

Automatic Sound Signals Quality Estimation

Sevana Oy

Endre Domiczi

Mobile: +372 53485178

E-mail: ceo@sevana.fi

Automatic Sound Signals Quality Estimation

- General outline and the basic components of the system – signal, synchronization and analytical components
- Test signal (including statistic speech model) and sound perception models
- Adequacy of analytical estimations based on the results of the comparison of received analytical and subjective MOS estimations
- Acoustic model further developments
- Available software
- System applications

Review of existing quality estimation methods

- **AI (Articulation Index)**. The idea is that the whole frequency range of speech signal is divided into 20 bands and the signal/noise ratio is determined within the band. The band broad is defined in such a way, that every band contributes equally in speech perception. The signal/noise ratio is calculated within every band. Articulation index is supposed to be equal the weighted total of the band values. **The articulation index does not take into account the properties of hearing and speech production, although it directs toward speech signal.**
- **SII (Speech Intelligibility Index)** is the evolution of AI method. The American Standard ANSI S3.5-1997 includes the speech intelligibility index. It provides for 4 measuring procedures on different band groups: 21 critical bands, 18 one third-octave bands, 17 equal by their contribution critical bands and 6 octave bands. The signal/noise ratio is calculated within every band and the total SII coefficient, ranged from 0 to 1 is computed. **The speech intelligibility index takes into account only the properties of hearing, not speech production.**

Review of existing quality estimation methods

- **STI (Speech Transmission Index).** We may approximately consider speech signal as broadband signal modulated by low-frequency signal. Articulation speed determines modulation frequency. When modulation depth decreases, speech signal becomes similar to noise and its intelligibility decreases. Accordingly, intelligibility decrease can be estimated according to modulation depth decrease also. Whole speech range is divided into 7 octave bands. An octave noise signal is the input. The test signal intensity distribution agrees with the distribution of speech signal intensities. The modulating signal frequencies vary from 0.5 to 12.5 Hz with one-third-octave interval (14 frequencies in all). The STI measuring method is stated in the International standard IEC 268-16.
- **RASTI/STIPA (Rapid Speech Transmission Index).** The STI method needs a lot of measuring procedures and calculations. A simplified method was developed, which provides for measuring only in 2 bands with 5 modulation frequencies and reduces the number of measuring procedures and calculations. For good intelligibility RASTI values must be not less than 0.6.

Speech transmission index as well as rapid speech transmission index imitates speech production process by means of noise model, but to take into account the properties of speech production and hearing in such a way is far from optimum.

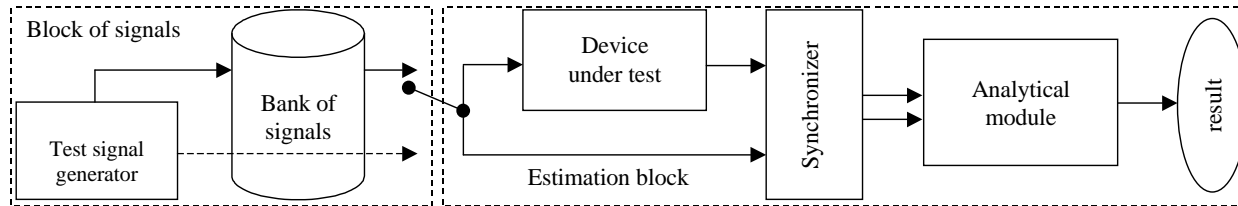
Review of existing quality estimation methods

- **C50 (factor of clearness)** determines sound clearness and clarity. It is computed as near echo/far echo ratio. The method is based on the fact, that echo reduces signal intelligibility. The near echo/far echo ratios in several frequency bands are calculated. They consider near echo (less than 33 ms) as useful signal and far echo (more than 33 ms) as disturbing signal. **The factor of clearness takes into account only the one kind of the possible distortions and it is worth to apply it only as one of the speech quality estimations.**

The need to develop new methods and to improve existing ones is caused by desire to bring together objective and subjective estimation of quality and to explicitly use in such systems our knowledge about hearing and speech production.

To use arbitrary or particularized signal as a source signal depends on the estimation purpose (speech intelligibility evaluation, sound reproduction quality, quality estimation of speech, transmitted through intercommunication channels, etc.) and allows increasing estimation objectivity.

Introduced System General Concept



Generator of test signals allows sound signal forming according to one of the sound flow models. It can be either a particularized set of sound signals or a signal, received in output of statistical speech model. Generator's signal can either be saved for follow-up usage or be exposed to processing and estimation. Bank of signals stores sound data, received as a result of signals' generator work or from some external sources.

Accordingly, an input of estimation block is a signal of generator directly or one of the bank of signals. Test signal is the input of the synchronizer or of the device under test, which can be for example, a vocoder or a communication channel. The output signal of the device under test is an input of synchronizer also.

The synchronizer matches in time an initial signal and a processed signal. The synchronized signals in chunks input in analytical module, which determines the degree of similarity for signals and issues the quality estimation as the measure of similarity between the initial and the processed signals.

Implementation

Algorithms described are implemented in Sevana Audio Codecs Analyzer for vocoder quality estimation and comparison of external initial signals and signals under test.

As the external signals arbitrary signals recorded with the sampling frequency of 8 kHz and the capacity of samples equal 16 bits can be used. Supposed, the signal under test is received from an initial signal as a result of some transformations (for example, compression/restoration, transmission through communication channels, filtration).

As internal initial signals (i.e. signals, which the user of the program has no access to) the signals generated according to the proprietary noise model and the signals, generated on the basis of the statistic model.

Internal input signals of sound data to the system, are implemented as DLL. One can use both DLLs provided within Sevana Audio Codecs Analyzer or developed by the others. The signal processed by means of methods contained in DLL is consider as the signal under test and is exposed to the proprietary quality estimation procedure.

Advantages

- It is a universal tool since it allows judging the quality of signals from various sources and processed in different ways;
- One can optimize quality estimation signal depending on the purposes:
 - in 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.

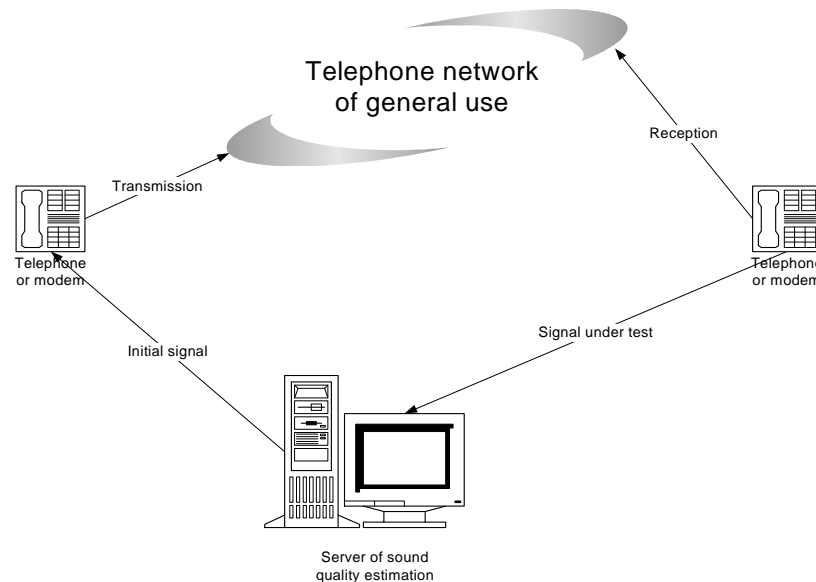
Test Results

Codec	MOS	Noise model						Statistic model		PhRT	
		Minimal		Reduced		Complete		-	Vc	-	Vc
		-	Vc	-	Vc	-	Vc	-	Vc	-	Vc
A-Law	4,10	4,79	4,73	4,78	4,78	4,78	4,78	4,79	4,80	4,80	4,84
Mu-Law	4,10	4,79	4,84	4,77	4,77	4,77	4,78	4,78	4,79	4,79	4,82
G.723.6.3	3,90	4,25	4,48	4,21	4,29	4,22	4,33	4,15	4,04	4,08	3,95
GSM.6.10	3,70	3,20	1,99	3,01	1,65	3,04	1,78	4,22	3,66	4,01	3,21
G.723.5.3	3,65	4,23	4,44	4,18	4,27	4,19	4,32	4,14	4,04	4,06	3,93

The table above represents quality estimations of several standard vocoders, received on various test signals using the proprietary method and Sevana Audio Codecs Analyzer. The table contains MOS estimations for comparison.

Estimations under the assumption, that bands are of equal probability, are in the column with «-» symbol and the estimation received under taking into account the coefficients of importance are in the column with «Vc»

Applications: quality estimation of sound transmission through telephone network of general use

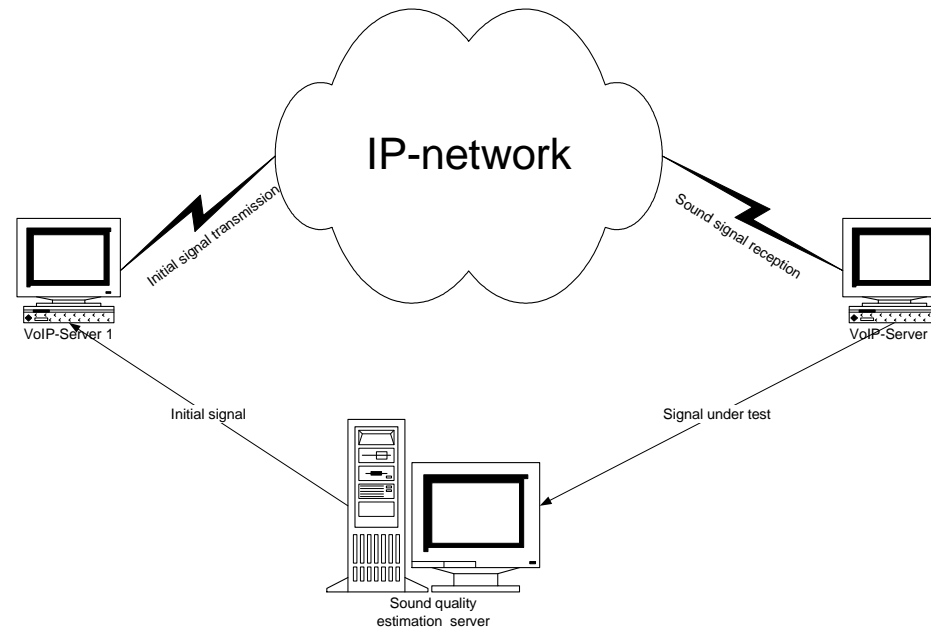


The picture represents applying the method described above for quality estimation of sound transmission through telephone network of general use. The given scheme is applicable both for local and for long-distance connections.

Server of sound quality estimation generates an initial signal (or chooses among signals prepared before) and transfers it to one of the telephone subscribers taking part in the testing. The subscriber received the signal establishes a standard connection with the second subscriber and reproduced the initial signal. The second subscriber records the receiving sound signal and transfers it to the sound quality estimation server.

The sound quality estimation server compares the initial and the test signals according to the suggested method and gives the quality estimation of the signal transferred through the telephone network. The received estimation can be used for improving the subscriber service, deciding about the necessity of equipment changing or setting (both on the side of the subscriber and on the station), as advertising and others.

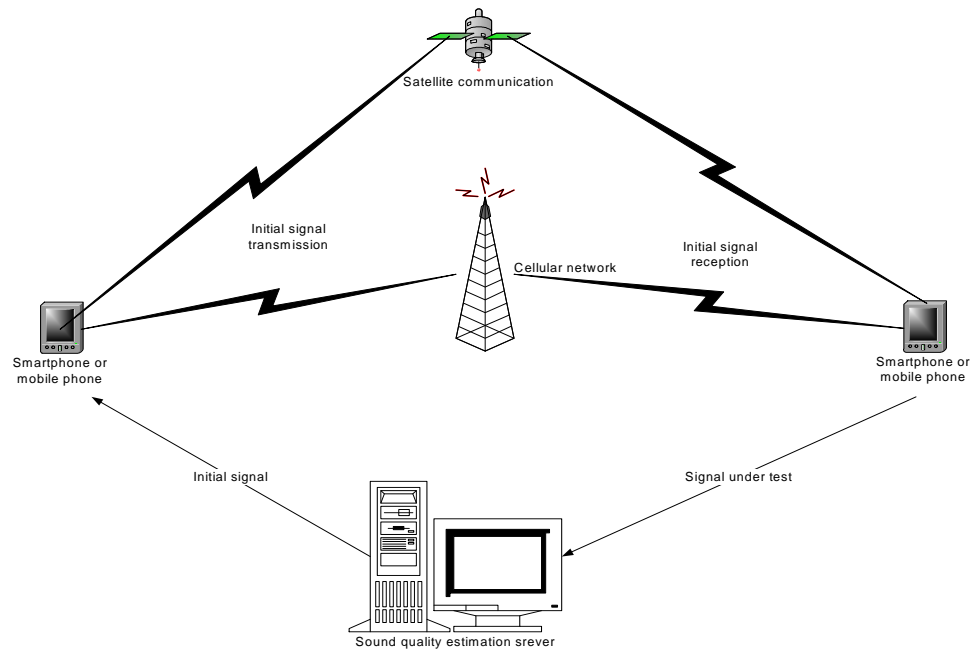
Applications: quality estimation of sound transferred through IP-network



Similarly to telephone network quality estimation of sound transferred through IP-network is performed. It differs from the previous application in the way of transferring the initial and the signal under test from the sound quality estimation server to subscribers and in the way of data transfer between subscribers.

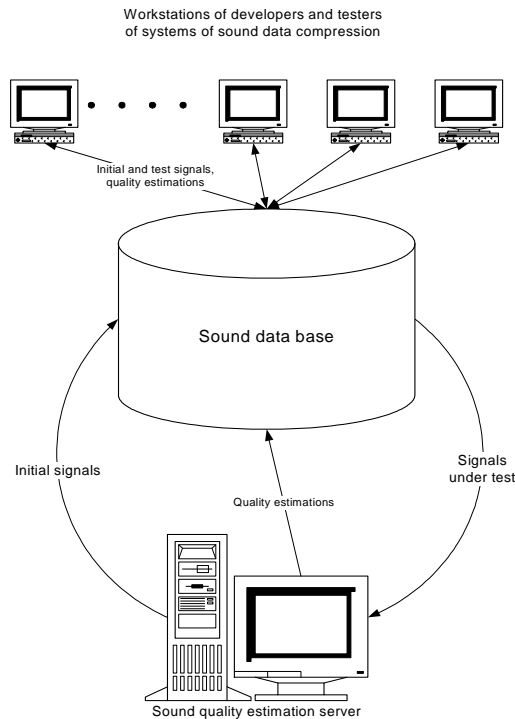
Quality estimations received can be used for choosing codecs used in VoIP-connection and when choosing operators, providing IP-telephony services.

Applications: sound quality estimation of cellular and satellite connection



The introduced method and software can be effectively used for quality estimation of cellular and satellite connections. Received estimations subscribers use for choosing operators and telephone models and operators use for optimization of base station locations.

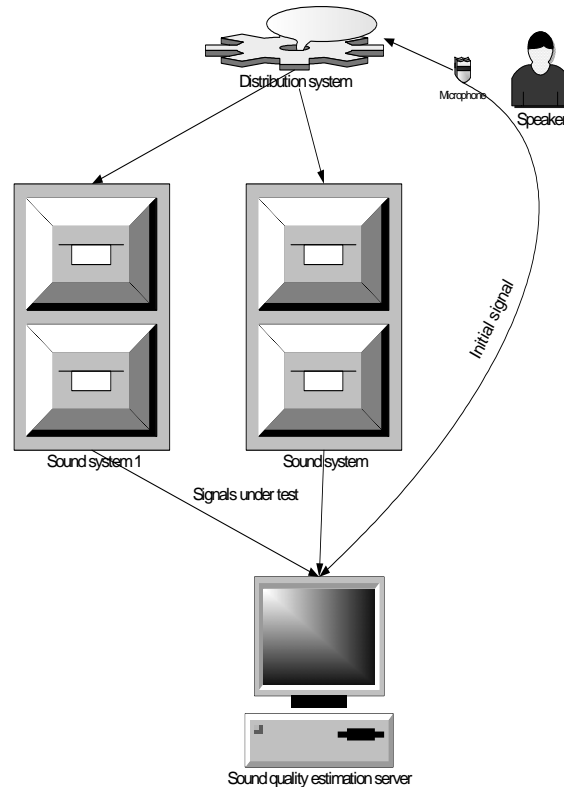
Applications: quality estimation of systems and algorithms (methods) for sound data compression



Every codec version (or codec with a set of parameters) requires estimation and comparison with the analogs. Every developer can refer to sound sample base, compress and restore a signal and receive objective quality estimation of the codec work.

Such a system allows managing the codec developing process and the optimization of their parameters much more effectively. Ultimate consumer will be able to receive not just functioning, but optimal algorithm.

Applications: rooms' sound quality estimation



In this case initial signal is a signal from the microphone located opposite the speaker and signals under test are those from microphones located in different parts of the room, in places where hearers and sound reproducing equipment are located.

The received estimations can be used for optimization of the location of sound reproducing equipment, furniture and spectators' places.

Further Development

- Integration with existing Quality of Service and Quality of Experience systems to increase their functionality and enrich test impact.
- Test signal model improvement. Here the noise model can be supplied with a set of multiband modulated noise signals; enrich the set of data and algorithms of the statistic speech model, increase the number of prepared test signals (such as records of PhRT);
- Development of more upgraded algorithms of synchronization, based, for example, on coincidence of maximums in signal energy spectrums;
- Acoustic model modernization with taking into account masking effects and the fact that pure tones and band noise cause difference in hearing;
-
- Signal comparison scheme modernization. Current distance measure can be more accurate for strongly different signals. For higher universality of the system it is desired to use the correlation analysis methods for comparison;
- Solve a number of practical problems the systems requires the possibility to work with multichannel (Stereo-, Quadro-, etc.) and to receive immediate quality estimations;
- Absolutely correct translation of the objective estimations into MOS estimation values requires further experimental researches.

THANK YOU!