# Measurement of Speech Transmission Suitability

Homeland Security

DHS-TR-PSC-07-01

**Department of Homeland Security**

**Public Safety Communications**

**Technical Report**

Homeland Security

Science and Technology

# Command, Control and Interoperability: Communication, Interoperability and Compatibility

## Defining the Problem

Emergency responders—police officers, fire personnel, emergency medical services-need to share vital voice and data information across disciplines and jurisdictions to successfully respond to day-to-day incidents and large-scale emergencies. Unfortunately, for decades, inadequate and unreliable communications have compromised their ability to perform mission-critical duties. Responders often have difficulty communicating when adjacent agencies are assigned to different radio bands, use incompatible proprietary systems and infrastructure, and lack adequate standard operating procedures and effective multi-jurisdictional, multi-disciplinary governance structures.

## OIC Background

The Department of Homeland Security (DHS) established the Office for Interoperability and Compatibility (OIC) in 2004 to strengthen and integrate interoperability and compatibility efforts to improve local, tribal, state, and Federal emergency response and preparedness. Managed by the Science and Technology Directorate, and housed within the Communication, Interoperability and Compatibility thrust area, OIC helps coordinate interoperability efforts across DHS. OIC programs and initiatives address critical interoperability and compatibility issues. Priority areas include communications, equipment, and training.

## OIC Programs

OIC programs, which are the majority of Communication, Interoperability and Compatibility programs, address both voice and data interoperability. OIC is creating the capacity for increased levels of interoperability by developing tools, best practices, technologies, and methodologies that emergency response agencies can immediately put into effect. OIC is also improving incident response and recovery by developing tools, technologies, and messaging standards that help emergency responders manage incidents and exchange information in real time.

## Practitioner-Driven Approach

OIC is committed to working in partnership with local, tribal, state, and Federal officials to serve critical emergency response needs. OIC’s programs are unique in that they advocate a “bottom-up” approach. OIC’s practitioner-driven governance structure gains from the valuable input of the emergency response community and from local, tribal, state, and Federal policy makers and leaders.

## Long-Term Goals

- Strengthen and integrate homeland security activities related to research and development, testing and evaluation, standards, technical assistance, training, and grant funding.
- Provide a single resource for information about and assistance with voice and data interoperability and compatibility issues.
- Reduce unnecessary duplication in emergency response programs and unneeded spending on interoperability issues.
- Identify and promote interoperability and compatibility best practices in the emergency response arena.

This page intentionally left blank.

Public Safety Communications Technical Report

# Measurement of Speech Transmission Suitability

**DHS-TR-PSC-07-01**
**November 2007**

**Reported for:** **The Office for Interoperability and Compatibility by NIST/OLES**

Homeland Security

This page intentionally left blank.

# Publication Notice

## Disclaimer

The U.S. Department of Homeland Security's Science and Technology Directorate serves as the primary research and development arm of the Department, using our Nation's scientific and technological resources to provide local, state, and Federal officials with the technology and capabilities to protect the homeland. Managed by the Science and Technology Directorate, the Office for Interoperability and Compatibility (OIC) is assisting in the coordination of interoperability efforts across the Nation.

Certain commercial equipment, materials, and software are sometimes identified to specify technical aspects of the reported procedures and results. In no case does such identification imply recommendations or endorsement by the U.S. Government, its departments, or its agencies; nor does it imply that the equipment, materials, and software identified are the best available for this purpose.

## Contact Information

Please send comments or questions to: S&T-C2I@dhs.gov

This page intentionally left blank.

# Contents

This page intentionally left blank.

# Abstract

This report describes a laboratory study on the suitability of speech transmission systems. Specifically, public safety first responders listened to and evaluated a large number of recordings of speech transmission systems. The packet loss requirements given in Section 2 of the Public Safety Statement of Requirements (PS SoR) Volume II [1] are based on the results of this laboratory study.

The study is a "human factors" study in the sense that human subjects experienced controlled stimuli and responded to the stimuli. The stimuli are combinations of sound recordings; thus the study could be more specifically described as a "human listening" study. This is a well-developed field in telecommunications research, and such studies are sometimes described as "listening tests," "subjective speech quality tests," or just "subjective tests." Where appropriate, the study follows the conventions and standards used in this field (e.g., P.800 [2] and P.830 [3]). But it also departs from these conventions in order to use messages, background sounds, listeners, and listening protocols that are specifically relevant to public safety operations.

**Key words**: background sound, human listening tests, PLC (packet loss concealment) algorithm, recorded message transcription, recorded message concatenation, record talker messages, signal-to-noise ratio (SNR), speech coding, speech packet loss, speech passband, speech transmission system, subjective speech quality tests

# 1 Introduction

In broad terms, the goal of the study is to determine how suitable or unsuitable different speech transmission systems would be for "mission-critical communications" in public safety operations. A key ingredient in meeting this goal is creating a well-controlled, yet realistic, simulation of the environment in which mission-critical speech communications occur. The twin goals of realism and control are generally at odds with each other, so the design of this study, like all studies in this field, required informed, calculated trade-offs or compromises.

Once the environment was simulated, first responders were recruited to enter the environment, listen to recordings, and provide their opinions on what was suitable and what was not suitable. The following describes the various steps of the study in some detail, and provides rationales and results.

# 2 Methods

## 2.1 Message Transcription

To help create a realistic environment, it was desired that recordings carried realistic messages related to public safety operations. Toward that end, Internet-based "scanners" were used to monitor actual public safety radio traffic in several different U.S. locations. The traffic was transcribed, filtered for appropriateness and to remove duplication, and edited as necessary to protect anonymity. The transcribed messages were then classified by length, resulting in four categories of messages, listed as follows with the number of messages for each type:

- Tiny (one word)—10 messages
- Short (several words)—64 messages

- Medium (typically a full sentence)—200 messages
- Long (typically more than one sentence)—200 messages

All messages were in the English language. Two example messages from each category follow:

- Tiny:

  “Negative.”

  “Copy.”

- Short:

  “They will not extradite.”

  “Sure go ahead.”

- Medium:

  “Do you need another unit there for traffic control?”

  “White males in their mid 50’s, one in a green baseball cap.”

- Long:

  “Meet the complainant at 145 Riverside. The citizen thinks he found a body in the woods near by. I also have fire responding.”

  “The suspect took a bag out of her vehicle, ran down the street, and put it in another vehicle.”

## 2.2 Message Recording

Two female and three male adults were selected to read the messages during recording sessions. Each person, also called a talker, read 40 messages from each category, or the maximum number of messages available. In the tiny category, each talker read all 10 messages. In the short category, some of the talkers read messages 1 through 40, and others read messages 25 through 64. In the medium and long categories, each message was read by one of five single talkers (5 talkers × 40 messages/talker = 200 messages).

The recording sessions were conducted in a sound-isolated chamber, with ambient noise level below 30 dBA (decibels A-weighted sound pressure level). A single, studio-quality microphone with a cardioid pickup pattern was used. This microphone includes an integral analog-to-digital (A/D) converter, and the digital output (48,000 samples per second, 24 bits per sample) was connected to a digital audio interface hosted by a personal computer. A sound editing software tool controlled the recording process. To preserve the most fundamental and pristine speech signal, no equalization was used in the recording process. The original digital recordings were then segmented into a sequence of shorter recordings that included just the desired messages, and excluded any mistakes in reading, or extraneous noises.

## 2.3 Addition of Transmit Location Background Sound

The recorded messages are of studio quality, and have virtually no background sound. However, background sound in actual public safety operations is often, if not always, present. Further, background sound types and levels can very greatly between locations. The use of messages with virtually no background sound would be unrealistic, yet including the effect of background sound at transmit locations as a factor would have been well beyond the scope of the study. As a compromise, background sound was added to the messages at a single low level, resulting in a 25 dB SNR (signal-to-noise ratio). This background sound was recorded using studio quality equipment in three different field locations: on a street corner, in a passenger car, and in an office.

The 25 dB SNR was selected empirically. It is high enough to be clearly perceptible, yet it is low enough that it does not interfere with judgments of the speech transmission systems of interest. Future studies might be conducted to characterize the effects of higher background sound levels, and specific noise types at transmit locations.

## 2.4 Message Concatenation

Recorded messages were concatenated to form 182 different meta-messages. A meta-message may contain 11, 12, 13, 14, or 15 messages. Each meta-message was at least 30 seconds long, and the average length was 32 seconds.

The messages in each meta-message were selected using a constrained randomization procedure. This procedure produced a good mixture of talkers and message lengths without resorting to a repetitive pattern.

All five talkers were heard once before any talker was heard a second time. No talker was heard twice in immediate succession. The messages within a meta-message were not connected logically, rather they produced the effect of scanning across multiple public safety communications channels. Following are two example meta-messages:

- Example Meta-Message 1:

  "552, criminal trespasses. Ladder 1, move up to Station 32. Standby Tom 33, they already have someone on the way. Sure, go ahead. 426 copy, a missing person. William 2 en route. Alright. Do you want me to call and have county get out of the area? Thanks. Transporting one adult male to West Precinct. Goodnight 1428. This is the Ladder 4, false alarm, going green and back in service. He should be getting a call right now. 073 Henry, I'm back. Engine 32 aid response."

- Example Meta-Message 2:

  "He also said yesterday it looked like there were two pit bulls inside with the suspect. Affirmative. Ladder 10, code red. Disregard. Engine 21 has arrived. A 15-year-old female missing from Riverside Lane. Disregard. If I take him down will they take him? Disregard. Can you see that person, and what house he went into? This is Engine 6, one truck for manpower please. Affirmed. Negative. They have an urgent call on the hill, units are asking for assistance. Please switch to East."

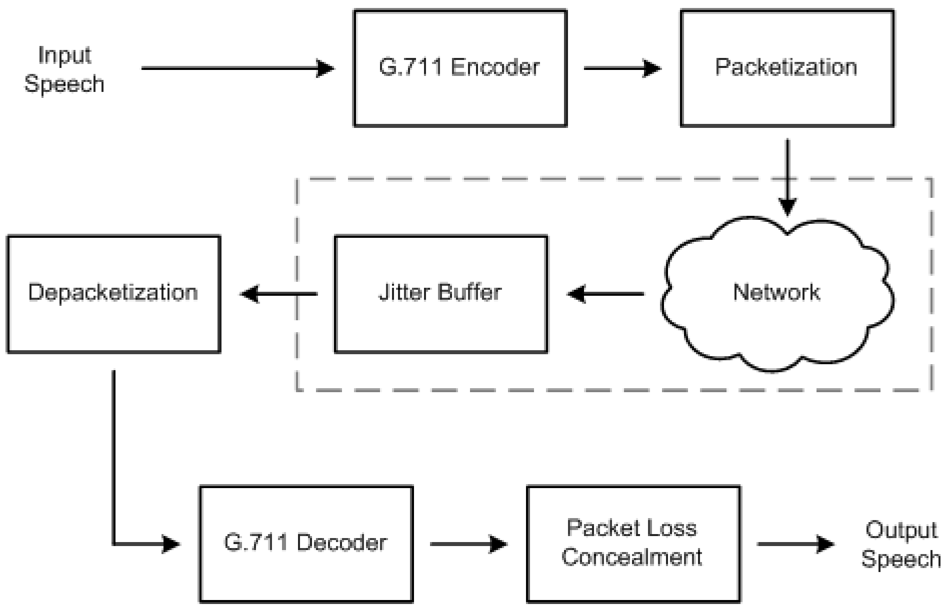
## 2.5 Speech Transmission Systems

The study focuses on G.711 speech coding accompanied by the packet loss concealment (PLC) algorithm specified in G.711 Appendix I. The study uses μ-law G.711, since that is the option used throughout North America. Note that G.711 is typically used to deliver a nominal speech passband that extends to 3400 Hertz (Hz). Potential future studies may investigate any possible benefit in transmitting a passband that extends beyond 3400 Hz (e.g., 7 kilo-Hertz (kHz) wideband speech coding, or 15 kHz audio coding).

The speech coding and data transmission processes simulated in the laboratory study are shown in Figure 1. As shown in this figure, incoming speech signals are encoded and the resulting data is placed into packets which are passed to the network. Packets that emerge from the network are processed to extract the G.711 data stream, which is placed into a jitter buffer to accommodate delay variation in the network.

Following G.711 decoding, the PLC algorithm attempts to hide the effects of lost channel data, and generates the final output speech signal.

Figure 1: Simulated Speech Coding and Data Transmission Processes




The simulation of these processes treats the network and the jitter buffer together as a single black box that can be parameterized (for at least tens of seconds) by a pair of packet loss parameters. These parameters describe fundamental properties and thus they provide a basic yet relevant model.

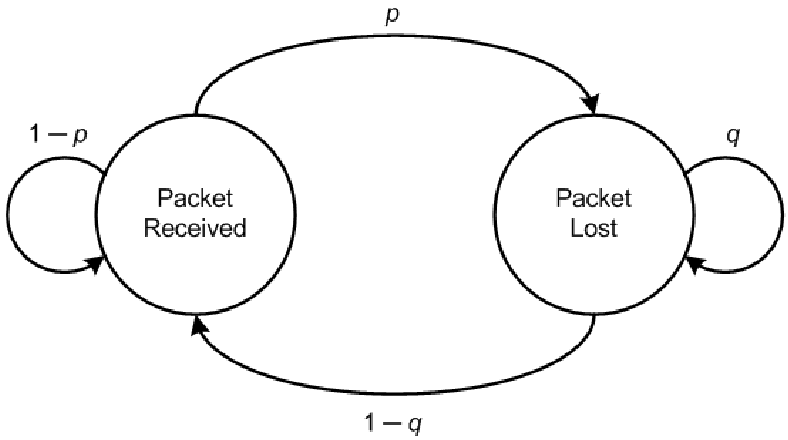
The two packet loss parameters are packet loss ratio and packet loss correlation. When packet loss correlation is zero, the packet loss process is random. As packet loss correlation is increased, the loss of packets becomes more bursty, and it becomes more likely that multiple packets will be lost in succession. In practice, packetized data networks can exhibit random or bursty packet loss patterns.

When the data network is supporting speech transmission, random packet losses generally are less offensive than bursty packet losses. When packet losses are random, it is more likely that isolated packets (not consecutive packets) will be lost. PLC algorithms can hide much of the effect of an isolated lost packet. As packet losses become more bursty, it becomes more likely that multiple consecutive packets will be lost. PLC algorithms loose effectiveness when asked to conceal extended periods of missing data, and they generally stop working altogether after 60 to 90 ms.

In this study, packet loss is modeled by a two-state Markov channel model [4] as shown in Figure 2. Packet loss ratio $(0 < \mu < 1)$ and packet loss correlation $(0 \le \gamma < 1)$ determine the transition probabilities in the Markov model according to $p = (1-\gamma) \times \mu$, and $q = (1-\gamma) \times \mu + \gamma$. Conversely, we have $\mu = p/(1+p-q)$, and $\gamma = q-p$.

When $\gamma = 0$, we have $q = p = \mu$. That is, the unconditional probability of loss $(\mu)$ and both conditional probabilities of loss ($p$ and $q$) are identical, as they must be for random (independent) losses. Note also that the mean length of a loss is $1/(1-q)$ packets, and the standard deviation of a loss is $\sqrt{q}/(1-q)$ packets.

Figure 2: Two-State Markov Channel Model Used in Laboratory




Future studies may incorporate more detailed network models, perhaps specifically tuned to the hybrid network environment (containing both wired and wireless links) that is likely to be used in future systems.

This study uses packet sizes 10 and 40 ms, meaning that G.711 data corresponding to either 10 or 40 ms of speech signal is assigned to a single packet. In practical systems, packet headers add overhead data to each packet, so data efficiency (ratio of payload data per packet to total data per packet) can be increased by using larger packets. On the other hand, larger packets induce larger packetization delay and cause worse speech impairments when a single packet is lost. The choice of a packet size involves balancing these two effects. Based on choices used in current packetized speech transmission systems, 10 and 40 ms packet sizes may approximately represent the endpoints of a viable range.

The meta-messages were passed through software implementations of various speech transmission systems of interest. The majority of these systems were G.711 speech coding, followed by packetization, packet loss, decoding, and PLC, reflected conceptually in Figure 1. The actual implementation followed these steps:

1. Used software to perform bandpass filtering (nominally 200-3600 Hz) of the speech consistent with G.712. [5] This filtering is the standard preparation for G.711 speech coding.
2. Used software to perform G.711-compliant speech encoding.
3. Used software to implement the Markov packet loss model shown in Figure 2, and accordingly deleted blocks of channel data consistent with one of two packet sizes: 10 ms or 40 ms.
4. Used software to perform G.711-compliant speech decoding.
5. Used software to perform G.711 Appendix I-compliant packet loss concealment.

In Table 1, combinations of packet loss ratios and packet loss correlation values used in the study are indicated by "X."

Table 1: Laboratory Packet Loss Ratios and Packet Loss Correlation Values

| | Packet Loss Ratio | | | | | | |
|---|---|---|---|---|---|---|---|
| Packet Loss Correlation | 2 Percent | 5 Percent | 10 Percent | 20 Percent | 30 Percent | 40 Percent | 50 Percent |
| 0.0 | X | X | X | X | X | X | X |
| 0.1 | X | X | X | X | X | | |
| 0.2 | X | X | X | X | X | | |
| 0.3 | X | X | X | X | X | | |
| 0.4 | X | X | X | X | | | |
| 0.5 | X | X | X | X | | | |
| 0.6 | X | X | X | | | | |
| 0.7 | X | X | X | | | | |
| 0.8 | X | X | | | | | |
| 0.9 | X | | | | | | |

The Markov model software uses a pseudo-random number generator to drive the state transitions. Thus the finite-length loss patterns produced have approximately, but not exactly, the desired packet loss ratio. To tightly control this factor, the actual packet loss ratio of each loss pattern was checked. A loss pattern was retained and used only when the measured packet loss ratio was within 5 percent of the target packet loss ratio. For example, when the target packet loss ratio is 0.1000 (commonly denoted as 10 percent), the actual packet ratio is guaranteed to be between 0.095 and 0.105 (commonly denoted as 9.5 percent and 10.5 percent).

Table 1 shows 39 different combinations of packet loss ratios and packet loss correlation values. When these 39 cases are crossed with 10 ms and 40 ms packet sizes, the result is 78 speech transmission systems. Another system arises from the case of no packet loss, which is unaffected by packet loss correlation and packet size.

Nine additional "speech transmission systems" were simulated in software for a total of 88 systems in this study. These nine systems were included for their potential use in future calibration processes.

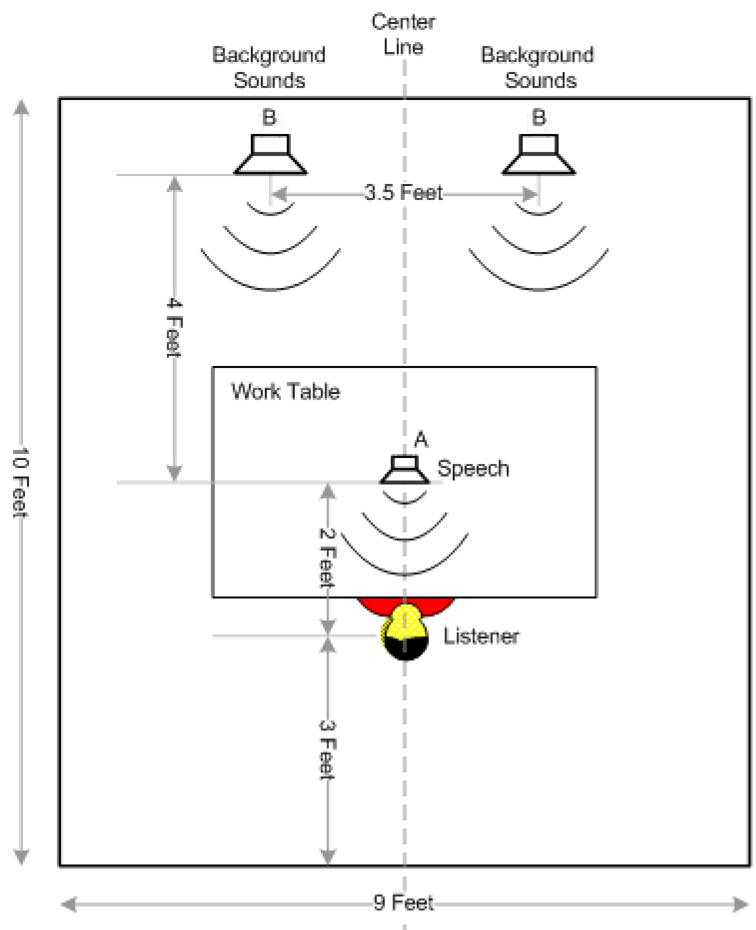
In all cases, proper care was taken to present the proper sample rate (8000 samples per second) and active speech levels (26 dB below the point at which amplitude clipping occurs) to the software implementations. The result of these steps is a second set of meta-messages—the processed meta-messages. Each of these processed meta-messages represents the output of a specific speech transmission system that a public safety first responder would hear at a receiving location. For each system under consideration, eight different processed meta-messages, containing a total of about four minutes of messages, were produced.

## 2.6 Laboratory Conditions

Public safety first responders evaluated the processed meta-messages in acoustically controlled laboratory conditions. We attained the necessary acoustic control through the use of sound-isolated chambers, located inside of an already quiet laboratory space. The inside dimensions of each chamber, effectively a room within a room, were about 9-feet wide, 10-feet long, and 7-feet high. The background noise level inside of these chambers was below 30 dBA.

Figure 3 illustrates the layout of the sound-isolated chambers (top view). Processed meta-messages are played through speaker marked A. Background sound is played through speakers marked B.

Figure 3: Laboratory Layout




The processed meta-messages were played through a small, studio-quality loudspeaker, equipped with a volume control, located on a table. In each evaluation, a single first responder was seated at this table, and was encouraged to adjust the volume level at any time. Through this arrangement, the processed meta-messages could be reproduced at nominal levels up to about 80 dBA at the seating location. The default starting position of the volume control resulted in a nominal level of about 65 dBA at the seating location. The majority of the first responders did not change this volume control (or at least it was in the default position at the end of the session.) Some first responders did leave the volume control above the default position by 4 to 7 dB; no first responder left the volume control in the extreme upper position.

While the chambers isolated first responders from virtually all uncontrolled background noises, such a quiet environment is certainly not typical for public safety operations. In actual public safety operations, background sound types and levels can vary greatly between locations, and may even be high enough to make communications difficult. The effect of high levels of background sound at listening locations is an operational factor that is outside the scope of the present study of equipment factors. On the other hand, any evaluation of equipment factors should certainly be made in an environment that is at least somewhat realistic. Thus to better represent the acoustic environments in which first responders would actually listen to messages, we introduced additional prerecorded background sound into each chamber in a controlled way.

The two background sound levels chosen reflect this compromise: they are high enough to be clearly audible, yet low enough to have only a minor effect on the votes submitted by the first responders. This prevents the background sound from obscuring the desired information regarding equipment factors. The higher level was nominally 60 dBA measured at the first responder head location, and the majority of the level readings for this setting fell in the 55 to 65 dBA range. Preliminary trials with informal self-reporting revealed that this level does not create significant annoyance, fatigue, or distraction. The lower of the two sound levels was 15 dB lower. A significant step in level was desired, yet it was also desirable for the background sound to still be clearly audible. The lower level was nominally 45 dBA and the majority of the sound-pressure level (SPL) readings for this setting fell into the 40 to 50 dBA range.

The background sounds were prerecorded on location using studio-quality equipment. They included street corner sounds, a fire truck at idle, and a distant siren. These recordings were mixed and processed to attain a monophonic signal with the desired level characteristics, and this signal was then transferred to digital audio tape (DAT) for storage. While first responders were in the chamber, the monophonic DAT was played through a pair of studio monitor speakers (identical signal fed to each speaker) marked B as shown in Figure 3.

## 2.7 Evaluation by First Responders

Thirty-five public safety first responders were recruited to participate in the laboratory listening evaluations. The first responders came from across the country. Various local jurisdictions were represented a total of 29 times, state jurisdictions 6 times, and Federal jurisdictions 2 times. Public safety first responders represented include firefighters, law enforcement officers, and Emergency Medical Services (EMS), with 18, 13, and 9 first responders, respectively. (Some first responders represented more than one jurisdiction or discipline.) Three of the first responders had less than 10 years experience, 11 had between 10 and 20 years experience, 14 had between 20 and 30 years experience, and 7 had more than 30 years of experience. Thirty-four of the first responders were males and one was female. Roughly 25 percent were in their thirties, about 50 percent were in their forties, and 25 percent were in their fifties. Thirty of the first responders appeared to be of European heritage, four of Hispanic heritage, and one of African-American heritage.

First responders participated one at a time. This allows each to proceed at his or her desired pace, and prevents first responders from exerting any intentional or unintentional influence on each other. A laboratory study administrator read identical instructions to each first responder, the key portion of which follows:

> Thank you for taking time to participate in this listening experiment. The experiment involves no risk or discomfort, and you are free to end your participation at any time with no penalty. This experiment is one small part of a major process that is being used to design future public safety communications systems.

> In this experiment you will hear a recorded sequence of messages and you are asked to decide whether or not the speech quality in the sequence of messages is suitable for mission-critical communications, given the job that you do. Once you have decided, you will press the appropriate button on the screen in front of you. Each sequence of messages is about 30 seconds in length, but you can make your decision at any time.
>
> If you press the "replay" button, the exact same recording will start again, and it will have the exact same distortions in the exact same places. (Note that this is different than asking a coworker to repeat him or herself.) You can press the "replay" button at any time.
>
> After you have indicated your decision on a recording, the next recording will start immediately.
>
> You are welcome to adjust the volume of the recordings to any level at any time using the knob on the speaker in front of you. Other speakers in this room will play background noise. Several different levels of background noise will be used, and you will not be able to control the volume of the background noise. However, if you start to feel any discomfort from the background noise, please let me know, and I will adjust the level.
>
> Please note that there are no right or wrong answers, we are seeking your opinion.
>
> We want to know what speech quality you personally would consider suitable for mission-critical communications in the job that you do. We ask that you base your judgments on any distortions, noises, or other impairments caused by the communications system. We ask that you not include the pleasantness and diction of the voices, the content of the messages, nor the quality or content of the background noise.

No first responder asked to have the background sound level adjusted, so the same nominal background sound levels defined in Section 2.6 were used for all 35 first responders.

Specialized software aided the actual listening process. This software initiates the playback of a processed meta-message, and waits for the first responder to respond. The first responder used a Personal Digital Assistant (PDA) with a wireless network connection. The following question was prominent on the screen: "Is the speech quality suitable for mission-critical communications?" Also prominently on the display were two response buttons marked "yes" and "no." First responders were allowed to vote on each processed meta-message at any time after it began playing, and the playback could be restarted at any time with a "replay" button. Once a vote was collected, the playback of the next processed meta-message would begin.

First responders did not know what speech transmission systems they were hearing, and a different random order was used in each listening session. First responders did receive text indicating their progress through each session (e.g., "Finished with 10 of 88 trials.") Each first responder first completed a training session that exposed the first responder to a wide range of speech transmission systems, and allowed the first responder to become familiar with the test procedure and equipment. Votes from these training sessions were discarded.

After the training session, each first responder participated in four sessions, each containing 88 processed meta-messages, and thus provided a total of 352 votes. Breaks were offered between each of the sessions so that subjects could refresh themselves if they wished to. The 35 first responders moved at different paces through the 352 recorded processed meta-messages. The total time spent listening ranged from 18 to 107 minutes, with an average total listening time of 58 minutes, which is about 14.5 minutes per session, or 10 seconds per processed meta-message.

Sessions 1 and 2 used the 45 dBA nominal background sound level, while Sessions 3 and 4 (and the training session) used the 60 dBA nominal background sound level. Sessions 3 and 4 used the same processed meta-messages as Sessions 1 and 2 respectively, but they were played in different random orders. All first responders heard each system once in each session. To maximize the number of processed meta-messages heard with each system, four different versions of each session were used, each containing different sets of processed meta-messages. Since Sessions 3 and 4 contained the same processed meta-messages as Sessions 1 and 2 respectively, a total of eight different processed meta-messages (four versions per session × two distinct sessions) were used to evaluate each system. On average, eight meta-messages contain 104 messages.

# 3 Analysis of Votes

In this study, a total of 12,320 votes were collected (88 processed meta-messages per session × 4 sessions per first responder × 35 first responders). Of these votes, 7,452 (60.5 percent) were "yes" votes indicating a good overall balance (suitable versus not suitable) of the systems included in the study. The requirements given in Section 2 of the PS SoR Volume II are based on a statistical analysis of the 12,320 votes. More specifically, the 12,320 votes break into 140 votes (35 first responders × 4 votes per first responder) for each system considered. These 140 votes per system can include variation due to the use of eight different processed meta-messages, the use of 35 different first responders, and the use of two different levels of controlled background sound.

The analysis exercised here views these 140 votes to be samples of the pool of all possible votes that could be cast for that system by the entire body of public safety first responders in this country, hearing all possible messages. Since each vote can be "yes" or "no," Bernoulli trials and the underlying binomial distribution provide a model for this voting process. If the samples (the votes collected in this study) are representative with respect to parameters that affect the votes of the larger pool, then we can use the 140 votes to find an estimate of the votes in that larger pool. Specifically, we can find a maximum likelihood estimate of the underlying parameter $p$ (interpreted as the probability of success, or a "yes" vote), in the binomial distribution. Note that this is different from the variable $p$ used in Section 2.5. Further, we can calculate an interval $[p_{low}, p_{high}]$ that is 95 percent certain to contain the true value of $p$ in the larger pool [6].

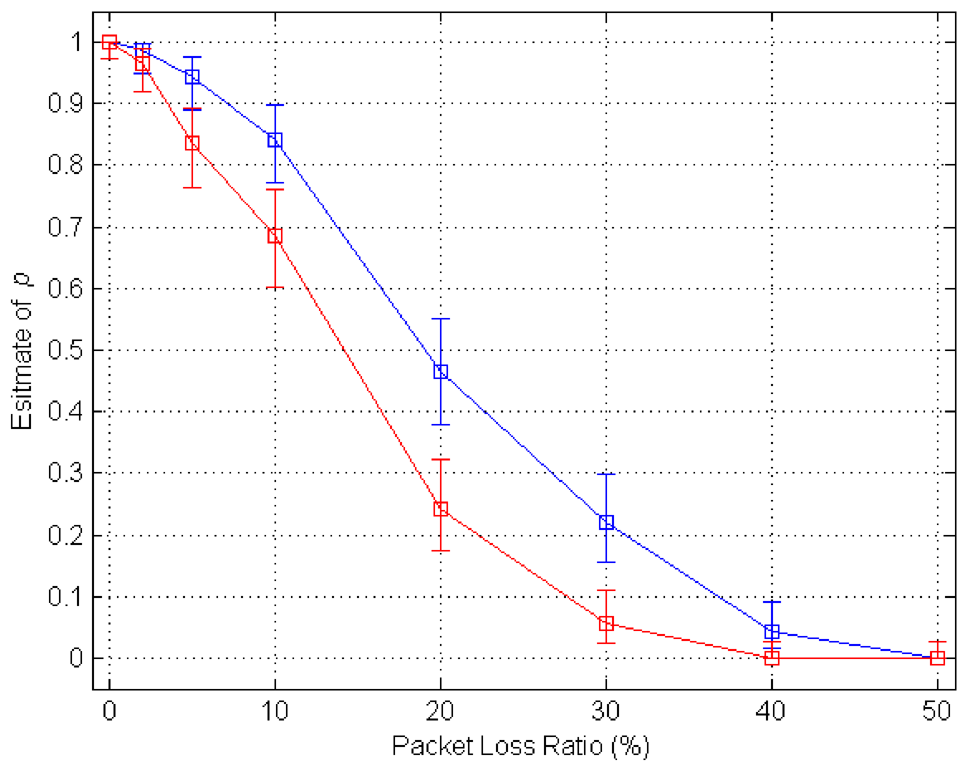
First we consider the effect of the two different background sound levels used during the first responder evaluations. When we compare statistical variations between sessions, we find that the variations between sessions with different background sound levels are no greater than the variations between sessions with the same background sound level. This analysis concludes that changing receive location background sound level does not have a statistically significant effect in this particular experiment. This is a helpful result. It suggests that the background sound levels used are not influencing the results significantly, yet they certainly increase the realism of the study.

Based on the background sound level result, the remaining analyses aggregate votes from both background sound levels, thus corresponding to a nominal background sound level between 45 and 60 dBA. Compared to many operating environments, this would be a conservative background sound level, and in light of the results above, we would expect the resulting requirements to be conservative as well. The remaining analyses use the 140 votes collected for each system to produce estimates on the true value of $p$ in the larger pool of all first responders and all messages.

For each system, we compute a range $[p_{low}, p_{high}]$ that is 95 percent certain to contain the true value of $p$ in the larger pool, if the sample is representative in all relevant aspects. From this calculation, subject to the constraints and assumptions detailed above and in Section 2 of the PS SoR Volume II [1], we then conclude that “it is expected that at least $p_{low}$×100% of the public safety first responders will find the resulting speech transmission to be suitable for mission-critical communications.” Section 2 provides results of this type, organized according to three different thresholds.

Figure 4 provides an example of the results attained. It shows the estimated values of $p$ (fraction of “yes” votes), and 95 percent confidence intervals [$p_{low}$, $p_{high}$], for those estimates for the systems associated with random packet loss $(\gamma = 0)$. The upper (blue) line corresponds to 10 ms packet size, and the lower (red) line corresponds to 40 ms packet size. As expected, estimates of $p$ decrease as packet loss ratio increases. In addition, the 40 ms packet size generally results in lower estimates of $p$ since the loss of 40 ms of speech is harder to conceal than the loss of 10 ms.

Figure 4: Example Results for Random Packet Loss



# 4 References


[1] *Public Safety Statement of Requirements for Communications & Interoperability*, Volume II: Quantitative, Version 1.1, November 2007. Available at: http://www.safecomprogram.gov/SAFECOM/.

[2] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality," Geneva, 1996.

[3] ITU-T Recommendation P.830, "Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs," Geneva, 1996.

[4] H. Sanneck and N. Le, "Speech Property-Based FEC for Internet Telephony Applications," *Proc. SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000*, San Jose, CA, January 2000.

[5] ITU-T Recommendation G.712, "Transmission Performance Characteristics of Pulse Code Modulation Channels," 2001.

[6] N. Johnson, S. Kotz, and A. Kemp, *Univariate Discrete Distributions, Second Edition*. New York: Wiley, 1992.