- [ ] 2700 Series
- [x] APx555
- [x] APx585 Series
- [x] APx525 Series
- [x] APx515

ABC-MRT IN APX AUDIO ANALYZERS

AN OBJECTIVE ESTIMATE OF SPEECH INTELLIGIBILITY

by Joe Begin

About This Technote

In this technote we describe ABC-MRT, a new speech processing algorithm that provides an objective estimate of the level of speech intelligibility in communication systems. We also describe an ABC-MRT measurement for APx500 audio analyzers, provided in the form of a software option.

ABC-MRT is an acronym for Articulation-Band Correlation Modified Rhyme Test. It follows the paradigm of the Modified Rhyme Test, a well-known technique for assessing speech intelligibility that is based on subjective assessment by human subjects. ABC-MRT was developed by the Institute for Telecommunication Sciences (ITS) of the United States National Telecommunications and Information Administration (NTIA). It uses a form of automatic speech recognition (ASR) to conduct a type of automated Modified Rhyme Test that does not involve human subjects.

Speech Intelligibility

The purpose of a communication system is to transmit information via speech. As such, the *intelligibility* of the transmitted speech is of utmost importance. Speech intelligibility is sometimes defined as a measure of how comprehensible or understandable a speech signal is. A more precise definition is the proportion of speech items (words and/or speech sounds) uttered by a talker that can be recognized by the listener.

The measurement of speech intelligibility is a time-honored problem and many techniques have been developed over the years for this purpose.

Modified Rhyme Tests

One class of methods to evaluate speech intelligibility is based on the direct approach of listening by human subjects. Carefully prepared speech source materials are read or played through a communication system to trained listeners in a controlled laboratory environment. The listeners then respond by answering questions or repeating what they heard. Based on analysis of these responses, conclusions are drawn regarding the speech intelligibility of the system.

ANSI/ASA standard S3.2 [1] prescribes a method for evaluating speech intelligibility that is based on this direct approach of human listening. In the standard, the intelligibility of speech over a speech communication system is measured by comparing monosyllabic words that trained listeners receive and identify with the words that trained talkers (or speech coders) speak into the system that connects the talkers with the listeners.

Three alternative tests are outlined in ANSI S3.2, each based on a different set of English language speech test materials. These are the Phonetically Balanced Word List, the Modified Rhyme Test, and the Diagnostic Rhyme Test. The standard states, *"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, and with bandwidth compression and encryption devices. The three tests are highly intercorrelated with each other and with other tests of speech intelligibility."*

A key factor to the success of these tests is the elimination of context, because context can help listeners to guess at words they do not recognize, resulting in an incorrect estimate of intelligibility. Rhyme testing minimizes context by presenting listeners with sets of rhyming single-syllable words in random order. The Modified Rhyme Test (MRT) described in [1] is one such test that is specified by the US National Fire Protection Association (NFPA) for critical communications testing.

MRT is based on 50 lists of six rhyming words of the form consonant–vowel–consonant. The six words in each list

differ only in the beginning or ending consonant. A trial consists of presenting the listener with a sentence containing a carrier phrase and the keyword, e.g., "Please select the word *bent.*" The listener than selects the word that he or she heard from a list of six options presented on a visual display (e.g., *went, sent, bent, dent, tent,* and *rent*). The words are presented in random order and the rate of correct word identification is used to calculate the MRT Intelligibility Score ϕ, a value between 0 and 1, where 1 represents a perfect score.

Speech intelligibility tests based on human subjects provide meaningful results if prescribed standards are carefully followed. However, these tests are expensive, because they require multiple trained talkers and listeners, as well as specialized test facilities.

Articulation Band Correlation MRT (ABC-MRT)

Researchers at ITS have developed a very effective objective estimator of speech intelligibility that follows the paradigm of the MRT. Known as Articulation Band Correlation Modified Rhyme Test (ABC-MRT), the method, first published in 2013, was limited to narrow band communication systems (250 Hz to 3.85 kHz bandwidth) [2]. In 2016, the method was extended for use with speech systems having bandwidths up to full-bandwidth (20 Hz to 20 kHz), and it was enhanced by adding an attention model to the algorithm. This enhanced method is referred to as ABC-MRT16 [3].

At its core, the ABC-MRT method uses a highly specialized automatic speech recognition (ASR) algorithm that emulates the MRT task – recognizing keywords transmitted through a communication system. Many ASR tools are available today, and these systems are complex in the sense that they are required to work with huge vocabularies, with any speaker and to be robust to impaired speech. ASR in the context of MRT is much simpler: The required vocabulary is tiny, it involves only a few speakers who are known in advance, and context need not be considered.

The ABC-MRT16 Algorithm

ABC-MRT uses frequency bands called Articulation Index (AI) bands that are known to be important to speech articulation. There are 17 AI bands for narrow band speech and 21 AI bands for full band speech.

The algorithm uses time-frequency (T-F) representations of the keywords in the MRT. The T-F pattern of an impaired speech signal is correlated with corresponding patterns of the six unimpaired word options from the list of 6 rhyming words in multiple AI bands. This is illustrated graphically in the figures below.

Figure 1[1] is a spectrogram of the sentence, "Please select the word *went,*" as spoken by Female 1 in the ITS ABC-MRT speech database. It shows frequency from 0 to 10 kHz on the vertical axis, and a time span of approximately 2.0 seconds on the horizontal axis, with color representing the signal level from low (blue) to high (red). The portion of the spectrogram containing the keyword *"went"* is highlighted with a red rectangle in Figure 1. Figure 2 shows the spectrograms of the six keywords from the same list as the word *"went"*, plotted on the same time and frequency scales as Figure 1. By studying the images, you can see how the correct word could be identified based on the T-F pattern.

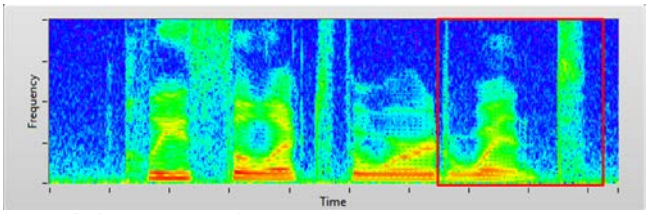


Figure 1. Spectrogram of the text *"Please select the word went".* The red rectangle shows the portion containing the keyword "went".

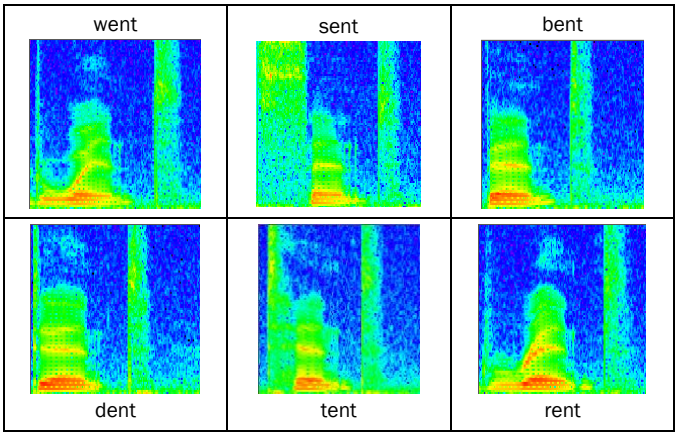


Figure 2. Spectrograms of the 6 keywords in the list.

[1] *Note: The spectrograms in Figures 1 and 2 are for illustration purposes only.*

ABC-MRT16 uses all six words from all 50 MRT lists in ANSI/ASA S3.2. At ITS, each of these 300 words was read (with a carrier phrase) by two female and two male talkers and recorded to .wav files at 48 kHz sample rate, resulting in 1,200 recordings. For each recording, the keyword was isolated, and a T-F pattern was created and stored in a template file.

The ABC-MRT algorithm is described in detail in [2] and [3]. A high-level description of the algorithm follows. See Figure 3 for a block diagram representation.

For each trial:

1. Generate a sentence containing the carrier phrase and the keyword and pass it through the system or device under test (DUT)[1]. Note: At this point in the process, the talker and the keyword are known.
2. Acquire the signal from the output of the DUT.
3. Transform the acquired signal to a T-F pattern using the same technique applied to the 1,200 isolated keyword recordings in the database.
4. Locate the keyword within the acquired T-F pattern from the DUT by cross-correlating it in two of the AI bands with the T-F pattern of the original keyword and talker from the T-F template file. The two AI bands that typically contain the most speech power (from approximately 500 to 800 Hz frequency) are used for this purpose.
5. Extract the portion of the DUT's T-F pattern containing the acquired keyword.
6. Attempt to match the keyword acquired from the DUT with one of the six rhyming keywords from the same list, as follows. For each of the six keywords:
   a. Cross-correlate the DUT's keyword T-F pattern with the T-F pattern of the keyword and talker in all the AI bands of interest.
   b. Accumulate the correlation results across AI bands and eliminate any negative results. This results in 17 (narrow band) to 21 (full band) AI band correlations for each candidate keyword.
   c. Sort the AI band correlations in descending order. This organizes the AI band "evidence" from most convincing to least convincing.
   d. Select the 16 AI band correlations with the highest correlation values.
7. This results in a selected keyword for 16 of the 17 to 21 AI bands analyzed. These selected keywords are then compared to the known correct keyword for the trial. The success rate c for the trial is then calculated as the average number of correct keyword selections over the 16 AI bands.

The success rate c is then averaged over all the trials for a combination of test condition and DUT to produce the average success rate $\bar{c}$. Following standard MRT procedure, this average success rate is then corrected for guessing, to yield the ABC-MRT16 Intelligibility Score c'.

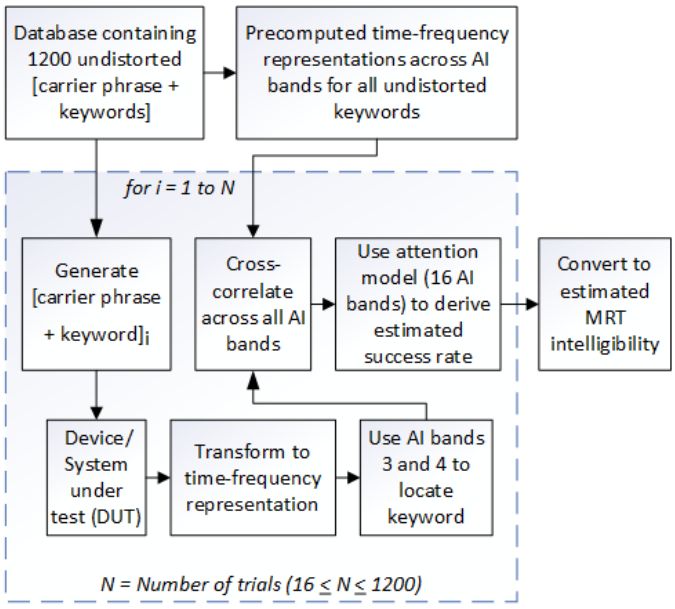


Figure 3. Block diagram of ABC-MRT algorithm.

Evaluation of ABC-MRT and ABC-MRT16

Researchers at ITS have evaluated the performance of ABC-MRT and ABC-MRT16 using real MRT tests with human subjects. These results are listed in Table 1, where the symbol ρ represents the Pearson correlation coefficient, a normalized measure of the covariance between c' (the ABC-MRT or ABC-MRT16 estimated intelligibility score) and ϕ

[1] In this document, the acronym DUT is used to refer to either a system under test or a device under test.

(the human-based MRT score) on a per condition basis. The Pearson coefficient reports how well the relative scoring of ABC-MRT16 and MRT agree, with 1.0 representing perfect correlation. The column headed RMSE in Table 1 shows an absolute measure of the difference between the two methods and has the same units as the MRT scores ϕ, with 0.0 representing perfect agreement.

The first row in Table 1 shows the values for 139 conditions involving narrow band speech systems, for which MRTs were conducted to support the land-mobile radio (LMR) needs of public safety officials, especially firefighters. In this case, original MRT recordings were mixed with high levels of recorded background noise from sources including alarms, saws, pumps and crowds, then passed through protective masks and various components of digital and analog LMR systems.

The second row in Table 1 shows the values for 367 total conditions, comprised of the 139 narrow band conditions supplemented by 228 additional conditions including many different codec modes, four speech bandwidths from narrow band to full band, six different noise environments and additional impairments described in [13].

	ABC-MRT		ABC-MRT16	
	ρ	RMSE	ρ	RMSE
139 NB Conditions	0.955	0.073	0.971	0.061
367 Conditions	0.874	0.124	0.954	0.066

Table 1. Agreement with MRT for ABC-MRT and ABC-MRT16. 367 conditions including NB, WB, SWB, and FB.

As shown in Table 1, ABC-MRT16 offers extremely high correlation with subjective MRT and low estimation error, especially considering the large diversity of conditions tested. This is shown graphically in Figure 4, which is a per condition plot of ABC-MRT16 intelligibility score c' versus subjective MRT intelligibility score ϕ for the 367 test conditions. Table 1 also shows that ABC-MRT16 performs significantly better than ABC-MRT for the 139 NB conditions, indicating the importance of the attention model in ABC-MRT16.

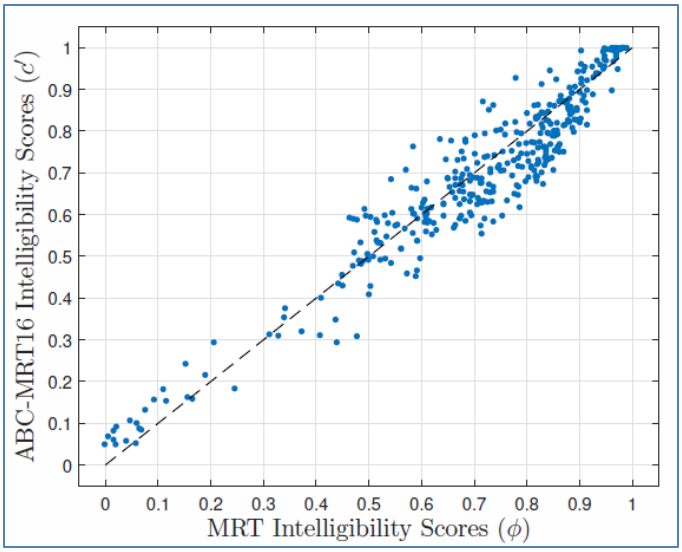


Figure 4. ABC-MRT16 compared with MRT on 367 conditions (NB, WB, SWB and FB), from [3]

ABC-MRT16 Measurements in APx

The ABC-MRT measurement for Audio Precision APx audio analyzers is distributed as a software measurement option, meaning it is not built into the APx500 software. To add the capability to APx, users must first download and run an installer.

> Note: The ABC-MRT option installer is a large download (approximately 677 MB), in part because it includes the 1,200 .wav files of original MRT recordings and the T-F template file containing the 1,200 T-F patterns of the keywords.

Once the option has been installed, a new measurement named ABC-MRT appears in the Add Measurement dialog in Sequence mode (Figure 5), and the dependent files are deployed to the required locations on the user's PC.

The measurement in APx is named ABC-MRT, but it implements the ABC-MRT16 algorithm described in [3].

> Note: The ABC-MRT measurement is a software option and requires a software key to function; the option can be demonstrated without a software key in APx Demo Mode. The APx ABC-MRT option also requires v4.5.2 or later of APx500 measurement software.

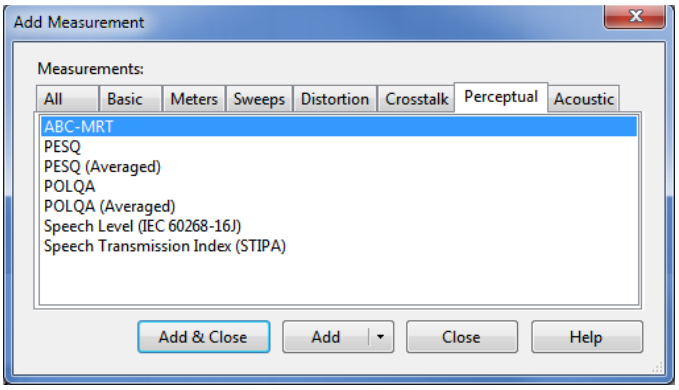


Figure 5. The ABC-MRT measurement appears in the Perceptual and All measurement tabs.

The Generator and Analyzer controls for the ABC-MRT measurement are shown in Figure 6.

Generator Controls

The field labeled **Source File** in the generator section is a read-only field that is used to display the name of the recorded speech signal currently being generated, as the measurement progresses through the specified number of trials. The source file names are encoded to include the talker ID, the MRT word list (1 to 50) and the number of the word in the list (1 to 6). For example, the source file name shown in Figure 6 ("M4_b2_w5_orig") indicates that it is the undegraded source recording of talker Male 4 speaking word #5 from list #2.

The **Level** field in the Generator section is used to specify the signal level. Note that, like all APx measurements, the Level control is calibrated for sine waves. Since this application uses speech signals, which have a much higher crest factor than sine signals, the specified generator level must be increased accordingly. You can select the appropriate level by measuring the level of the generated signal and adjusting the Generator Level control as needed.

The numbered buttons labeled **Channels** in the Generator section are used to enable or disable output channels. If a channel is disabled, no signal is generated on that channel.

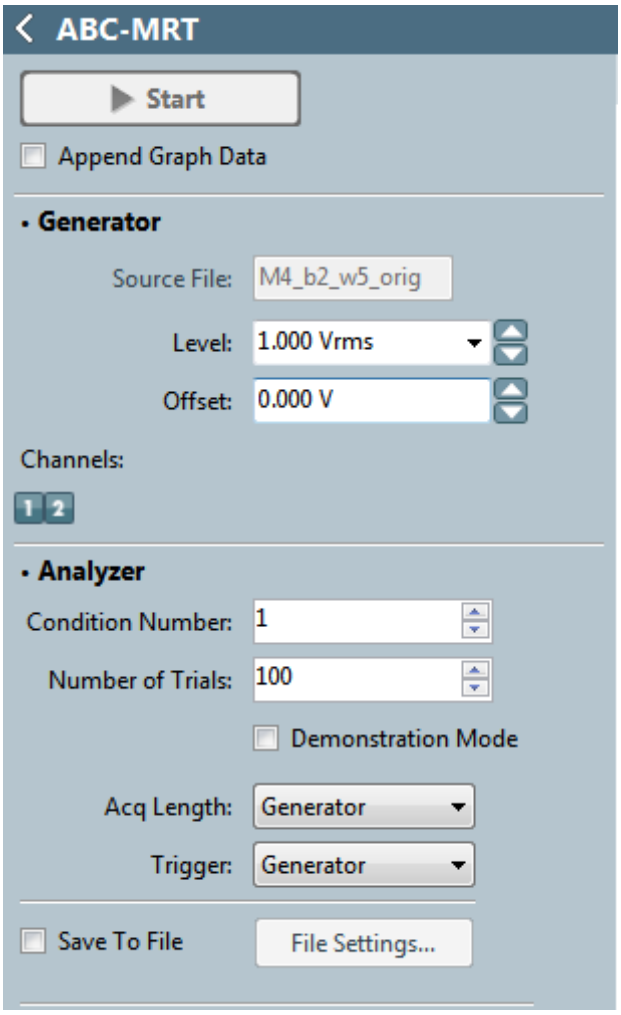


Figure 6. ABC-MRT measurement controls.

Analyzer Controls

The **Condition Number** field in the Analyzer section is used to specify a number from 1 to 99, identifying a test condition for a DUT.

The **Number of Trials** field in the Analyzer section is used to specify the number of MRT trials to be tested for the specified condition.

If the **Demonstrate Mode** checkbox in the Analyzer section is checked, the measurement will be run in a special mode used to demonstrate the ABC-MRT16 algorithm applied to degraded speech samples. In this case, the Condition Number is constrained to a value between 1 and 6, and the Number of Trials is constrained to a value between 1 and 16. In demonstration mode, the measurement generates up to 16 degraded speech files from one of six conditions provided by ITS for demonstration purposes. This mode is

intended for demonstrating the analysis with the analyzer's outputs channels looped back to its inputs.

The **Acq. Length** control in the Analyzer section is used to specify the length of the signal acquisition. The default value of "Generator" is the best choice for this control. It will set the acquisition length to match the length of the speech signal being generated.

The **Trigger** control in the Analyzer section is used to specify the acquisition trigger. It should also be set to the default value of "Generator". This will cause the acquisition to start at the same instance that the speech signal is generated.

Note: If there is a significant delay through the DUT, you can compensate for this by using either the DUT Delay field in Signal Path Setup, or the Delay Time field in the Advanced Settings section of the ABC-MRT measurement. This is necessary if the delay causes the keyword to be truncated, which would cause an incorrect score.

Running the ABC-MRT Measurement

Like any Sequence Mode measurement in APx, there are two options for running the ABC-MRT measurement:

1. Click the Start button in the measurement, or
2. Check the ABC-MRT measurement in the Navigator and run the Sequence.

With either option, the ABC-MRT measurement will loop over the specified Number of Trials, generating a different MRT speech file and analyzing the acquisition each time. The MRT speech files are selected from a master list prepared by ITS. The same order of samples from the master list is repeated for each test.

The ordering of samples in the master list may appear somewhat random, but it is not. The order was carefully selected by ITS for good truncation properties and a good balance of talkers. In other words, the deviation from the full result using 1200 trials is minimized when a lower number than 1200 is used.

When the ABC-MRT measurement is run from the APx Sequencer, in addition to running the measurement, any enabled Sequence Steps are executed, results are evaluated against limits (if limits have been specified) and results are accumulated in the APx report.

ABC-MRT Measurement Results in APx

The ABC-MRT measurement in APx contains four primary results. The first is a meter (single valued) result named Overall ABC-MRT Score (Figure 7). This result contains the average ABC-MRT16 score over all trials, corrected for guessing, (*c'*) and is updated on each trial.

The second primary result is an XY graph named ABC-MRT Score by Trial (Figure 8). This result shows the ABC-MRT16 success rate (*c*) for each trial by trial number. It is also updated on every trial.

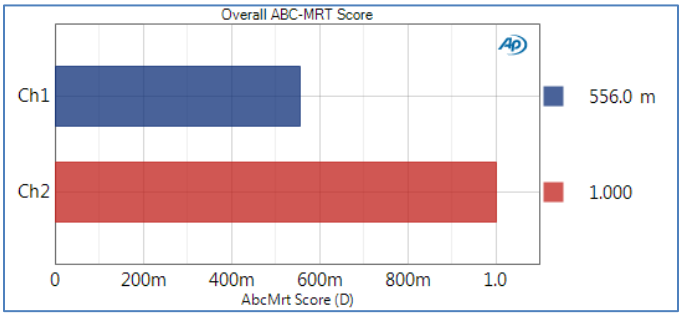


Figure 7. The Overall ABC-MRT Score meter result.

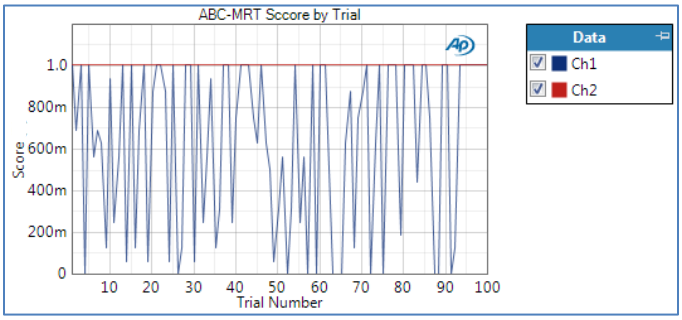
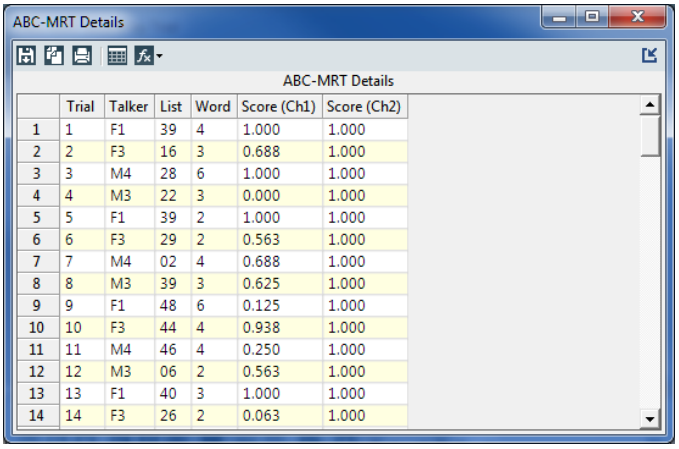


Figure 8. The ABC-MRT Score by Trial result.

The third result is a table named ABC-MRT details (Figure 9). This result contains details for each trial including the Talker ID, the List number and Word number, as well as the ABC-MRT16 score for each channel.

The fourth result is the Acquired Waveform result (Figure 9). This result is not present by default when an ABC-MRT measurement is added to the sequence but may be added using the Add Primary Result button on the toolbar below the result graphs. The Acquired Waveform result is a simple plot of the last signal waveform acquired for all enabled input channels.



	Trial	Talker	List	Word	Score (Ch1)	Score (Ch2)
1	1	F1	39	4	1.000	1.000
2	2	F3	16	3	0.688	1.000
3	3	M4	28	6	1.000	1.000
4	4	M3	22	3	0.000	1.000
5	5	F1	39	2	1.000	1.000
6	6	F3	29	2	0.563	1.000
7	7	M4	02	4	0.688	1.000
8	8	M3	39	3	0.625	1.000
9	9	F1	48	6	0.125	1.000
10	10	F3	44	4	0.938	1.000
11	11	M4	46	4	0.250	1.000
12	12	M3	06	2	0.563	1.000
13	13	F1	40	3	1.000	1.000
14	14	F3	26	2	0.063	1.000

Figure 9. The ABC-MRT Details tabular result.

The example shown in Figure 7 through Figure 10 Is for a test in which 100 trials were analyzed. For Channel 1, a type of noise (speech babble) was mixed with the original speech signals at a speech to noise ratio of approximately 0 dB. For Channel 2, the generated speech signals were acquired directly (looped back from output to input) with no impairments. Note that Channel 2 has a perfect ABC-MRT score of 1.0, as expected, and for Channel 1, the ABC-MRT score by trial varies from 0.0 to 1.0, with a mean value (corrected for guessing) of 0.556. This highlights the need for many trials.

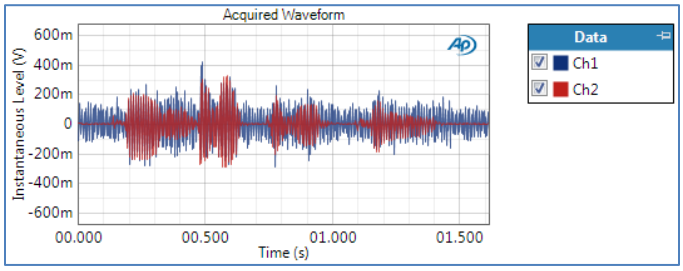


Figure 10. The Acquired Waveform result.

Saving ABC-MRT Results in APx

Like all XY graphs in APx, the ABC-MRT Score by Trial result will be saved with the APx project file, by default. APx does not save meter or tabular results in project files. If a record of these results is required (beyond the APx500 report), results can be exported either manually, or using a series of Sequence Steps.

The Number of ABC-MRT Trials

Recommendations on conducting ABC-MRT and ABC-MRT16 tests from ITS indicate that, to be most thorough, the number of trials should be set to 1,200 (i.e., all MRT speech samples should be used). For expediency, a subset of the 1,200 speech samples can be used, but the minimum number of trials should not be fewer than 16.

ITS characterized the deviation in ABC-MRT results caused by using fewer than 1,200 trials, relative to results using all 1,200 trials for the 139 narrow band conditions described above. Results of this characterization are shown in Table 2. The column headed Sample Time in Table 2 lists the total time required to pass the N speech samples through the DUT. Note that ABC-MRT scores range from 0 to 1, so a deviation of 0.01 corresponds to 1% of the intelligibility scale.

The results listed in Table 2 are for ABC-MRT and narrow band speech systems. Values for ABC-MRT16 are not available, but are expected to be comparable in magnitude.

N	RMS Deviation	Maximum Deviation	Sample Time (MM:SS)
16	0.017	0.053	00:28
32	0.009	0.033	00:57
64	0.005	0.014	01:55
128	0.003	0.008	03:51
256	0.002	0.004	07:42
512	0.001	0.002	15:22
1,200	0.000	0.000	36:10

Table 2. Deviation of ABC-MRT results using N < 1200 trials from results with N = 1200 trials.

Saving the Acquired Speech Waveforms

The Save to File checkbox located below the Analyzer section in the ABC-MRT measurement controls (Figure 6) can be used to save the acquired MRT speech waveforms to a series of .wav files. As shown in Figure 11, the measurement uses APx User Defined Variables representing the Output Folder, Condition Number and Output File Name to specify the path of the saved files. The value of the Condition Number and Output File variables is updated automatically by the ABC-MRT measurement for each trial. The Output Folder variable is initially set to a sub-folder of the user's My Documents system folder, as shown in Figure 12, but users may change the value of this variable if they want to save the acquired waveform files to a different folder.

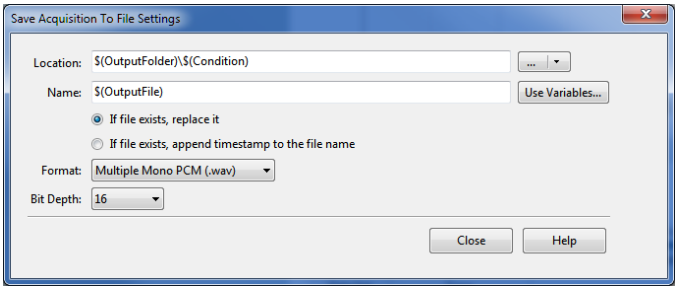


Figure 11. The Save Acquisition to File Settings dialog.

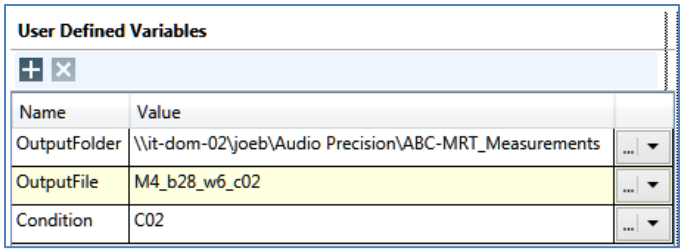


Figure 12. User defined variables used to manage the saving of ABC-MRT speech waveforms.

As shown in the example in Figure 12, the ABC-MRT measurement in APx uses essentially the same file naming scheme as that used by ITS, whereby the Talker, List number, Word number and Condition number are encoded into the file name. There is one difference: When the number of input channels is greater than 1, the string "_n" is appended to the file name, where n represents the channel number. For the example in Figure 12, the file saved for the signal acquired on Input Channel 16 would be named M3_b48_w3_c01_16.wav.

Measurement Procedures

When measuring intelligibility, it is important to use an appropriate signal level for the DUT. Basic methods for measuring the level of real speech are based on removing the silent parts of the speech signal (e.g., the gaps between words). One such method, the Speech Level measurement, defined in IEC 60268-16 [8], is included in the STI option for APx [9]. Another comparable method is the Active Speech Level described in ITU-T P.56 [10].

In addition to setting the required level, additional precautions may be required depending on whether the DUT is purely electrical or has acoustic input and/or output.

Electrical Input and Output

For purely electrical systems (e.g., evaluating a codec or parts of a communication system excluding acoustic transducers), the signal should be generated at the reference operational level of the system. It may also be advisable to perform measurements at different input levels (e.g., ±10 dB relative to the reference level) to quantify the influence of level on speech intelligibility.

Acoustic Input

For systems that use a microphone for input, a special loudspeaker known as a mouth simulator is required to generate the speech signal at the required sound pressure level. In some cases, the mouth simulator (or a comparable small, wide range powered speaker) can be used on its own. However, if the microphone interacts with the sound field around the talker (e.g., a headset microphone or a microphone built in to a face mask), a Head and Torso Simulator (HATS) with an integral mouth simulator should be used.

The mouth simulator should be equalized to have a flat frequency response (typically ±1 dB) over the frequency range of interest. The required speech level depends on the application. In IEEE standard 269 for telephony tests [11], the level of speech for normal vocal effort is specified as 89 dBSPL at the Mouth Reference Point (MRP), a point located 25 mm (1 in) in front of the lip plane of the mouth simulator. For loud and quiet vocal effort, levels of 94 and 74 dBSPL are specified. When a level is not specified, reference [8] suggests a default equivalent speech level of 60 dBA at 1 m in front of the mouth simulator (or loudspeaker). This is typically regarded as a conversational speech level.

Acoustic Output

When the output of the DUT terminates at a loudspeaker in a room, a measurement microphone with free field response placed at the listener position is typically used. In this case, once the microphone is calibrated, the DUT output can be measured and analyzed directly.

When the output from the DUT terminates in headphones or earphones, a HATS or acoustic test fixture equipped with ear simulators like the one shown in Figure 13 is typically used. In this case, the signal measured in the ear simulators should be corrected to the free field before analysis. In APx, this is easily accomplished by applying an Input EQ curve equivalent to the inverse of the free field Head Related Transfer Function (HRTF) [12].



Figure 13. Audio Precision AECM206 headphone test fixture.

ABC-MRT Measurement Applications

The ABC-MRT measurement is ideally suited to the same application as subjective MRT, namely for use in measuring the intelligibility of English speech in either indoor or outdoor spaces over a communication system. Examples of spaces include offices, vehicles, enclosures and auditoriums. Communication systems may be as simple as unamplified face to face communication, or they could involve public address systems, telephone networks, intercoms, radio links, broadcast systems or special systems for speech communication. It is also intended for use in evaluating the contribution of individual components, such as microphones, loudspeakers, headsets, codecs, encryptors, etc., to the intelligibility of speech communication systems.

ABC-MRT was a critical component used in experiments conducted at ITS to evaluate the noise resistance of candidate voice codecs for use in public safety communications networks. It helped to accelerate speech intelligibility testing that would have taken a full year, into two months.

Criteria for ABC-MRT Scores

Some examples of specified minimum intelligibility scores are as follows:

1. An international standard covering sound systems for emergency purposes (IEC 60849 [6]) specifies a minimum Common Intelligibility Score (CIS) of 0.7, which equates to an MRT score of approximately 0.77.
2. The 2007 edition of NFPA standard 1981 [7] on self-contained breathing apparatus (SCBA) for emergency purposes (e.g., for fire fighters) specified minimum MRT scores of 0.80 mechanical voice diaphragm systems and 0.85 for electronic voice communication systems.

ABC-MRT Compared to Other Speech-Related Measurements in APx

Readers may be wondering how ABC-MRT compares to other speech-related measurements available in the APx500 audio analyzer platform, such as PESQ, POLQA and STI.

PESQ and POLQA

PESQ, and its newer cousin POLQA, are measures of speech *quality*. In simple terms, intelligibility is a measure of how "understandable" a speech message is, whereas quality is more concerned with how pleasing a speech message sounds; for example, does it sound natural and is it free of distortion? The two are related, but they are fundamentally different perspectives. An example of a transmission channel with high quality but low intelligibility is a system with high bandwidth but a tendency to drop entire words, as can occur in some VoIP (voice over internet protocol) systems. The extended bandwidth provides higher quality, enabling the speech to sound much more natural, but when words are dropped, intelligibility is poor. In contrast, a lower bandwidth system without dropouts, such as a POTS (plain old telephone service) telephone line, will have lower quality but higher intelligibility.

Both PESQ and POLQA are intended to measure speech quality over telephone transmission networks, in ideal listening conditions, with very low levels of background noise. In the case of POLQA, one reference [4] states that POLQA scores obtained in noisy listening environments are not reliable, and should not be considered for any kind of speech *quality* evaluation.

Concerning PESQ, there are some documented cases of it being used for speech intelligibility testing. One such study [5] reported a correlation coefficient of 0.77 between PESQ and subjective tests like MRT, and in [2], a correlation coefficient of 0.836 between PESQ and MRT is reported for narrow band speech. For comparison, reference [3] indicates a correlation coefficient of 0.971 between ABC-MRT16 and MRT for narrow band speech.

Speech Transmission Index (STI)

STI and ABC-MRT share the common goal of providing objective estimates of speech intelligibility over communication systems. STI is an effective predictor of speech intelligibility, but there are certain conditions where it cannot be used. These include:

1. Systems containing voice codecs or noise suppression algorithms that do not transmit noise (because they will not transmit the STI signal, which is essentially broadband noise).
2. Systems with strong center clipping (also called crossover distortion).

ABC-MRT does not have the same limitations.

In addition, STI has the disadvantage that most people find it hard to rationalize how its "odd sounding" STI signal, which is based on modulated noise, can be used to predict speech intelligibility. ABC-MRT is much better in this regard, because it uses real speech and observers can easily do a "sanity check" of intelligibility scores against their own perception.

Summary

The ABC-MRT16 algorithm is an objective estimator of speech intelligibility that performs remarkably well compared to MRT, a well-established intelligibility test based on testing with human subjects. ABC-MRT can be used for any application where MRT would be used, including communication systems containing voice codecs or noise suppression. The ABC-MRT16 algorithm can also be used for any of the standard speech bandwidths from narrow band to full band. And ABC-MRT tests are much less expensive than MRT, because there is no need to use trained talkers and listeners in special facilities.

Having the ABC-MRT measurement in the APx500 software makes it very simple to use, and it provides access to all the features of the platform, including multiple audio interfaces (analog, acoustic, AES3- SP/DIF, Digital Serial, Bluetooth, PDM, HDMI, and ASIO), a built-in test sequencer and reporting engine, test limits, channel counts from 1 to 16 and a wide variety of audio quality measurements.

References

1. ANSI/ASA S3.2-2009. Method for Measuring the Intelligibility of Speech over Communication Systems. (Reaffirmed 2014).
2. Voran, S. (2013). Using Articulation Index Band Correlations to Objectively Estimate Speech Intelligibility Consistent with the Modified Rhyme Test. 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. (https://www.its.bldrdoc.gov/publications/2738.aspx)
3. Voran, S. (2017). A Multiple Bandwidth Objective Speech Intelligibility Estimator Based on Articulation Index Band Correlations and Attention. Proceedings of the IEEE 42nd International Conference on Acoustics, Speech and Signal Processing. (https://www.its.bldrdoc.gov/publications/3169.aspx)
4. Cotanis, I. (2010). Voice Service Quality Evaluation Techniques and the New Technology, POLQA (Whitepaper NT11-1037, Ascom).
5. J. Ma, Y. Hu, and P. Loizou, (2009) Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions, JASA, vol. 125, no. 5, pp. 3387–3405.
6. IEC 60849: 1998. Sound Systems for Emergency Purposes.
7. Standard on Open-Circuit Self-Contained Breathing Apparatus (SCBA) for Emergency Services, NFPA 1981, 2007 (NFPA is the National Fire Protection Association, Quincy, Massachusetts, USA).
8. Sound System Equipment Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, IEC 60268-16, 2011.
9. Measuring Speech Transmission Index with APx Audio Analyzers. Technote 128, Audio Precision, 2016.
10. Objective Measurement of Active Speech Level ITU-T P.56, 2011.
11. IEEE Standard Method for Measuring Transmission Performance of Analog and Digital Telephone Sets, Handsets, and Headsets. IEEE Std 269-2010.
12. Headphone Electroacoustic Measurements. Audio Precision Application Note. 2016.
13. S. Voran and A. Catellier, "Speech codec intelligibility testing in support of mission-critical voice applications

for LTE," Tech. Rep. TR-15-520, NTIA, 2015.
https://www.its.bldrdoc.gov/publications/2811.aspx

5750 SW Arctic Drive, Beaverton, Oregon 97005 | 503-627-0832
www.ap.com
Copyright © 2018 Audio Precision
XVIII0326105725