

Contents

Contents	1
Introduction	2
Functionality	2
Requirements	2
Generate test signals	2
Test voice codecs	2
Compare wav files	2
Testing parameters	2
Scientific background	3
AQuA Command Line parameters	4
Print sounds' names: -h sndn	4
Print samples of program usage: -h exam	4
Define program mode> -mode <mod>	5
Command line argument: -clibf <file>	5
Use initial sound file as source or internal signal generator: -src <file <fname> gen <mode>>	5
Set type of weight coefficients: -ct <ctype>	5
Set name of the file being tested: -tstf <fname>	5
Set name of the file generated by the speech model: -dst <fname>	5
Generate full speech sounds distribution or synthesized sound: -sn <all <sname>>	5
Set voice type: -voit <female male>	5
Set duration: -slen <num>	5
Set quality loss or naturalness: -qt <quality naturalness>	5
Enable indication of codec speed performance: -power	5
Enable energy normalization -enorm	5
Set number of link points: -npnt <num>	5
Set envelope smoothing level: -miter <num>	5
Turn on "waiting for key press" after showing voice quality output: -gch	6
Print reason of quality loss: -fau <fname>	6
Set voice quality output type: -ratem <% m p>	6
AQuA Command Line usage	6
Compare two audio files and learn about reasons for voice quality loss	6
Compare two audio files and receive percentage of their similarity	7

Introduction

AQuA is a simple but powerful tool to provide perceptual voice quality testing and audio file comparison in terms of audio quality. This is the easiest way to compare two audio files and test voice quality loss between original and degraded files. Besides this a most demanded functionality of the software can also test audio codecs and generate audio signals for voice quality testing.

AQuA gives a unique opportunity to design your own voice quality testing solution not being dependent on particular hardware and software. It is available for Windows as a DLL and also available for Linux, portable to Java and mobile devices.

Functionality

Requirements

This version of AQuA works with audio files represented in .wav or .pcm formats. Audio files should have the following characteristics: 8Khz, 16 bit, Mono.

Generate test signals

Aqua allows generating test signals with the following parameters:

- Choosing voice type for synthesized sound: male or female
- Allows to define name of the synthesized sound or to run generation of full speech sounds distribution
- Allows choosing duration for synthesized sound equal in amount of samples
- Signals are built according to internal speech model
- Allows test sound signals generated as short, normal and long

Test voice codecs

- Allows output of codec speed performance indicator
- Allows testing any codec library

Compare wav files

- Allows intrusive testing of original wav file against degraded wav file
- Allows testing using internally generated audio test files
- Allows direct wav files comparison quality wise
- Allows measuring voice quality for any language

Testing parameters

AQuA supports the following parameters for voice quality testing:

- Choosing type of quality measurement: overall quality loss or voice naturalness
- Choosing filenames for original, test and generated audio files
- Define type of weight coefficients: uniform, linear or logarithmic
- Allows energy normalization

- Allows setting envelope smoothing level from 1 to 10
- Allows choosing source for original sound (external file or generated internally)
- Contains audio synchronization and voice activity detection
- Provides reasons for voice quality loss (quite unique feature on the market)
 - Duration distortion
 - Changes in signal spectrum
 - Distortion detection in low, medium and high frequency bands
- Provides voice quality feedback in:
 - Percentage of similarity
 - MOS-like value
 - PESQ-like value

Scientific background

The human ear is a non-linear system, which produces an effect named masking. Masking occurs on hearing a message against a noisy background or masking sounds.

As result of the research of the harmonic signal masking by narrow-band noise Zwiker has determined that the entire spectrum of audible frequencies could be divided into frequency groups or bands, recognizable by the human ear. Before Zwiker, Fletcher, who had named the selected frequency groups as critical bands of hearing, had drawn a similar conclusion.

Critical bands determined by Fletcher and Zwiker differ since the former has defined bands by means of masking with noise and the latter – from the relations of perceived loudness.

Sapozhkov has determined a critical band as “a band of frequency speech range, perceptible as a single whole”. In his earlier researches he even suggested that sound signals in a band could be substituted by an equivalent tone signal, but experiments did not confirm this assumption. Critical bands determined by Sapozhkov differ from those determined by Fletcher and Zwiker since Sapozhkov proceeded from the properties of speech signal.

Pokrovskij has also determined critical bands on the basis of speech signal properties. According to his definition the bands provide equal probability of finding formants in them.

The value of spectrum energy in bands can be used for different purposes; one of which is the sound signal quality estimation. However, using only one author’s critical bands (for example, Zwiker’s critical bands are used in prototype) does not allow getting an estimation objective enough, since they show only one of the aspects of perception or speech production. AQuA can determine energy in various critical bands as well as in logarithmic and resonator bands, that allows taking into consideration more properties of hearing and speech processing.

Taking into account that the bands determined by Pokrovskij and Sapozhkov are better for speech signal and not for sound signal, in general allows increasing the accuracy of estimation depending on it purpose.

AQuA utilized research results of the above mentioned scientists implementing different algorithms in one software solution. AQuA also has several advantages compared to other existing voice quality measurement software:

- it is universal since it allows measuring signals quality from various sources and processed in different ways;
- one can optimize quality estimation depending on the purposes:
 - for speed (for example, it is possible to receive rough estimation quickly);
 - in signal type (using different bands for speech signals and sound signals in general);
- resulting estimations correlate well with that of MOS;
- quality estimations received for speech signals can be translated in values of various kinds of intelligibility.

AQuA Command Line parameters

Print sounds' names: -h sndn

sndn - prints list of sounds names;
There are 54 sounds in the database

Num	Name	Type	Num	Name	Type
000	a0 <<---	Vocal	001	a1 <<---	Vocal
002	a2 <<---	Vocal	003	a4 <<---	Vocal
004	e0 <<---	Vocal	005	e1 <<---	Vocal
006	i0 <<---	Vocal	007	i1 <<---	Vocal
008	i4 <<---	Vocal	009	o0 <<---	Vocal
010	o1 <<---	Vocal	011	o4 <<---	Vocal
012	u0 <<---	Vocal	013	u1 <<---	Vocal
014	u4 <<---	Vocal	015	y0 <<---	Vocal
016	y1 <<---	Vocal	017	l <<---	Sonor
018	l' <<---	Sonor	019	m <<---	Sonor
020	m' <<---	Sonor	021	n <<---	Sonor
022	n' <<---	Sonor	023	j <<---	Sonor
024	v <<---	Noised	025	v' <<---	Noised
026	zh <<---	Noised	027	z <<---	Noised
028	z' <<---	Noised	029	r <<---	Noised
030	r' <<---	Noised	031	b <<---	Voiced Explosiv
032	b' <<---	Voiced Explosiv	033	g <<---	Voiced Explosiv
034	g' <<---	Voiced Explosiv	035	d <<---	Voiced Explosiv
036	d' <<---	Voiced Explosiv	037	f <<---	UnVoiced
038	f' <<---	UnVoiced	039	h <<---	UnVoiced
040	h' <<---	UnVoiced	041	s <<---	UnVoiced
042	s' <<---	UnVoiced	043	sh <<---	UnVoiced
044	sch <<---	UnVoiced	045	k <<---	Occlusive
046	k' <<---	Occlusive	047	p <<---	Occlusive
048	p' <<---	Occlusive	049	t <<---	Occlusive
050	t' <<---	Occlusive	051	c <<---	Occlusive
052	ch <<---	Occlusive	053	_ <<---	Silencer

Print samples of program usage: -h exam

exam - prints samples of program usage;

In order to test voice quality between original and reference files use the following set of parameters:

VQcmd.exe -mode files -src file ORIGINAL_FILE -tstf REFERENCE_FILE

To test voice codec provided as a DLL library use the following set of parameters:

VQcmd.exe -mode codec -clibf <DLL_LIBRARY_NAME> -src file <TEST_AUDIO_FILE>, e.g.
VQcmd.exe -mode codec -clibf GSM610.dll -src file short.wav

Define program mode> -mode <mod>

Defines AQuA mode of operation. The following modes are available:

<mod> :

codec - codec testing mode;

files - audio file comparison mode;

generate - test signals generation mode

For example: *-mode codec*

VQcmd.exe -mode files -src file ORIGINAL_FILE -tstf REFERENCE_FILE

VQcmd.exe -mode codec -clibf <DLL_LIBRARY_NAME> -src file <TEST_AUDIO_FILE>

VQcmd.exe -mode codec -clibf GSM610.dll -src file short.pcm

Command line argument: -clibf <file>

- codec library file name;

Use initial sound file as source or internal signal generator:

-src <file <fname> | gen <mode>>

Determines source of initial sound: <file> - external sound file, or internal signal generator <gen>

In <file> mode one should specify name of audio file.

The signal generator mode has one of the following: short, normal or long

Set type of weight coefficients: -ct <ctype>

uniform, linear or logarithmic;

Set name of the file being tested: -tstf <fname>

Set name of the file generated by the speech model: -dst <fname>

Generate full speech sounds distribution or synthesized sound:

-sn <all | <sname>>

runs generation of full speech sounds distribution or defines name of the synthesized sound

E.g. *VQcmd -sn all -dst SPEECH_MODEL_FILE*

Set voice type: -voit <female | male>

- sets voice type for synthesized sound;

Sets duration: -slen <num>

sets duration for synthesized sound equal to <num> samples

Set quality loss or naturalness: -qt <quality | naturalness>

- sets type of quality measurement **overall quality loss** or **voice naturalness**

Enable indication of codec speed performance: -power

enables output of codec speed performance indicator;

Enable energy normalization -enorm

- enables energy normalization;

Set number of link points: -npnt <num>

sets number of link points;

Set envelope smoothing level: -miter <num>

smoothing level is in the range of (1..10);

Turn on “waiting for key press” after showing voice quality output: -gch

turns on waiting for a key press after output of voice quality

Print reason of quality loss: -fau <fname>

- prints reasons for quality loss to the file specified;

Set voice quality output type: -ratem <% | m | p>

%: voice quality in percentage,

M: MOS-like estimation

P: PESQ-like estimation.

AQuA Command Line usage

Most of our customers can be divided into two main categories:

- those who would like to compare original audio against degraded (test voice quality between the original and reference files) and obtain quantitative reasons for voice quality loss.
- those who like to compare two audio files and find out how similar they are quality wise

To show how this works we are going to use WAV files one can find on the site of Microtronix (<http://www.microtronix.ca/pesq-disc.html>). However, one can use any audio files that are recorded at 8Khz sampling and are 16 bit mono.

Compare two audio files and learn about reasons for voice quality loss

To compare two audio files in AQuA Command Line version when one is interested to get extensive feedback from the software we suggest to invoke AQuA in the following manner:

VQcmd.exe -mode files -src file Or272.wav -tstf Dg002.wav -ratem %mp -fau log.txt

As result you will received the following output:

Sevana Voice Quality Analyzer.

Copyright (c) 2009 by Sevana Oy, Finland. All rights reserved.

File Quality is

Percent value 36.11

MOS value 1.49

PESQ value 2.56

Thus one can see that file comparison gives only 36.11% of similarity what corresponds to 1.49 MOS and about 2.56 PESQ. By the way, this is an example of when AQuA does detect voice quality loss and PESQ does not (please read more details about this test case on Microtronix page).

After test was excicute log.txt file contains quantitative reasons for voice quality loss:

Duration distortion.

Audio stretching corresponds to 1.42 percent.

Delay of audio signal activity.

Signal delayed by 100.000000 ms.

Audio signal activity mistiming (unsynchronization) is 1.13 percent.

Corrupted signal spectrum.

Overall spectral energy distortion approaches 64.47 %

Vibration along the whole spectrum [-16.51, 47.97] %

Significant distortion in low frequencies band.

Energy distortion approaches 27.15 %

Spectrum vibration in low frequency band [-13.93, 13.22] %

Significant distortion in medium frequencies band.

Energy distortion approaches 29.41 %

Amplification approaches 26.88 %

When invoking AQuA with the parameters above one receives the most detailed response from the software on voice quality and reasons for audio being degraded.

Compare two audio files and receive percentage of their similarity

In case we like to simply compare two audio files and get feedback on how similar they are quality wise we suggest invoking AQuA in the following manner:

VQcmd.exe -mode files -src file Or272.wav -tstf Dg002.wav -ratem %

Result will be:

Sevana Voice Quality Analyzer.

Copyright (c) 2009 by Sevana Oy, Finland. All rights reserved.

File Quality is

Percent value 36.11

or invoking it for the other degraded file:

VQcmd.exe -mode files -src file Or272.wav -tstf Dg001.wav -ratem %

Sevana Voice Quality Analyzer.

Copyright (c) 2009 by Sevana Oy, Finland. All rights reserved.

File Quality is

Percent value 50.47

It is easy, fast and very effective!