oz” technique. This technique would have a test evaluator behind the scenes acting upon the commands given by a test subject. The test subject would have various tasks and scenarios to complete. While the test subject runs through the tasks, the words and phrases used by the subject are collected for evaluation. After running this process numerous times, the recorded spoken words and phrases will be used to construct a vocabulary list and command syntax, commonly referred to as a grammar. The vocabulary could be refined in further tests by only allowing those contained words and phrases to be valid, and have test subjects again run through a suite of tasks. Observations would be made as to how well the test subjects were able to complete the tasks using the defined vocabulary and syntax. Based on these tests, and the evaluation results, the vocabulary is modified as required.

A paper version of the evaluation process could be administered by giving the pilot a list of tasks, and then asking them to write out what commands they would use to perform the task. Following this data collection step, a second test could be generated having the pilot choose from a selected list of words and commands what he would likely say to complete the task. As a rule, pilots will tend to operate in a predictable manner, and this lends itself to a reduced vocabulary size and structured grammar.

8.2.3 Modes of Operation for Speech Recognizers

There are two modes of operation for a speech recognizer: continuous recognition, and discrete or isolated word recognition.

8.2.3.1 Continuous Recognition

Continuous speech recognition systems are able to operate on a continuous spoken stream of input in which the words are connected together. This type of recognition is more difficult to implement due to several inherent problems such as determining start and stop points in the stream and the rate of the spoken input.

The system must be able to determine the start and endpoint of a spoken stream of continuous speech. Words will have varied starting and ending phonemes depending on the surrounding phonemes. This is called “co-articulation.” The rate of the spoken speech has a significant impact on the accuracy of the recognition system. The accuracy will degrade with rapid speech.

8.2.3.2 Discrete Word Recognition

Discrete or isolated word recognition systems operate on single words at a time. The system requires a pause between saying each word. The pause length will vary, and on some systems the pause length can be set to determined lengths. This type of recognition system is the simplest to perform because the endpoints are easier for the system to locate, and the pronunciation of a word is less likely to affect the pronunciation of other words (co-articulation effects are reduced). A user of this type of system will speak in a broken fashion. This system is the type most people think of in terms of a voice recognition system.

8.2.4 Methods of Error Reduction

There are no real standards by which error rates of various speech recognizers are measured and defined. Many systems claim accuracy rates in the high 90% range, but under actual usage with surrounding noise conditions, the real accuracy level may be much less. Many factors can impact the accuracy of SR systems.

Some of these factors include the individual speech characteristics of the user, the operating environment, and the design of the SR system itself.

There are four general error types impacting the performance of a SR system; these are substitution errors, insertion errors, rejection errors, and operator errors,

- Substitution errors occur when the SR system incorrectly identifies a word from the vocabulary. An example might be the pilot calling out “Tune COM one to one two four point seven” and the SR system incorrectly recognizes that the pilot spoke “Tune NAV one to one two four point seven.” The SR system substituted NAV in place of COM. Both words may be defined and valid in the vocabulary, but the system selected the wrong word.
- Insertion errors may occur when some source of sound other than a spoken word is interpreted by the system as valid speech. Random cockpit noise might at some time be identified as a valid word to the SR system. The use of noise-canceling microphones and PTT can help to reduce this type of error.
- Rejection errors occur when the SR system fails to respond to the user’s speech, even if the word or phrase was valid.
- Operator errors occur when the user is attempting to use words or phrases that are not identifiable to the SR system. A simple example might be calling out “change the radio frequency to one one eight point six” instead of “Tune COM one to one one point eight six,” which is recognized by the vocabulary.

When designing a speech recognition application, several design goals and objective should be kept in mind:

- **Limitations of the hardware and the software** — Keep in mind the limitations of the hardware and the software being used for the application. Will the system need to have continuous recognition and discrete word recognition? Will the system need to be speaker independent, or will the reduced accuracy in using a speaker-independent recognizer be acceptable. Will the system be able to handle the required processing in an acceptable period of time? Will the system operate acceptably in the target environment?
- **Safety** — Will using SR to interface with a piece of equipment compromise safety? Will an error in recognition have a serious impact on the safety of flight? If the SR system should fail, is there an alternate method of control for that application?
- **Train the system in the environment in which it is intended to be used** — As discussed earlier, a SR system that has a 99% accuracy in the lab, may be frustrating and unusable in actual cockpit conditions. The speech templates or the training of the SR system needs to be done in the actual environment, or in as similar an environment as possible.
- **Don’t try to use SR for tasks that don’t really fit** — The problem with a new tool, like a new hammer, is that everything becomes a nail to try out that new hammer. Some tasks are natural candidates for using SR, many are not. Do not force SR onto a task if it is not appropriate for use of SR. Doing so will add significant risk and liability. Good target applications for SR include radio tuning functions, navigation functions, FMS functions, and display mode changes. Bad target applications for SR would be things that can affect the safety of flight, in short, anything that will kill you.
- **Incorporate error correction mechanisms** — Have the system repeat, using either voice synthesis or through a visual display, what it interprets, and allow the pilot to accept or reject this recognition. Allow the system to be able to recognize invalid recognition. If the recognizer interprets that it heard the pilot call out an invalid frequency, it should recognize it as invalid and possibly query the pilot to repeat, or prompt the pilot by saying or displaying that the frequency is invalid.
- **Provide feedback of the SR system’s activities** — Allow the user to interact with the SR system. Have the system speak, using voice synthesis, or display what it is doing. This will allow the user to either accept or reject the recognizer interpretation. This may also serve as a way to prompt a user for more data that may have been left out of the utterance. “Tune COM 1 to...” After a delay,

the system might query the user for a frequency: “Please select frequency for COM1.” If the user selects some repeated command, the system may repeat back the command as it is executed: “Tuning COM 1 to ...”

8.2.4.1 Reduced Vocabulary

One way to dramatically increase the accuracy of a SR system is to reduce the number of words in a vocabulary. In addition to the reduction in words, the words should be carefully chosen to weed out words that sound similar.

Use a trigger phrase to gain the attention of the recognizer. The trigger phrase might be as simple as “computer...” followed by some command. In this example, “computer” is the trigger phrase and alerts the recognizer that a command is likely to follow. This can be used with a system that is always on-line and listening.

Speech recognition errors can be reduced using a noise-canceling microphone. The flightdeck is not the quiet, sterile place a lab or a desktop might be. There are any number of noises and chatter that could interfere with the operation of speech recognition. Like humans, a recognizer can have increased difficulty in understanding commands in a noisy environment. In addition to the use of noise-canceling microphones, the use of high-quality omnidirectional microphones will offer further reduction in recognition errors. Using push-to-talk (PTT) microphones will help to reduce the occurrence of insertion errors as well as recognition errors.

8.2.4.2 Grammar

Grammar definition plays an important role in how accurate a SR application may be. It is used to not only define which words are valid to the system, but what the command syntax will be. A grammar notation frequently used in speech recognition is Context Free Grammar (CFG). A sample of a valid command in CFG is

$$\langle \text{start} \rangle = \text{tune}(\text{COM} | \text{NAV}) \text{ radio}$$

This definition would allow valid commands of “tune COM radio,” and “tune NAV radio.” Word order is required, and words cannot be omitted. However, the grammar can be defined to allow for word order and omitted words.

8.3 Recent Applications

Though speech recognition has been applied to various flightdeck applications over the past 20 years, limitations in both hardware and software capability have kept the use of speech recognition from serious contention as a flightdeck tool. Even though there have been several notable applications of speech recognition in the recent past, and there are several current applications of speech recognition in the cockpit of military aircraft, it will likely be several more years before the civilian market will see such applications reach the level of reliability and pilot acceptance to see them commonly available.

In the mid 1990s, NASA performed experiments using speech recognition and voice control on an OV-10A aircraft. The experiment involved 12 pilots. The speech recognizer used for this study was an ITT VRS-1290 speaker-dependent system. The vocabulary used in this study was small, containing 54 words. The SR system was tested using the 12 pilots under three separate conditions: on the ground, 1g conditions, and 3g conditions. There was no significant difference in SR system performance found between the three conditions. The accuracy rates for the SR system under these three test conditions was 97.27% in hangar conditions, 97.72 under 1g conditions, and 97.11% under 3g conditions.³

A recent installation that is now in production is a military fighter, the Eurofighter, Typhoon. This aircraft will be the first production aircraft with voice interaction as a standard OEM configuration with speech recognition modules (SRMs). The speech recognizer is speaker dependent, and sophisticated enough to recognize continuous speech. The supplier of the voice recognition system for this aircraft is Smiths Industries. In addition, the system has received general pilot acceptance. Since the system is speaker

dependent, the pilot must train the speech recognizer to his unique voice patterns prior to its use. This is done at ground-based, personal computer (PC) support stations. The PC is used to create a voice template for a specific pilot. The created voice template is then transferred to the aircraft prior to flight, via a data loader. Specifications for the recognizer include a 250-word vocabulary, a 200-ms response time, continuous speech recognition, and an accuracy rate of 95–98%.²

Another recent application of speech recognition technology is in the Joint Strike Fighter (JSF) being developed by Boeing and BAe Systems. Continuous speech recognition is being integrated into the cockpit. The speech recognition system will provide selected cockpit controls sole operation by using voice commands. The JSF speech recognition system will be used to allow the pilot to avoid the distraction of selected manual tasks while remaining focused on more critical aspects of the mission. The supplier of the speech recognition system for this aircraft is ITT Industries' Voxware (formerly VERBEX) voice recognition system. The Voxware system was chosen for this application due its recognized and previously proven ability to perform in a noisy cockpit environment.¹

8.4 Flightdeck Applications

The use of speech recognition, the enabling technology for voice control, should not be relied on as the sole means of control or entering data and commands. Speech recognition is more correctly defined as an assisted method of control; and should have reversionary controls in place if the operation and performance of the SR system is no longer acceptable. It is not a question of whether voice control will find its way into mainstream aviation cockpits, but a question of when and to what degree. As the technology of SR continues to evolve, care must be exercised so that SR does not become a solution looking for a problem to solve. Not all situations will be good choices for the application of SR. In a high workload atmosphere, such as the flightdeck, the use of SR could be a logical choice for use in many operations, leading to a reduction in workload and heads-down time.

Current speech recognition systems are best assigned to tasks that are not in themselves critical to the safety of flight. In time, this will change as the technology evolves. The thought of allowing the speech recognition system to gain the ability to directly impact flight safety brings to mind an example that occurred at a speech recognition conference several years ago. While a speech recognition interface on a PC was being discussed and demonstrated before an audience, a member of the audience spoke out “format C: return,” or something to that effect. The result was the main drive on the computer was formatted, erasing its contents. Normally an event such as this impacts no one's safety, however, if such unrestricted control were allowed on an aircraft, there would be serious results.

Some likely applications for voice control on the flightdeck are navigation functions; communications functions such as frequency selection, toggling of display modes, checklist functions, etc.

8.4.1 Navigation Functions

For navigation functions, SR could be used as a method of entering waypoints and inputting FMS data. Generally, most tasks requiring the keyboard to be used to enter data into the FMS would make good use of a SR system. This would allow time and labor savings in what is a repetitive and time consuming task. Another advantage of using SR is that the system is able to reduce confusion and guide the user by requesting required data. The use of SR with FMS systems is being evaluated and studied by both military and civilian aviation.

8.4.2 Communication Functions

For communication functions, voice control could be used to tune radio frequencies by calling out that frequency. For example, “Tune COM1 to one one eight point seven.” The SR system would interpret this utterance, and would place the frequency into stand-by. The system may be designed to have the SR system repeat the recognized frequency back through a voice synthesizer to the pilot for confirmation

prior to the frequency being placed into standby. The pilot would then accept the frequency and make it active or reject it. This would be done with a button press to activate the frequency. Another possible method of making a frequency active would be to do this by voice alone. This does bring about some added risk, as the pilot will no longer be physically making the selection. This could be done by a simple, “COM one Accept” to accept the frequency, but leave it in pre-select. Reject the frequency by saying, “COM one Reject,” and to activate the frequency by saying, “COM one activate.”

The use of SR would also allow a pilot to query systems, such as by requesting a current frequency setting; “What is COM one?” The ASR system could then respond with the current active frequency and possibly the pre-select. This response could be by voice or by display. Other possible options would be to have the SR respond to ATC commands by moving the command frequency change to the pre-select automatically. Having done this, the pilot would only have to command “Accept,” “Activate,” or “Reject.” The radio would never on its own, place a frequency from standby to active mode.

With the use of a GPS position-referenced database, a pilot might only have to call out “Tune COM one Phoenix Sky Harbor Approach.” By referencing the current aircraft location to a database, the SR systems could look up the appropriate frequency and place it into pre-select. The system might respond back with, “COM one Phoenix Sky Harbor Approach at one two oh point seven.” The pilot would then be able to accept and activate the frequency without having to know the correct frequency numbers or having to dial the frequency into the radio. Clearly a time-saving operation. Possible drawbacks are out-of-date radio frequencies in the database or no frequency listing. This can be overcome by being able to call out specific frequencies if required. “Tune COM one to one two oh point seven.”

8.4.3 Checklist

The use of speech recognition is almost a natural for checklist operations. The pilot may be able to command the system with “configure for take-off.” This could lead to the system bringing up an appropriate checklist for take-off configuration. The speech system could call out the checklist items as they occur and the pilot, having completed and verified the task, could press a button to accept and move on to the next task. It may be possible to allow a pilot to verbally check-off a task, vs. a button selection; however, that does bring about an opportunity for a recognition error.

Defining Terms

Accuracy: Generally, accuracy refers to the percentage of times that a speech recognizer will correctly recognize a word. This accuracy value is determined by dividing the number of times that the recognizer correctly identifies a word by the number of words input into the SR system.

Continuous speech recognition: The ability of the speech recognition system to accept a continuous, unbroken stream of words and recognize it as a valid phrase.

Discrete word recognition: This refers to the ability of a speech recognizer to recognize a discrete word. The words must be separated by a gap or pause between the previous word and successive words. The pause will typically be 150 ms or longer. The use of such a system is characterized by “choppy” speech to ensure the required break between words.

Grammar: This is a set of syntax rules determining valid commands and vocabulary for the SR system. The grammar will define how words may be ordered and what commands are valid. The grammar definition structure most commonly used is known as “context free grammar” or CFG.

Isolated word recognition: The ability of the SR system to recognize a specific word in a stream of words. Isolated word recognition can be used as a “trigger” to place the SR system into an active standby mode, ready to accept input.

Phonemes: Phonemes are the fundamental parts of speech. The English language is made up from 45 to 50 individual phonemes.

Speaker Dependent: This type of system is dependent upon the speaker for operation. The system will be trained to recognize one person’s speech patterns and acoustical properties. This type of system will have a higher accuracy rate than a speaker-independent system, but is limited to one user.

Speaker Independent: A speaker-independent system will operate regardless of the speaker. This type of system is the most desirable for a general use application, however the accuracy rate and response rate will be lower than the speaker-dependent system.

Speech Synthesis: The use of an artificial means to create speech-like sounds.

Text to Speech: A mechanism or process in which text is transformed into digital audio form and output as “spoken” text. Speech synthesis can be used to allow a system to respond to a user verbally.

Tri-Phones: These are groupings of three phonemes. The sound a phoneme makes can vary depending on the phoneme ahead of it and after it. Speech recognizers use tri-phones to better determine which phoneme has been spoken based upon the sounds preceding and following it.

Verbal Artifacts: These are words or phrases, spoken with the intended command that have no value content to the command. This is sometimes referred to simply as garbage when defining a specific grammar. Grammars may be written to allow for this by disregarding and ignoring these utterances, for example, the pilot utterance, “uhhhhhmmmmmm, select north up mode.” The “uhhhh-hmmmmmm” would be ignored as garbage.

Vocabulary: The vocabulary a speech recognition system is made up of the words or phrases that the system is to recognize. Vocabulary size is generally broken into four sizes; small, with tens of words, medium with a few hundred words, large with a few thousand words, very large with up to 64,000 words, and unlimited. When a vocabulary is defined, it will contain words that are relative, and specific to the application.

References

1. Boeing JSF to feature voice-recognition technology, [On-Line]. Available: www.boeing.com/news/releases/2000/news_release_000222o.htm.
2. The Eurofighter Typhoon Speech Recognition Module, [On-Line]. Available: www.smithsind-aerospace.com/PRODS/CIS/Voice.htm
3. Williamson, David T., Barry, Timothy P., and Liggett, Kristen K., Flight test results of ITT VRS-1290 in NASA OV10A. Pilot-Vehicle Interface Branch (WL/FIGP), WPAFB, OH.

Bibliography

- Anderson, Timothy R., Applications of speech-based control, in *Proc. Alternative Control Technologies: Human Factors Issues*, 14-15 Oct., 1998, Wright-Patterson AFB, OH, (ISBN 92-837-1003-7).
- Anderson, Timothy R., The technology of speech-based control, in *Proc. Alternative Control Technologies: Human Factors Issues*, 14-15 Oct., 1998, Wright-Patterson AFB, OH, (ISBN 92-837-1003-7).
- Bekker, M. M., “A comparison of mouse and speech input control of a text-annotation system,” Faculty of Industrial Design Engineering, Delft University of Technology, Jaffalaan 9, 2628 BX Delft, The Netherlands.
- Boeing, JSF to feature voice-recognition technology, [On-Line]. Available: www.boeing.com/news/releases/2000/news_release_000222o.htm.
- Eurofighter Typhoon Speech Recognition Module, Available: www.smithsind-aerospace.com/PRODS/CIS/Voice.htm.
- Hart, Sandra G., Helicopter human factors, in *Human Factors in Aviation*, Wiener, Earl L. and Nagel, David C., Eds., Academic Press, San Diego, 1988, chap. 18.
- Hopkin, V. David, Air traffic control, in *Human Factors in Aviation*, Wiener, Earl L. and Nagel, David C., Eds., Academic Press, San Diego, 1988, chap. 19.
- Jones, Dylan M., Frankish, Clive R., and Hapeshi, K., Automatic Speech Recognition in Practice, *Behav. Inf. Technol.*, 2, 109–122, 1992.
- Leger, Alain, Synthesis and expected benefits analysis, in *Proc. Alternative Control Technologies: Human Factors Issues*, 14-15 Oct., 1998, Wright-Patterson AFB, OH, (ISBN 92-837-1003-7).
- Rood, G. M., Operational rationale and related issues for alternative control technologies, in *Proc. Alternative Control Technologies: Human Factors Issues*, 14-15 Oct., 1998, Wright-Patterson AFB, OH, (ISBN 92-837-1003-7).

- Rudnick, Alexander I. and Hauptmann, Alexander G., Models for evaluating interaction protocols in speech recognition, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1991.
- Wickens, Christopher D. and Flach, John M., Information processing, in *Human Factors in Aviation*, Wiener, Earl L. and Nagel, David C., Eds., Academic Press, San Diego, 1988, chap. 5.
- Williamson, David T., Barry, Timothy P., and Liggett, Kristen K., Flight test results of ITT VRS-1290 in NASA OV10A. Pilot-Vehicle Interface Branch (WL/FIGP), WPAFB, OH.
- Williges, Robert C., Williges, Beverly H., and Fainter, Robert G., Software interfaces for aviation systems, in *Human Factors in Aviation*, Wiener, Earl L. and Nagel, David C., Eds., Academic Press, San Diego, 1988, chap. 14.

Further Information

There are numerous sources for additional information on speech recognition. A search of the Internet on "speech recognition" will yield many links and information sources. The list will likely contain companies and corporations that deal primarily in speech recognition products. Some of these companies include:

- Analog Devices (800) 262-5643 www.analog.com
- AT&T Adv Speech Products Group (800) 592-8766 www.att.com/aspg
- Brooktrout Technology (617) 449-4100 www.techspk.com
- Dialogic (201) 993-3000 www.dialogic.com
- Dragon Systems (800) 825-5897 www.dragonsys.com
- Entropic Cambridge Research Labs (202) 547-1420 www.entropic.com
- IBM Speech Products (800) 825-5263 www.software.ibm.com/is/voicetype
- Kurzweil Applied Intelligence (617) 883-5151 www.kurzweil.com
- Lernout & Hauspie (617) 238-0960 www.lhs.com
- Nuance Communications (415) 462-8200 www.nuance.com
- Oki Semiconductor (408) 720-1900 www.oki.com
- Philips Speech Processing (516) 921-9310 www.speech.be.philips.com
- PureSpeech (617) 441-0000 www.speech.com
- Sensory (408) 744-1299 www.SensoryInc.com
- Smith Industries (610) 296-5000 www.smithsind-aerospace.com/
- Speech Solutions (800) 773-3247 www.speechsolutions.com
- Texas Instruments (800) 477-8924 x 4500 www.ti.com