Fall 1999

The Impact of the Simulated Digital Data Radio Communication Induced Interference on Voice Radio Communication

Daniela T. Kratchounova
Embry-Riddle Aeronautical University - Daytona Beach

THE IMPACT OF THE SIMULATED DIGITAL DATA RADIO COMMUNICATION INDUCED INTERFERENCE ON VOICE RADIO COMMUNICATION INTELLIGIBILITY

by

Daniela T. Kratchounova

A Thesis

Presented to the Human Factors and Systems Department

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Human Factors

Embry-Riddle Aeronautical University

Daytona Beach, Florida

Fall 1999

THE IMPACT OF THE SIMULATED DIGITAL DATA

RADIO COMMUNICATION INDUCED INTERFERENCE ON VOICE RADIO

COMMUNICATION INTELLIGIBILITY

by

Daniela T. Kratchounova

This thesis was prepared under the direction of the candidate's thesis chair, Dr. John A. Wise, Department of Human Factors and Systems, and has been approved by the members of his thesis committee. It was submitted to the Department of Human Factors and Systems and was accepted in partial fulfillment of the requirements for the degree of

Master of Science of Human Factors & Systems Engineering

THESIS COMMITTEE:

John A. Wise, Ph.D, Chair

Albert Helfrick, Ph.D, Member

Eric Vaden, Member

HFS Graduate Program Coordinator

Department Chair, Department of Human Factors and Systems

ACKNOWLEDGMENTS

I wish to express special thanks to my thesis chairman, Dr. John A. Wise, whose valuable guidance, practical suggestions, help and patience were critical for the successful outcome of this thesis. I would like to thank Dr. Albert Helfrick for the great ideas, support, expertise, unlimited ingenuity, and boundless patience. I would like to thank Eric Vaden for his practical suggestions, expertise, constant encouragement and support. This experiment would not have been possible without Steve Hall's valuable help and advice. I would like to thank Peter S. Pierpont, Jennie Gibbs and all the wonderful people from Department of Engineering Technology for giving me the opportunity to be a part of the team for the last two years. I wish to express my thanks to Monica Frapier for her support and encouragement.

This statement of acknowledgement would be incomplete without a formal expression of appreciation and gratitude to all my friends, here in the USA, back home in Bulgaria, my family, and especially to my wonderful children, Elitza and Kaloyan, for providing me the inspiration to complete the task.

Abstract

Author: Daniela T. Kratchounova

Title: The Impact of the Simulated Digital Data Radio Communication Induced Interference on Voice Radio Communication Intelligibility

Institution: Embry-Riddle Aeronautical University

Degree: Master of Science in Human Factors

Date: 1999

The transition of air/ground communications, from its present analog structure to a new digital communications architecture, will span a 10-year time frame. Testing to date indicates that the potential for interference from the digital system to the analog system would be critical, especially for General Aviation (GA) aircraft. This type of interference can be described as short, random bursts of noise capable of completely obliterating parts of the voice communication. The subsequent degrading effects on voice radio communications could jeopardize flight safety. The masking of parts of important information would result in distracted attention, debilitated cognitive performance, high level of annoyance, and stress. The goal of the proposed experiment was to examine the degree to which such noise impacts voice radio communications intelligibility. A classic, well established psycho-acoustic method of measuring intelligibility was used. It was anticipated that the digital data radio communication interference would influence voice communication intelligibility and the ratio between the length of

the burst and the length of the affected consonant would be critical for voice communication intelligibility. Based upon the results of the statistical analysis that was exactly what happened. Those words treated with higher LR were more difficult to identify that those treated with lower LR or with no interference at all.

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

INTRODUCTION

Pilots have many tasks in a very complex environment in flight and on the ground. Part of the complexity of this environment can be attributed to the airborne radio communications equipment. Because the results of miscommunication can be catastrophic, voice communication is such a vital aspect of the aviation system. Advanced aircraft communications systems also involve digital data radio communication techniques as a promising technology that may bring many benefits to the aviation community.

The Federal Aviation Administration (FAA), in conjunction with airlines and avionics manufacturers, is undertaking the development and deployment of a set of datalink services to enhance air/ground communications in the en route environment. In this globally coordinated effort, digital data communication between controllers and flight crews is being introduced. Out of necessity there will be a transition period when digital data and conventional analog voice radios must coexist.

Problem statement

Digital Very High Frequency (VHF) data radio transmissions interfere with VHF AM receiving equipment used for voice communications. The potential for this type of interference is critical because of the inability to separate antennas enough, especially for GA aircraft. Pilots perceive digital data transmission interference as short noise-like bursts capable of distorting parts of the incoming speech.

Review of Related Literature

Communications

In the most fundamental sense, communication involves implicitly the transmission of information from one point to another through a succession of processes as:

The generation of a thought pattern or image in the mind of an originator;

-The description of that image by a set of aural or visual symbols;

-The encoding of these symbols in a form that is suitable for transmission over a physical medium;

-The transmission of the encoded symbols to the desired destination;

-The decoding and reproduction of the original symbols; and

-The recreation of the original thought pattern or image, with a determinable degradation in quality caused by imperfections in the system, in the mind of the recipient.

Irrespective of the form of the considered communication process, there are three basic elements to every communication system: transmitter, channel, and receiver as shown in Figure 1. The transmitter is located at one point in space, the receiver is located at some other point separate from the transmitter, and the channel is the physical medium that connects them. The purpose of the transmitter is to transform the message signal produced by the source of information into a form suitable for transmission over the channel.

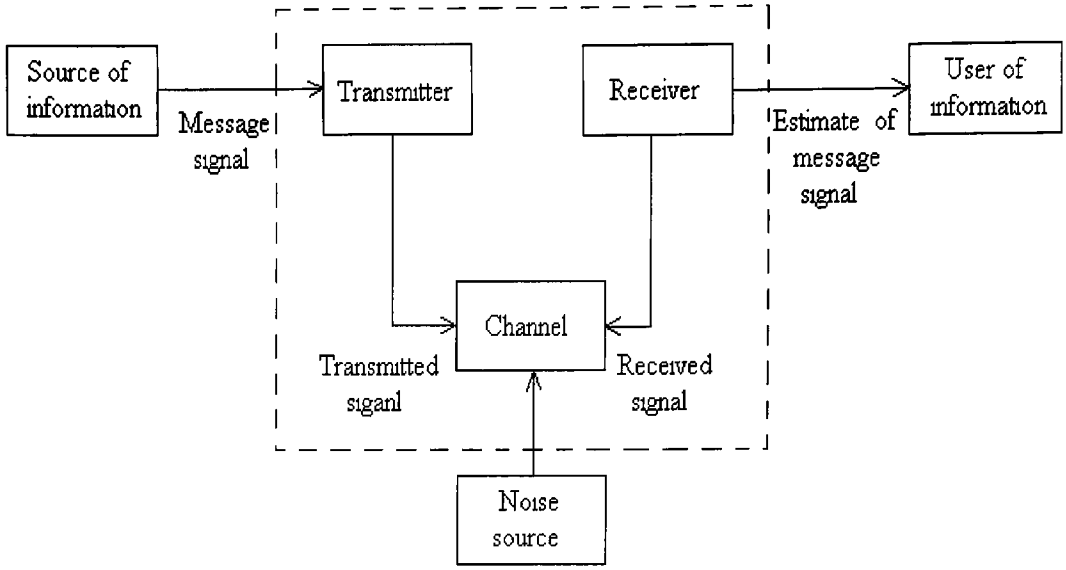


Figure 1 Elements of a communication system

As the transmitted signal propagates along the channel, it is distorted due to channel imperfections, even more, noise and interfering signals are added to the channel output. As a result of that the received signal is a corrupted version of the transmitted signal. The receiver role is to reconstruct a recognizable form of the original signal and to deliver it to the user. The signal processing role of the receiver is the reverse of that of the transmitter.

In the context of a communication system, a signal of primary interest is the message signal delivered by a source of information. This signal is also referred to as baseband signal, with the term "baseband" being used to designate the band of frequencies representing the message signal. Baseband signals can be of an analog or digital type. Another signal of primary interest is the transmitted signal, the characterization of which is determined by the type of channel used in the communication system. In this context, we speak of baseband or passband transmission. In baseband transmission, the band of transmission frequencies supported by the channel closely matches the band of transmission frequencies occupied by the message signal. In passband transmission, the transmission band of the channel is centered at a frequency much higher then the highest frequency component of the message signal. In that case, the transmitted signal is called passband signal, the generation of which is accomplished in the transmitter using a process known as modulation. The process of modulation involves varying some parameter of a carrier wave in accordance with the message signal. The receiver recreates the original message from a degraded version of the transmitted signal after propagation through the channel. This recreation is accomplished by using a process known as demodulation, which is the reverse of the modulation process used in the transmitter. However, because of the unavoidable presence of noise and distortion in the received signal, the receiver cannot recreate the original message exactly.

The communication process is of a probabilistic nature. To appreciate this fundamental property, we merely have to recognize that if the receiver of a communication system were to know the composition of a message exactly, there would be no point in having the system transmit that message. A major source of "uncertainty" in the operation of a communication system is noise that originates naturally at the front end of the receiver. The two most common forms of noise are thermal noise, produced by the random motion of electrons in conducting media, and shot noise, produced by random fluctuation of current flow in electronic devices. The received signal can be further corrupted by interference due to undesirable sources or external effects. The result is that the received signal is randomlike in appearance, in other words, we cannot predict the exact value of the received signal. The received signal, however, can be described in terms of its statistical properties. Another factor contributing to uncertainty in the communication process is the source of information itself. The characterization of a random process and the evaluation of the effect produced by passing it through a linear system are essential to a thorough understanding of the operation of a communication system.

Digital Data Radio Communication

During the past several years, the electronic communications industry has undergone some remarkable technological changes. Traditional electronic communication systems that use conventional analog modulation techniques are gradually replaced with more modern digital communication systems. Digital communication systems offer several outstanding advantages over traditional analog systems: ease of processing, ease of multiplexing, and noise immunity. Information is propagated through an electronic communication system in the form of symbols that can be analog, such as the human voice, or digital, such as binary-coded numbers, alpha/numeric codes or database information.

The term digital communication covers broad area of communication techniques, including digital radio. Digital radio is the transmittal of digitally modulated analog carriers between two or more points in a communication system.

Data communication is the process of transferring digital information between two or more points. Information that has been processed, organized, and stored is called data. At both the source and destination, data are in digital form. Advanced aircraft communication systems also involve Digital Data Radio Communication, for example VHF Data Link (VDL), as a promising technology that supports navigation and surveillance applications (RTCA DO-224) and may bring many benefits to the aviation community.

Digital Data Radio Communication Interference with Voice Radio Communication

The transition of air/ground communications, from its present analog structure to a new digital communications architecture, will span a ten-year time frame. Testing to date indicates that the potential for interference from the digital system to the analog system would be critical, especially for general aviation aircraft (P., Huisman, at al.,1998.) This type of interference can be described as short, random bursts of noise capable of completely obliterating parts of the voice communication. The subsequent degrading effects on voice communications could jeopardize flight safety. The obliterating of parts of important information would result in distracted attention, debilitated cognitive performance, high level of annoyance, and stress.

In the case of digital radio data communication induced interference with voice communication, the pilot perceives the VHF data transmission that interferes with the VHF AM (Amplitude Modulation) receiving equipment used for Voice Communication as a short burst of noise. The source of this interference is, the so called co-site interference from on-board or same site digital radio data transmitters, operating simultaneously with the reception of VHF signals virtually anywhere within the 118-137 MHz band.

The Federal Communications Commission (FCC) defines harmful interference (CFR 47) as interference which endangers the functioning of a radionavigation service or other safety services or seriously degrades, obstructs or repeatedly interrupts a radio communication service operating in accordance with the Radio Regulations.

Sound

Language in the form of speech is the primary method for human communication. The speech communication process involves the transfer of information from a speaker to a listener, which takes place in three stages: production, propagation and perception. An intended message in the speaker's mind is represented by a speech signal that consists of sounds generated inside the speaker's mouth and governed by the rules of language. The sound waves propagate through the air, reaching the listener's ears. The listener into a received message interprets the incoming sounds.

The sounds we hear are created from a form of mechanical energy. They are patterns of successive pressure disturbances occurring in a medium (ordinarily the sounds we hear are transmitted in the air). Sound may be described in terms of three variables:

1. Amplitude
2. Frequency
3. Time pattern

Sound pressure is the amplitude or measure of the difference between atmospheric pressure (with no sound present) and the total pressure (with sound present).

The unit of sound level is the decibel (dB); this is a logarithmic scale with a particular reference level of sound. A logarithmic scale is used because the range of sound intensities is so great that it is convenient to compress the scale to encompass all the sounds that need to be measured. Considering the definition of a decibel:

$$dB = 20 \log (p_1 / p_0)$$

where p_1 refers to the air pressure amplitude of the sound under consideration and p_0 refers to a standard reference level of air pressure. This reference level is typically set at 20 microPascals (μ Pa), the microPascal being a measure of pressure. To signify when amplitudes are being given relative to this particular sound pressure level (SPL), we denote such amplitudes as "dB_{SPL}". We also can describe amplitudes relative to each other in terms of dB, omitting the subscript SPL.

The human ear has an extremely wide range of response to sound amplitude. Sharply painful sound pressure is 10 million times greater than the least audible sound. In decibels, this 10 million to 1 ratio is 140 dB.

The corresponding auditory sensation of pressure amplitude is loudness. High-amplitude sound waves express large pressure changes and create the experience of loud sounds, whereas low-amplitude waves reflect small pressure changes are heard as soft sounds.

The rate at which a sound source vibrates, or makes the air vibrate, determines frequency. The unit of time is one second and the term "Hertz" is used to designate the number of cycles per second. The human ear and that of most animals has a wide range of response. Humans can identify sounds with frequencies from about 16 Hz to 20,000 Hz. Because pure tones are relatively rare in real-life situations, most sounds consist instead of a complex mixture of many frequencies.

Frequency refers to a physical property of the sound wave. The auditory sensation – the psychological attribute of frequency – is called pitch, which refers to how high or low a sound is experienced by a listener.

Sounds produced by the human voice possess extremely complex cycles of pressure variations and are produced by the interaction of many different waves of varying frequencies. In general, a complex sound-emitting source vibrates simultaneously at a number of frequencies. The lowest frequency is called fundamental frequency. It determines the pitch of a complex sound. All additional vibrations, whose frequencies are multiples of the fundamental frequency, are called harmonics (overtones).

The characteristics sensory experience corresponding to sound's complexity is called timbre. Timbre refers to a sound's distinctive tonal quality produced by the number and intensity of the harmonics.

The temporal nature of sound may be described in terms of its pattern of time and level: continuity, fluctuation, impulsiveness, intermittence. Continuous sounds are those produced for relatively long periods at a constant level Intermittent sounds are those which are produced for short periods. Impulse sounds are produced in an extremely short span of time. Fluctuating sounds vary in level over time.

Speech Production

The human vocal and auditory organs form a very complex communication system. The human apparatus concerned with speech production and perception is complex and uses many elements-the lungs, mouth, nose, ears and their controlling muscles and the brain. Speech sounds are produced when breath is blow from the lungs and causes either a vibration of the vocal cords (for vowels) or turbulence at some point of constriction in the vocal tract (for consonants). The sounds are affected by the shape of the vocal tract, which influences the produced harmonics. The vibration of the vocal cords in voicing produces sound at a sequence of frequencies, the natural harmonics, each of which is a multiple of the fundamental frequency.

For voiced sounds like the vowels the vocal tract acts as a resonant cavity. This resonance produces large peaks in the resulting speech spectrum. These peaks are known as formants. The formants of the speech contain almost all the information contained in the signal. Nasals are produced when the oral cavity is blocked and the velum is lowered to couple the nasal cavity with the vocal tract. They are also voiced in nature but the coupling of the oral and nasal cavities introduces anti-resonances or nulls instead of resonances so the formants disappear. Unvoiced hiss-like sounds like "s,""f," and "sh" are generated by constricting the vocal tract close to the lips.

The formant frequencies for each of the vowel sounds are quite distinct but for each vowel sound generally have similar values regardless of who is speaking. The fundamental frequency will vary depending on the person speaking, mood and emphasis, but it is the magnitude and relationship of the formant frequencies, which make each voiced sound recognizable.

The sound power in speech is carried by the vowels, which average from 30 to 300 milliseconds in duration. Intelligibility is imparted chiefly by the consonants, which average from 10 to 100 milliseconds in duration and may be as much as 27 dB lower in amplitude than the vowels. The strength of the speech signal varies as a whole, and the strength of individual frequency ranges varies with respect to the others as the formants change.

The Acoustic Theory of Speech Production: the Source-Filter Model

Acoustic speech output in humans is commonly considered to result from a combination of a source of sound energy (e.g., the larynx) modulated by a transfer function (filter) determined by the shape of the supralaryngeal vocal tract. This combination results in a shaped spectrum with broadband energy peaks. This model is often referred to as the "source-filter theory of speech production" and stems from the Müller (1848) experiments in which a functional theory of speech was tested by blowing air through larynges excised from human cadavers. Müller noticed that the sound that came directly from the larynx differed from the sounds of human speech. Speech-like quality could be achieved only when he placed over the vibrating cords a tube whose length was roughly equal to the length of the airways that normally intervene between the larynx and a person's lips. In this model the source of acoustic energy is at the larynx – the supralaryngeal vocal tract serves as a variable acoustic filter whose shape determines the phonetic quality of the sound. When the larynx serves as a source of sound energy, voiced sounds are produced by a repeating sequence of events. First, the vocal cords are brought together, temporarily blocking the flow of air from the lungs and leading to increased subglottal pressure. When the subglottal pressure becomes greater than the resistance offered by the vocal folds, they open again. The folds then close rapidly due to a combination of factors, including their elasticity, and laryngeal muscle tension. If the process is maintained by a steady supply of pressurized air, the vocal cords will continue to open and close in a quasiperiodic fashion. As they open and close, puffs of air flow through the glottal opening. The frequency of these pulses determines the fundamental frequency of the laryngeal source and contributes to the perceived pitch of the produced sound. The rate at which the vocal folds open and close during phonation can be varied in a number of ways and is determined by the tension of the laryngeal muscles and the air pressure generated by the lungs. The shape of the spectrum is determined by details of the opening and closing movement, and is partly independent of fundamental frequency. In normal speech fundamental frequency changes constantly, providing linguistic information, as in the different intonation patterns associated with questions and statements, and information about emotional content, such as differences in speaker mood. In addition, the fundamental frequency pattern determines naturalness of utterance production.

The supralaryngeal vocal tract, consisting of both the oral and nasal airways, can serve as a time-varying acoustic filter that suppresses the passage of sound energy at certain frequencies while allowing its passage at other frequencies. Formants are those frequencies at which local energy maxima are sustained by the supralaryngeal vocal tract and are determined, in part, by the overall shape, length and volume of the vocal tract. The detailed shape of the filter (transfer) function is determined by the entire vocal tract serving as an acoustically resonant system combined with losses including those due to radiation at the lips.

Language

Human language is an complicated, symbolic, rule-governed, and creative communication system. As with any complex ability, language comprises a number of systems working together. Language is symbolic. It consists of linguistic units, sounds that form words and other meaningful units that symbolize or stand for the referent of the word. Language is shared by all speakers of a language culture. Different languages have different sounds called phonemes (the smallest unit of the speech sounds of a language that enables one expression to be distinguished from another.) Phonemes by themselves are meaningless and in isolation may not even be pronounceable, but with other phonemes they form syllables and words. The ways in which phonemes can be combined in any given language constitutes the study of phonology. The way the sounds are put together, identifying meaningful units of language is an aspect known as morphology. Morphemes are the smallest meaningful units of language. To determine the role that each word plays, the syntax, or structure of each sentence needs to be identified. A syntactically correct sentence does not by itself make for a good communication. The sentence must mean something to the listener. Semantics is the branch of linguistics devoted to the study of meaning. There also needs to be some flow in the communication for it to work. Listeners must pay attention and make certain assumptions, and speakers must craft their contribution in ways that will make the listener's job feasible. This aspect of language is called pragmatics. All of the encountered different linguistic rules make up the grammar of the language, and taken together, define the way a language works. Grammar refers to ways of speaking that form intelligible phrases recognizable as examples of language that a native speaker of the language might produce.

Speech Reception

The sense of hearing provides a unique source of information and many forms of social communication. Little is known about how exactly the brain decodes the acoustic information it receives. However quite a lot is known about the receiver it uses to detect these signals, the ear.

Auditory function is the process by which the stimulation of receptors in the inner ear is followed by transduction of the mechanical stimulus into neural energy, which upon reaching the brain gives us the sensation of hearing.

The human ear on the basis of their short time spectra and how these spectra evolve with time distinguishes different speech sounds. The effective bandwidth of speech is approximately 7 kHz. The human ear consists of three main sections, the outer, the middle and the inner ears. The outer ear consists of the ear lobe (pinna) and the external auditory canal. The function of the ear lobe is to channel sounds into the ear and aid in the localization of sounds. The external auditory canal channels the sound into the middle ear. The canal is approximately 2.7 cm in length and is closed at one end by the eardrum. It can be viewed as an acoustic tube that resonates at 3055 Hz.

The eardrum is a hard membrane, approximately 0.1 mm thick, which is flexible at the edge (like the diaphragm of a loudspeaker). When a sound wave strikes this membrane it vibrates. This vibration is then transferred to the three bone structure in the middle ear and from there to the inner ear. These bones act as a transformer and match the acoustic impedance of the inner ear with that of air. Muscles attached to these bones suppress the vibration if it is too violent and so protect the inner ear. This protection only works for sounds below 2 kHz and it does not work for impulsive sounds. The Eustachian tube connects the middle ear to the vocal tract and removes any static pressure difference between the middle ear and the outer ear. If a significant pressure difference is detected then the Eustacian tube opens and the difference is removed.

The inner ear is composed of the semicircular canals, the cochlea, and auditory nerve terminations. The function of the semicircular canals is to control balance. The cochlea is fluid-filled and helical in shape (it resembles the shell of a snail). Inside the cochlea there is a hair-lined membrane called the basilar membrane. This membrane converts the mechanical signal into a neural signal. Different frequencies excite different portions of this membrane allowing a frequency analysis of the signal to be carried out.

Like any receiver there is a limit to the sensitivity of the ear. If sounds are too weak they will not be detected. This is known as the threshold of audibility. This threshold varies with frequency and it can be increased at any given frequency by the presence of a large signal at a nearby lower frequency. This phenomenon is called masking and it is described in more detail in the next paragraph. Since hearing is largely a matter of stimulus reception over time, we would expect time to influence the perception of sound. Recognizable tonal quality requires some minimal duration of the sound. If a tone of an audible frequency and intensity is presented for only a few milliseconds, it will lose its tonal character and will either be inaudible or be heard as a click. According to Gulick et al. (1989), the length of time a given frequency must last in order to produce the perception of a stable and recognizable pitch is 250 msec. Loudness is also affected by the duration of the sound. As the duration of a sound becomes progressively briefer than 200msec, intensity must be increased to maintain a constant level of loudness.

Masking is a general term that covers a very wide variety of situations when the intrusion of unwanted sounds inevitably interferes with the speech signal. Technically, masking is defined as the rise in the threshold of one tone due to the presence of a second tone.

One relationship between the strength of the speech signal and the masking sound is called the signal-to-noise ratio expressed in decibels. Ideally, the S/N ratio is greater than 0dB, indicating that the speech is louder than the noise. The most uniformly effective mask is broadband noise. Although, narrow-band noise is less effective at masking speech than broadband noise, the degree of masking varies with frequency.

Speech perception is based on the interaction of an enormous number of complex psychological factors. It requires us to make very fine discriminations between sounds. A spoken word consists of a short pattern of sounds lasting less than a second. The perception of speech persists even when the sounds comprising words change greatly. Words retain their identity and are perceived accurately under many distorting conditions. Even when most physical characteristics of speech sounds have been changed to some degree, the sounds may still be intelligible. A complex cognitive integrative mechanism is required to perceive speech, but we appear to come specially equipped to perceive speech in very efficient ways. We continue to perceive intelligible speech even when the flow of speech is reduced or distorted.

One form of speech degradation involves filtering out whole ranges of frequencies from the speech signal. In a classic study by French and Steinberg (1947), subjects heard a series of recorded words from which whole bands of frequencies either above or below 1900 Hz were completely removed. When frequencies above 1900 Hz were filtered out, about 70% of the words were still intelligible. The same result was obtained with the elimination of all frequencies below 1900 Hz. This means that as much of the total intelligibility is carried by frequencies below 1900 Hz as is carried by frequencies above 1900 Hz. Thus, the critical stimulus information for speech perception is not confined to any frequency range. This procedure is known as frequency cutoff.

The perception of continuous speech does not depend only on the fixed acoustic stimulation that occurs at any given moment. It also depends on anticipations and expectations of what the stimulation should be based on the cognitive framework created by preceding and following speech sounds, in other words on the context. People sometimes "hear" phonemes that are not there (Warren,1970.) In the 1970 study, Warren presented participants with recording of a sentence in which a 120 ms portion had been replaced with a coughing sound. Nineteen out of twenty listeners demonstrated phoneme restoration effect, so called because listeners apparently "restore" the missing phonemes predicted by other linguistic information during the course of perception. The context directs the listener's perception of a sound – typically without the listener even being aware of this influence. The phoneme restoration effect demonstrates that people effortlessly "fill in" the gaps created in a stream of speech.

The role of context may also reduce inaccuracies in the perception of fluent conversational speech. This occurs especially when the acoustic input is ambiguous or the same articulation has readily accessible alternative perceptions (Garrett, 1982.) That is when the same acoustic input, received phonetically in the same manner, may be perceived differently, depending on the context in which that occurs.

A result similar to the perceptual filling-in-effect of context occurs when seconds of speech are eliminated by periodically turning the speech on and off or by systematically blanking out sections of speech flow by turning a masking noise on and off. This form of speech disruption is called speech blanking (Miller, 1947.)

When conversation was blanked out for 50% of the time (at a rate of nine blinks per second), speech intelligibility was altered, but only about 15% of the words were lost.

In general, the listener's challenge is to parse speech sounds into meaningful units of language - a complicated task. Gaps in the sound do not necessarily correspond to word or syllable breaks. Speech sounds also are not discrete events: rather, they merge and overlap in time, and the articulation of a given phoneme differs in different contexts and with different speakers. In fact, the precise ways in which the ear-brain mechanism decodes speech remain something of a mystery.

Speech perception is remarkably resistant to degradation. Speech remains reasonably intelligible even with background noise and even when chunks of linguistic elements, including phonemes and whole frequency bands, are far from ideal.

Speech Intelligibility

In the process of evaluating a speech communication system, well-established criteria and standards are needed. The major criterion for evaluating a speech communication system is intelligibility. Intelligibility is the degree to which speech can be understood. With specific reference to speech/voice radio communication system testing, intelligibility denotes the extent to which trained listeners can identify words or phrases that are spoken by trained talkers and transmitted to the listeners via the communication system. Diminished intelligibility is associated with a loss of information that is coded in a number of highly interactive elements, and many factors influence it. In general, intelligibility is highest for sentences, less for isolated words, and lowest for nonsense syllables. Therefore, intelligibility is very dependent on context and expectations. As any other communication system, a speech communication involves implicitly the three basic elements: transmitter (speaker), channel (the transmission system, noise environment), and receiver (the listener).

The intelligibility of speech depends in part on the character of the speaker's voice. Bilger, Hanley, and Steer (1955) found those "superior" speakers (in contrast with less intelligible speakers)

Had a longer speech syllable duration,

Spoke with greater intensity,

Utilized more of the total time with speech sounds (less pauses), and

Varied their speech more in terms of fundamental vocal frequencies.

The differences in intelligibility of speakers are due to the structure of their articulators and the speech habits people have learned

Characteristics that affect the intelligibility of the message include phonemes used, the words, and the context. Certain speech sounds are more easily confused than others. Intelligibility is greater for familiar words than for unfamiliar words. Intelligibility is higher for long words than for short words, because with short words there is less opportunity to pick up pieces of the word. In the aviation world we use word-spelling alphabet in place of single letters in the transmission of alphanumeric information. Intelligibility is higher for sentences than for isolated words because the context supplies information, as already discussed.

Speech transmission systems (telephones, radios) can produce various forms of distortion, interference and noise. Noise, be it external in the environment or internal to the transmission system, is the bane of speech intelligibility. Any unwanted introduced signal or sound in a communication system or speaking environment is called noise. The sources of noise are many, and can be both acoustical and electronic. The effects of noise are described in more details in one of the next paragraphs.

The listener is the last link in the communication chain. For receiving speech messages under noise conditions, the listener should have normal hearing, should be trained in the types of communications to be received, should be reasonably able to withstand stress of the situation, and should be able to concentrate on one of several conflicting stimuli.

Subjective speech intelligibility measurements

Speech intelligibility is an important measure of the effectiveness or adequacy of a communication system or of the ability of people to communicate in noisy environments. Speech intelligibility can be measured directly using a measurement process where a number of English-fluent talkers speak standardized word lists through the communication system to trained, English-fluent listeners who indicate what they hear. The word lists are crafted to evaluate specific aspects of speech transmission; the ability of the listeners to identify individual words indicates the quality of the transmission.

Tape recordings of the speakers are used rather than live speech, so that different communications systems can be compared with exactly the same speech material.

Subjective intelligibility measurements use human beings, rather than electronic test instruments, to assess speech communication systems. First proposed in 1910 and refined with the introduction of the telephone and the advent of electronic communication systems in World War II, such tests are still considered to be the most accurate and reliable measures of intelligibility.

Materials used in intelligibility testing must employ a representative sample of the critical speech sounds under all the conditions of speech communication under test. The materials must have demonstrated validity and reliability and must permit analysis of performance errors. Criteria that also are important, are economy of testing and the potential of automation to simplify administration of the tests and analysis of the results. Subjective intelligibility tests are based on various types of speech material evaluated in speaker-listener communication. Speech elements frequently used for testing include phonemes, words (digits, alphabet, meaningful words, or nonsense CVC-words (Consonant-Vowel-Consonant), sentences, and sometimes a free conversation. A sentence that is used to present test words in subjective intelligibility tests is called carrier sentence. The test word is spoken without emphasis, and the sentence is the same for each test word. The percentage correctly recalled items of the set presented gives the score. The recall procedure can be based on a given limited set of responses or on an open response design in which all possible alternatives are allowed as a response. A limited response set is used with the so-called rhyme tests. These types of tests are easy to administer and do not require extensive training by the listeners in order to arrive at stable scores. Rhyme tests may, depending on the design, disregard specific phoneme confusions. Open response tests, especially those, which make use of nonsense words, require an extensive training of the listeners. Additionally to the word and phoneme scores, possible confusions between phonemes are obtained. Redundant speech material (sentences, rhyme tests) suffers from ceiling effects (100% score at poor-to-fair conditions) while tests based on nonsense words may discriminate between good and excellent conditions.

The proposed experiment deals most directly with the American National Standards Institute's approved procedure (ANSI S3.2-1989, "Method for Measuring the Intelligibility of Speech Over Communication Systems").

The subjective measurement process uses trained, English-fluent talkers speaking standardized word lists through the communication system to trained, English-fluent listeners. The word lists are crafted to evaluate specific aspects of speech transmission; the ability of the listeners to identify individual words or word pairs indicates the quality of the transmission.

A number of specialized word lists are in common use for testing various aspects of speech communication. The ANSI standard specifies three:

1. Modified Rhyme Test (MRT)

The MRT is one of the ANSI standards (ANSI S3.2-1989) for measuring the intelligibility of speech over communications systems. The test materials consist of six 50 rhyming monosyllable word sets (i.e. pin,sin,tin,fin,din,win) that were selected with half to differ in the initial consonant and the other half in the final consonant. The listener's task is to respond to the stimulus by indicating which of the six rhyming words presented before him was spoken.

2. Diagnostic Rhyme Test (DRT)

The DRT is also one of the ANSI standards (ANSI S3.2-1989) for measuring the intelligibility of speech over communication systems. The test materials consist of 96 rhyming monosyllable word pairs (i.e. veal-feel) that were selected to differ in only their initial consonant. These differences are categorized into six distinctive features and scores in each of these categories provide information on diagnosing system deficiencies. These six scores are averaged together to provide an overall measure of system intelligibility. The listener's task is to respond to the stimulus by indicating which of the two rhyming words presented before him was spoken.

1. Set of twenty Phonetically Balanced Word Lists (PB)

The oldest of the three tests is the set of 20 phonetically balanced (PB) lists, each of which contains 50 monosyllabic English words. The PB materials were developed during World War II and have been used very widely for more than 30 years. The words are presented in a pseudo-open set, and listeners write down their responses. The words are included in the same carrier sentence, "Would you write <word> now" which all speakers use for all words.

In each list of words the frequency or occurrences of types of speech sounds are proportional to those in everyday life. Pre-recorded test material can be used. At a minimum, each participant is given three PB or MRT word lists. A set of percentage scores is calculated showing the number of times words were identified correctly by each listener. Taking an average of these can produce a single overall score. If either the DRT or MRT is used, the results are adjusted mathematically to account for guessing (no adjustment is required for the PB test). All three tests generally provide the same rank orders and magnitude of differences among systems or devices when used with a large number of communication systems under a wide range of conditions. Results obtained with the three tests have been demonstrated to be highly correlated with results obtained with vocabularies which are representative of various types of operational communications. All three tests have demonstrated sensitivity to degraded speech, and they have been used successfully with human talkers and coders, with analog and digital speech, in reverberation and noise. The three tests are highly intercorrelated with each other and with other tests of speech intelligibility. With appropriate control of extraneous factors, all three tests yield highly self-consistent and repeatable results.

Noise and its effects

Intruding noises can interfere with human activities by distracting attention and by making activities more difficult to perform, especially when concentration is needed. Interference from noise can even make some activities (such as communication) virtually impossible.

The effects of noise are seldom catastrophic, and are often only transitory, but adverse effects can be cumulative with prolonged or repeated exposure. There is some evidence that it can affect general health and well being in the same manner as chronic stress (Broadbent, 1971.) Noise can mask important sounds and disrupt communication between individuals in a variety of settings. This process can cause anything from a slight irritation to a serious safety hazard involving an accident or even a fatality because of the failure to hear important piece of information. Numerous accident reports have many "say again" exchanges between pilots and controllers, although neither side reports anything wrong with the radios. Noise can cause fatigue and vocal strain in those who need to communicate in spite of the noise. Interference with speech communication and other sounds is one of the most salient components of noise-induced annoyance. The resulting disruption can constitute anything from an annoyance to a serious safety hazard, depending on the circumstance. Noise can adversely affect task performance in a variety of circumstances (Broatbent, 1979) Even moderate noise levels can increase anxiety, decrease the incidence of helping behavior, and increase the risk of hostile behavior. Noise is considered a nonspecific biological stressor, gaining a response that prepares the body for action response. The physiological mechanism to be responsible for this reaction is the stimulation by noise (via the auditory system) of the brain's reticular activating system (Cohen, 1977). Neural impulses spread from the reticular system to the higher cortex and throughout the central nervous system. Noise can, therefore, influence perceptual, motor, and cognitive behavior, and also trigger glandular, cardiovascular, and gastrointestinal changes by means of the autonomic nervous system. Annoyance can be viewed, as the expression of negative feelings resulting from interference with activities, as well as disruption of one's peace of mind and the enjoyment of one's environment.

Experiment

The goal of the proposed experiment was to examine the degree to which simulated digital data radio transmission induced interference impacts voice radio communications intelligibility.

Hypothesis

The hypothesis is:

Voice radio communication intelligibility decreases as the ratio between the length of the noise-like burst and the length of the affected consonant (LR).

2. METHOD

Participants

Twenty-seven pilots were used as participants. All participation was on a volunteer basis. All participants were English-fluent and held a valid second class medical certificate, which includes a hearing test. Pilots were considered as highly trained listeners because they have received on-the-job training and have extensive experience in voice radio communications.

Instruments

The experiment consisted of one condition with each participant. A set of three prerecorded (Intelligibility and Measurement Test Discs 1 & 2, produced by PROSONUS in association with INTEREC Publishing, 1991) PB word lists (50 words each), were used with each participant, as a minimum requirement by the standard (ANSI S3.2-1989.) The PB test was chosen for this experiment because in each list of words the frequency or occurrences of types of speech sounds are proportional to those in everyday life. PB tests have alos demonstrated sensitivity to degraded speech and have been very successful with human participants. All three PB word lists were edited using SOUND FORGE XP 4.5 sound editor as follows:

- The first list contained no interference,
- The second list included lower LR (.2 < LR < .5), and
- The third list included higher LR (.7 < LR < 1.2)

A sample of white noise 40 msec (ARINC No. 96-CRDA-0090) long, was used to simulate the digital data radio transmissions. One of the consonants of each word in the second and the third word lists was obliterated by the burst as shown in Figure 2. The burst location in each word was assigned over one of the consonants because consonantal sounds are difficult to transmit successfully, and are, thus, more important than vowels to speech intelligibility. The words from the second and the third word lists were divided into two groups of 50 words each:

- words that contain long (lasting 80-200 msec) consonant, and
- words that contain short (lasting 30-80 msec) consonant

In both groups of words, consonants (one per word) were edited using SOUND FORGE XP 4.5 sound editor in a way so that the leading 40 msec of each consonant was cut off and then a 40 msec burst of white noise was pasted in its place. The test sequence of words, each carried by the carrier sentence, were saved as 150 separate files. Then those files were randomized using a table of random numbers. The actual sequence of words used in the study can be found in Appendix A of this paper. The test was computer based and self-paced. All participants wore headsets (Optimus PRO-50MX). They were given paper to write on.

Task

Words were presented one at a time. Participants then wrote down the word they heard. Words were not repeated.

Design

A within-subjects design was used for the study. A within-subject design was chosen because it allows the variation among participants to be removed prior to the assessment of factor effects. The independent variable (IV) was the ratio between the length of the burst and the length of the affected consonant (LR). The dependent variable (DV) was the speech intelligibility or percentage correctly recalled items out of 150.

In this experiment there were three levels of the factor (interference) defining the conditions - no interference, $.2 < LR < .5$, and $.7 < LR < 1.2$. Scores were obtained from each participant for each of the conditions.

Procedures

Participants were given verbal instructions. The participants then became thoroughly familiar with all the perception-discrimination tasks required to recognize the test words under the conditions of the actual test. An equipment familiarization trial of three words (each in a carrier sentence) with no interference was conducted. The participants were thoroughly familiar with the motor demands of the test and were briefed on the specific kinds of degraded communication that were investigated.

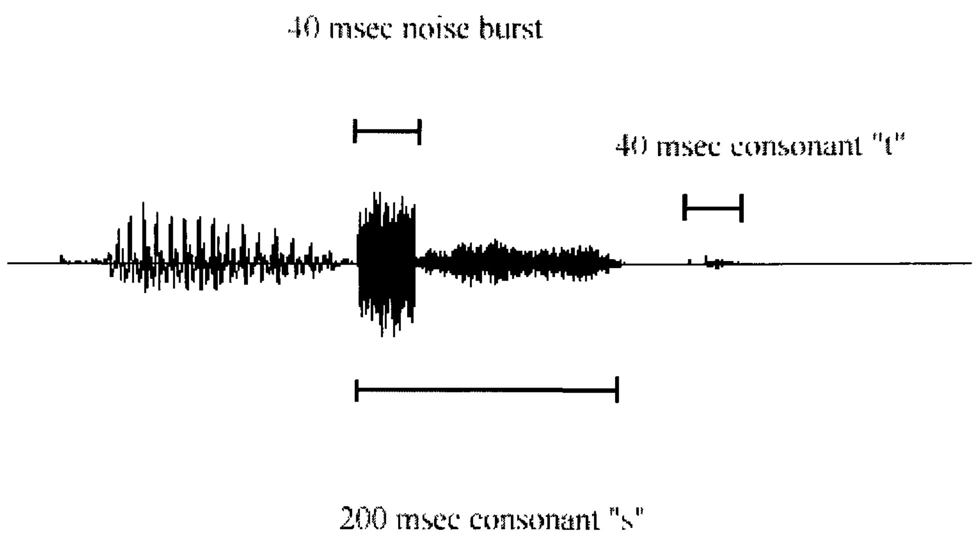


Figure 2 The waveform of the word "FEAST" with the burst of white noise pasted over the leading 40 msec of the consonant "s"

In both familiarization and testing no visual cues were allowed.

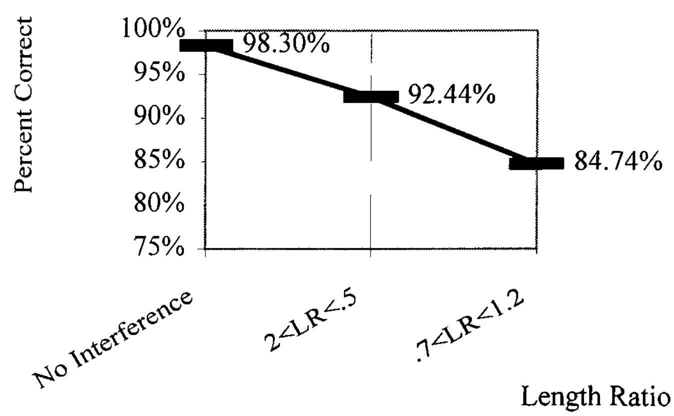
In order to prevent them to get any visual cues the participants were asked not to write anything down during the familiarization nor were they allowed to see printed copies or of the actual test sequence of words (Appendix A).

Results and data analysis

The evaluation of the data from this experiment consisted of a set of percentage scores and a statistical analysis and summary. The generally accepted method for measuring performance on intelligibility tests is to compute the arithmetic mean of percent of words correctly identified by the participants.

The mean test scores for no interference, .2 < LR < .5, and .7 < LR < 1.2 group of words were 98%, 92% and 84% correct, respectively (Figure 3).

Figure 3 Percent correctly identified words as a function of the LR



These means do differ significantly using One-Way ANOVA $F(2,52) = 117.59$. Thus, different Length Ratios significantly influence the percent of words correctly identified by the participants. Specifically those words treated with higher LR were more difficult to identify that those treated with lower LR or with no interference at all. The number of items correct was used as a measure of intelligibility The data was subject to a one-way within-subjects analysis of variance (ANOVA), the results of which are shown in Table1.

Table 1 Summary table of One-Way Within-Subjects Measures ANOVA

Source	Sum of squares	df	Mean square	F
Factor A (interference)	0.24961	2	0.124805	117.5909
Subjects	0.080958	26	0.003114	
Interaction A x S	0.05519	52	0.001061	
Total	0.385758	80		

* $p<.001$, $F_{crit} = 3.23$

From looking at the graphed means (Figure 3) it looks like that when there was no interference introduced, the scores are higher, while the percent of correctly identified items, where different levels of interference were present, was lower. Thus, when the LR was higher the scores were lower.

These apparent differences were examined using a Tukey HSD post hoc procedure, HSD (2,40) = .0215 (Table 2.)

Table 2 Tukey HSD post hocs

Tukey HSD			**No interference**	**.2 < LR < .5**	**.7 < LR < 1.2**
0.021564		**No interference**	0		
		.2 < LR < .5	0.058519	0	
		.7 < LR < 1.2	0.135556	0.077037	0

*p < .05

In fact the percent correctly identified items for the "no interference" group of words was significantly higher than the ".2 < LR < .5" or ".7 < LR < 1.2" groups. However the percent correctly identified words from the ".2 < LR < .5" group was significantly higher than the percent correct from the ".7 < LR < 1.2" group.

In summary different levels of interference or no interference at all produce significant effect on the percent correctly identified words.

Discussion and recommendations

Of primary concern for this study were the effects of digital data radio communication induced interference on voice radio communication intelligibility.

It was anticipated that the digital data radio communication interference would influence voice communication intelligibility and the length ratio (LR) would be critical for voice communication intelligibility. Based upon the results of the statistical analysis that was exactly what happened. Those words treated with higher LR were more difficult to identify that those treated with lower LR or with no interference at all.

In general, as the interfering burst obliterates a larger proportion of a consonant the intelligibility decreases.

To investigate the effect of this type of interference specifically the study was conducted in a very conservative environment. No other form of noise, distortion or interference was introduced in the experimental setting. All the context, particularly aviation context, was intentionally eliminated from the experimental study design, because earlier research (Miller,1947; Warren,1970) demonstrated that speech can be severely distorted without becoming unintelligible. The normal stream of speech contains many more discriminative clues than are necessary. If interference removes or confuses some of these clues others carry the message. If the speech stimulus contains many more clues than the listener needs, distortions in the stimulus must be extreme in order to obliterate all traces of the message. Even the context of each word itself was enough for some of the participants in this study to correctly identify most of the words, especially when a longer consonant was obliterated.

There were recognizable shortcomings in that study. First, the conservative environment used in the experiment reduced the ecological validity of the study.

Given what we know about the influence of context on speech intelligibility, research similar to this should be conducted in a higher fidelity environment. It could well be the context of a live cockpit that reduces the negative effect of this type of interference to a negligible level. But, it also could be that the typically very noisy environment in the live cockpit makes the negative effect of the burst even stronger.

It should be noted that scores were negatively influenced not only by the higher LR, but also by the location of the burst. Specifically scores worsen when the burst occurred over the first consonant.

Future research on the impact of the use of English as a second language on voice radio communication intelligibility would bring some insights about the way we should approach the growing diversity of cultural backgrounds in today's "busy" skies.

Interference caused by digital data radio transmission may affect intelligibility differently over the various phases of flight. This may be influenced by the specific terminology/phraseology commonly used by controllers. A better understanding of the intellectual strategies pilots use to cope with the already noisy environment of the live cockpit may be required.

Communication is an essential element of human society, and speech is its most convenient form of expression. Interference with speech can degrade living directly, by disturbing normal social and work-related activities, and indirectly, by causing annoyance and stress. Sometimes the communications, disturbed by noise are of vital importance, especially in the aviation environment.

References

1. Ashcraft, Mark H., *Human Memory and Cognition*, Harper Collins College Publishers

2. Bashford, J.A., Reiner, K.R. & Warren, R.J. (1992) *Increasing the intelligibility of speech through multiple phonemic restorations.* Perception & Psychophysics, 51, 211-217.

3. Bilger, R.,Hanley, T., and Steer, M. (1955) A further investigation of the relationships between voice variables and speech intelligibility in high level noise. Lafayette, IN: Purdue University

4. Bordens, Kenneth S., Abbot, *Bruce B. Research Design and Methods A Process Approach*, Mayfield Publishing Company, Montain View, California

5. Broadbent, D.E. (1979) *Human performance and noise.* In Harris, C.M., ed., Handbook of Noise Control, 2nd ed., NY, McGraw-Hill.

6. Broadbent, D.E. (1983) *Recent advances in understanding performance in noise.* In: Rossi, G., ed. Proceedings of the Fourth International Congress on Noise as a Public Health Problem, Vol. 2

7. Code of Federal Regulations.47, *Telecommunications*, Office of the Federal Register, National Archives and Records Service, General Services Administration; Washington, D.C. Subpart A, Terminology ,p 291

8. Cohen, S. and Theinstein N. (1981) *Nonauditory effects of noise on behavior and health.* Journal of the Acoustical Society of America, 37, 36-70.

9. *Cooperative Research and Development Agreement No. 96-CRDA-0090,* ARINC

10. Dijk, P., Huisman, P., Limbosch, A., Scheidema, P., (1998) *Differential Eight-level Phase Shift Keying Interference to VHF AM-DSB Aviation Communication Receivers.* Final report of the measurements taken at ERAU, Daytona Beach, Florida

11. Federal Aviation Administration (1994) *Airman's Information Manual: Official Guide to Basic Flight Information and ATC Procedures,* Washington, D.C.: U.S. Printing Office

12. French, N.R.,& Steinberg, J.C.(1947) *Factors governing the intelligibility of speech-sounds.* Journal of the Acoustical Society of America, 19, 90-119

13. Gardner, B., *Say Again Please Guide to Radio Communications,* Aviation Supplies & Academics, Inc, Newcastle, Washington

14. Garrett, M.F. Production of speech: *Observations from normal and pathological language use.* In A.W. Willis(Ed.) Normality and pathology in cognitive functions. New York: Academic Press, 1982

15. Gulick, W.L., Gescheider, G.A., & Frisina, R.D. (1989) *Hearing: Physiological acoustics, neural coding, and psychoacoustics,* New York: Oxford University Press11. Miller, G.A. *The masking of speech.* Psychological Bulletin, (1947), 44, 105-129

16. Sanders, M.S.,McCormick E.J., *Human Factors in Engineering and Design,* McGraw-Hill,Inc.

17. Schiffman, H.R., *Sensation and Perception,* John Wiley & Sons, Inc.

18. Shannon, Claude E., Warren Weaver,(1947) *The mathematical theory of communication,* University of Illinois Press

19. Warren, R.M (1970) *Perceptual restoration of missing speech sounds,* Science, 167, 392-393

20. Warren, R.M.,& Warren, R.P (1970) *Auditory illusions and confusions.* Scientific American, 223, 30-36

APPENDIX A

List of the one hundred and fifty PB test words used in the study

range	check	jug
scout	folk	pants
mote	air	south
pulse	class	dose
him	feast	rich
such	not	box
path	chop	fraud
grove	shaft	dig
bad	pig	grade
leave	fling	slip
plush	sit	wheat
raid	rub	fake
why	phase	pile
fuss	gnaw	hunt
hive	flush	fig
cast	cook	ride
wire	sin	thrash
drop	fame	barb
mange	rat	dill
death	then	mud
nest	hid	bead

take	siege	cape
far	neck	there
crime	sniff	vow
dwarf	but	pent
gasp	ache	cut
fort	off	please
cleanse	rate	trip
knit	act	smile
woo	strife	pounce
toil	woe	fern
turf	jam	rouse
aim	dike	sag
dish	wharf	bask
rash	nine	comes
is	rag	gun
sped	pan	cane
sledge	by	muck
no	bar	heap
stag	lush	whiff
roar	end	toe
quiz	deed	size
use	plod	wedge

am	bald	coast
creed	are	who
sob	nook	though
deck	crave	pest
dope	clove	rise
crash	oak	shout
ford	law	hurl