BongioviAcousticLabs

Improving Speech Intelligibility Using Advanced Audio Processing Techniques

Introduction

Telemedicine is one of the fastest growing areas in medicine. It allows matching a provider with a patient or an expert consultant regardless of location. Telemedicine systems have advanced greatly in the last few years, with much-improved video quality and reliability. However, audio quality is still problematic for both verbal communication between the participants and for internet-enabled stethoscopic sounds.

- Telemedicine is being conducted in “hostile” audio environments. Background noise can exceed 70 dB SPL.
- Clinical spaces may be made of materials with poor acoustic properties (e.g. tiled walls).
- The audio hardware (microphones and speakers) may be of low grade.
- The communications (internet, satellite, telephone) bandwidth may be low so audio is degraded due to data compression.

These challenges combine to reduce the level of speech intelligibility.

Purpose

The following is the procedure and results of a test to analyze the effects of MDPS processing when applied to a communications system. The MDPS algorithm was calibrated to provide maximum voice clarity on an HMDX P420 portable speaker in a variety of environments including:

- Reverberant spaces.
- High-noise simulated by pink noise and crowd noise.
- Listening distance beyond 2 meters.

Several industry-recognized speech intelligibility tests were performed to compare the performance of the system with and without MDPS.

About MDPS Processing

The MDPS digital audio processing algorithm is based on patented methods used to deliver full spectrum audio in a high noise environment. It incorporates a unique method of dynamic range control and high-quality equalization. A standard audio dynamic range controller senses the amplitude of incoming audio then applies a given amount of attenuation based on a ratio and other parameters. However, the MDPS dynamic range controller's sensitivity is based on frequency. Specifically, the frequencies at which the human ear is most sensitive. This allows the MDPS dynamic range controller to have a much wider operating range without adverse effects compared to other methods. The resulting headroom created by the dynamic range controller allows the MDPS equalizers to significantly optimize the output signal for any desired application.

Quantifying Speech Intelligibility

Speech intelligibility testing, on a fundamental level, quantifies the ability of an average person to understand human speech within a given system or environment. Similar tests may be used for intercoms, hearing aids, public address systems, and radio communications. These tests seek to provide common benchmarks for a communication system and provide useful information for improving the system if necessary.

Methods

In this study, care was taken to simulate real-world impediments to good speech intelligibility. In order to keep distances and volume levels at a manageable level, the HMDX P420 portable speaker was chosen. It is a full range powered speaker that can run on 4 AA batteries or AC power adaptor. The frequency response is similar to many desktop and ceiling mounted speakers used for voice communications.

These tests were used to empirically evaluate the system:

- ANSI Modified Rhyme Test
    - *Polycom RealPresence Desktop*[1] software was used between two remote locations via an internet connection.
    - Standard webcams and their built-in microphones were used at each endpoint.

[1] http://www.polycom.com/products-services/hd-telepresence-video-conferencing/realpresence-desktop/realpresence-desktop.html

- STIPA in a controlled environment
  - Noise simulated with crowd loop.
- STIPA in an outdoor environment
  - Natural and mechanical urban background noise.

It is important to note that care was taken to ensure minimal change (<1dBA) in SPL between control (no audio processing) and processed signal tests.

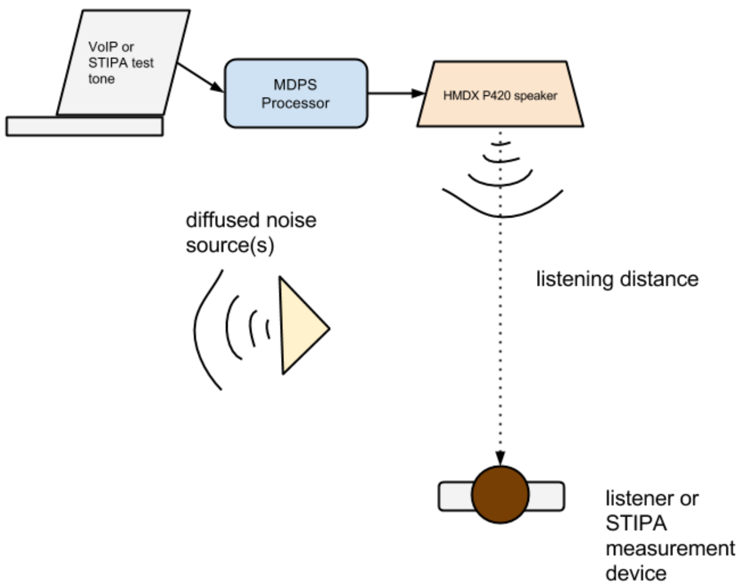
The ANSI Modified Rhyme Test[2]

This ANSI standard is used to measure the intelligibility of English speech as transmitted by a communication system in both outdoor and indoor spaces. Speech intelligibility is measured by comparing monosyllabic words received by trained listeners spoken by trained talkers. The requirement of both listeners and talkers is that they are native speakers of English with no hearing or speech defects.

The test consists of 50 sets with six words in each set. Each word is monosyllabic and most words have three sounds in consonant-vowel-consonant order. The talker speaks a carrier sentence that includes the test word in an obvious location. Example; "Please mark the word *peach*." The set of words for that questions would include *peas, peace, peak, peal, peat* and the correct answer, *peach*.

In this study, the listeners sat in a room 2 meters from the HMDX speaker. Crowd noise was generated and diffused in the room so the level of noise at the listening position was nominally 70dBA SPL. The communication system was calibrated so the SPL level in the environment did not increase when MDPS was switched on. The HMDX speaker and MDPS processing module were connected to a laptop computer running the Polycom software.

[2] *ANSI S3.2-2009 (R2014) Method for Measuring Intelligibility of Speech Over Communication Systems.*



The talker administered the test from a remote location via an internet connection. An off-the-shelf webcam with a built-in microphone was used to capture the talker's voice.

Modified Rhyme Results

Raw data:

Listener #	MDPS OFF	MDPS ON
1	58%	94%
2	72%	94%
3	56%	94%
4	70%	88%
5	44%	82%

Analysis:

Results	**Without DPS**	**With DPS**	**% Improvement**
Mean Accuracy	60%	90%	50%
Median Accuracy	58%	94%	62%

The STIPA Test[3]

The Speech Transmission Index of Public Address Systems (STIPA) is a standardized test used to measure transmission quality of Public Address (PA) systems in their actual or intended environment. STIPA is a variation of an STI (Speech Transmission Index) test which is the international standard for objective rating of speech transmission. STIPA gives a decimal score and corresponding letter grade. These scores range from 0-1 which correspond to letters from U (unusable) to A+ (recording studio quality). STIPA test results are derived from variables such as acoustic reproduction equipment, room anomalies, and ambient noise with regards to how they affect the transmission of the human voice.

The STIPA test signal consists of multiple tones with various modulations that simulate sounds of human speech. A STIPA test is conducted by sending the official STIPA test signal through the system in the location (airport, train station etc) with appropriate ambient noise (conversations at rush hour, mechanical noise from air conditioning etc). While the tone is playing, the audio is captured at the desired listening location and analyzed to create the STIPA value.

[3] *Standard IEC60268-16 ED.4.0 B:2011 Sound System Equipment - Part 16:*

Objective Rating of Speech Intelligibility by Speech Transmission Index

STIPA Results[4]

The indoor test environment was stressed by diffusing crowd noise into the room. The ambient noise was measured at 73dBA SPL with or without MDPS processing engaged. This emulated the environment used during the Modified Rhyme Test.

Indoor Test	**Distance**	**STIPA value**	**STIPA score**
DPS OFF	2 meters	0.35	U
DPS ON	2 meters	0.61	D

Please refer to the STIPA Score Key at the end of this document for a description of each value/score.

Outdoor Test	**Distance**	**STIPA value**	**STIPA score**
DPS OFF	2.3 meters	.744	A
DPS ON	2.3 meters	.974	A+
DPS OFF	4.6 meters	.494	G
DPS ON	4.6 meters	.692	B
DPS OFF	9.1 meters	.360	U
DPS ON	9.1 meters	.618	D

Ambient noise levels did fluctuate during the outdoor test which affected the STIPA value. Each test was performed 5 times and averaged to produce these results. Ambient noise averaged 69dBA SPL with and without MDPS processing engaged.

[4] *Test conducted with 'iSTI Standard' IOS App created by Embedded Acoustics using built-in iPhone 6 microphone.*

Conclusion

Any normally functioning communication system should have a good speech intelligibility score in a quiet controlled environment. However, environments are often chaotic and noisy when information transfer is most critical and lives are at stake.

This study indicates MDPS audio processing has a demonstrable effect on speech intelligibility when applied to a real-world communication system. The improvement delta increases as the listening environment degrades. The effect is large enough (>50% improvement in most cases) to rise above the noise in these tests and warrants further detailed study on a variety of systems.

Applications for MDPS audio processing include these scenarios:

- Public Address systems in high noise environments such as airports, subways and bus terminals.
  - MDPS optimizes the output of a speaker to overcome noise without raising SPL.
- Public Address systems where the area to be covered is large and speaker placement is difficult.
  - MDPS improves the effective range of a speaker so fewer speakers are needed to cover a larger area.
- Telemedicine Systems where doctor's endpoint is noisy and/or speaker/headset quality is not good.
- Telemedicine Systems where clinic endpoint is large and microphone placement is not ideal.
  - MDPS increases the sensitivity of the system so talkers that are far from the microphone are as easy to hear as those close to the microphone.
- Any communication system in a reverberant environment such as a shopping mall or other space with reflective walls and ceilings.
  - The optimized audio signal created by MDPS creates reflections in a room that are coherent across the voice spectrum. This increases intelligibility in reverberant spaces.
- Any communication system experiencing issues with inconsistent volume levels resulting in inconsistent speech intelligibility.

MDPS is currently available for WebRTC-based voice chat systems as [IntelliSpeech™ for Telemedicine](#).

BongioviAcousticLabs

Points of Contact

Bongiovi Medical CMO Daniel N Weiss, MD, FACC, FHRS - dweiss@bmhtllc.com

Bongiovi Medical CTO Joseph Butera III - xaqmusic@bmtdps.com

Bongiovi Medical Business Development Joseph Butera Jr. - jbutera@bmtdps.com

STIPA Score Key[5]

Category	Nominal STI value	Type of message information	Examples of typical uses (for natural or reproduced voice)	Comment
A+	>0,76		Recording studios	Excellent intelligibility but rarely achievable in most environments
A	0,74	Complex messages, unfamiliar words	Theatres, speech auditoria, parliaments, courts, Assistive Hearing Systems (AHS)	High speech intelligibility
B	0,7	Complex messages, unfamiliar words		
C	0,66	Complex messages, unfamiliar words	Theatres, speech auditoria, teleconferencing, parliaments, courts	High speech intelligibility
D	0,62	Complex messages, familiar words	Lecture theatres, classrooms, concert halls	Good speech intelligibility
E	0,58	Complex messages, familiar context	Concert halls, modern churches	High quality PA systems
F	0,54	Complex messages, familiar context	PA systems in shopping malls, public buildings offices, VA systems, cathedrals	Good quality PA systems
G	0,5	Complex messages, familiar context	Shopping malls, public buildings offices, VA systems	Target value for VA systems
H	0,46	Simple messages, familiar words	VA and PA systems in difficult acoustic environments	Normal lower limit for VA systems
I	0,42	Simple messages, familiar context	VA and PA systems in very difficult spaces	
J	0,38		Not suitable for PA systems	
U	<0,36		Not suitable for PA systems	

NOTE 1 These values should be regarded as minimum target values.

NOTE 2 Perceived intelligibility relating to each category will also depend on the frequency response at each listening position.

NOTE 3 The STI values refer to measured values in sample listening positions or as required by specific application standards.

PA = Public Address

VA = Voice Alarm

[5] *IEC 60268-16:2011 Annex G*

