Comparison of vu-meter-based and rms-based calibration of speech levels

Mead C. Killion
Etymotic Research, Inc., 61 Martin Lane, Elk Grove Village, Illinois 60007
m_killion@etymotic.com

Abstract: A difference of approximately 5 dB exists between the level of spoken English determined using the ANSI standard vu-meter method compared to the common root-mean-square (rms) method. If the rms method is substituted for the present ANSI standard method for calibrating a speech audiometer, for example, the reported speech reception thresholds will improve 5 dB: Speech levels read approximately 5 dB less using rms. Similarly, the reported signal-to-noise ratio required to understand speech in a speech-spectrum noise will be 5 dB better using rms. A simple method for obtaining a close approximation to traditional calibrations using a modified rms method is given.

PACS numbers: 43.71.An, 43.66.Yw [QJF]
Date Received: March 5, 2009 **Date Accepted:** July 9, 2009

1. The historical method

In 1940, Chinn *et al.* 1940 published the paper describing the vu meter. The vu meter was fairly soon adopted for speech research and audiology, and has been the basis for the standard method for setting the calibration tone on a speech recording, which is that "…the rms sound pressure level of a 1000 Hz signal [is] adjusted so that the vu meter deflection produced by the 1000 Hz signal is equal to the average peak vu meter deflection produced by the speech signal." These instructions go back to Z24.13-1953 (American National Standard Specifications for Audiometers) and continue unchanged through ANSI Standard S3.6-1969 (1969) to the present ANSI Standard S3.6-2004 (2004). The above quote is found in Sec. 6.2.11 of the latter standard. The standard calibration method was employed in the classic NU-4 and NU-6 speech tests (Tillman *et al.* 1963; Tillman and Carhart, 1966) and the more recent MIT female-talker recordings (Rabinowitz *et al.*, 1992) used in the QuickSIN test (Killion *et al.*, 2004).

As taught by Tillman, the most accurate readings require two readings of the vu meter for a given segment of recorded material. In the first pass, the reader notes about where on the meter each peak occurs. On the second pass (or third, as needed), the reader's eyes are fixed on the approximate location of a given peak. Using this method, the exact reading for each peak can be obtained within 0.1–0.2 dB. The author has personally trained skeptical colleagues and found that their "average of frequent peaks" results agree with the author's within 0.2 dB, even though no prescription of which peaks to use was given other than "typically two to three peaks per sentence."

2. Experiment 1

Ludvigsen (1992) recently compared various measures of speech level. The vu-meter method was not included, leading the present author to exchange DAT recordings of running speech for measurement and correlation with Ludvigsen (1992). Ludvigsen (1992) later simulated in software the vu-meter ballistics and 1.4 power characteristic of the copper-oxide rectifier used in the standard vu meter.

Figure 1 shows the output of Ludvigsen's software-simulated vu meter for 30 s of running speech. The 0 dB reference value on the ordinate is the normalized rms value for this 30 s of running speech after pauses and silent periods had been removed.

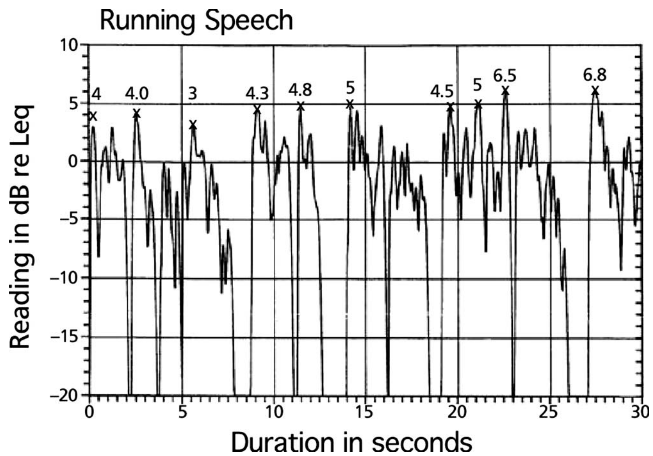

Fig. 1. Simulated (×) and actual vu-meter readings (numbers) for running speech with rms value of 0 dB.

After measuring the rms value of the same speech sample in the same way, the author made vu-meter readings as described above on a 25-year-old laboratory-reference vu meter belonging to the author. For this purpose, the analog output of the DAT recorder was adjusted so that the maximum peak stayed on the vu-meter scale, after which the individual vu-meter readings were normalized by the rms level. A 1 kHz sine wave was used to transfer the rms level measured on COOLEDIT to the vu meter.

Ludvigsen's vu-meter simulation and the author's direct reading of a vu meter with standard ballistics showed the average of the frequent peaks of the vu meter as 4.8 dB above the rms reading of the speech.

Because of the possibility of background noise in the speech sample, the removal of pauses and silent periods was done in this experiment using both eye and ear: The waveform cuts were determined from the appearance of the waveform on the computer screen, combined with listening for an error in the selection. If the background noise level is low enough and the talker does not make extraneous noises, automatic methods may be employed.

3. Experiment 2

Sixteen of the sentences recorded by Rabinowitz *et al.*, (1992) were chosen for a second comparison of the vu and rms methods. The original recordings of IEEE sentences of Rabinowitz *et al.* 1992 were made during their study of the contribution of visual cues to speech intelligibility. The audio recording of one of their female talkers was used, with permission, for the target speech in the "QuickSIN" speech-in-noise test later described by Killion *et al.* (2004).

Out of the 96 sentences on the QuickSIN CD, the 16 sentences with a 25 dB signal-to-noise ratio (SNR) (or more precisely, talker-to-babble ratio) were selected for analysis. The peak vu-meter reading and rms value of each of those sentences was determined as described above. The presence of the four-talker babble on the QuickSIN sentences with 25 dB SNR was estimated to have a negligible effect (less than 0.02 dB) on the composite level, compared to the original recordings of Rabinowitz *et al.* (1992).

The average of the peak vu-meter readings across sentences was 0.2 dB above the vu-meter reading of the calibration tone, consistent with the ANSI standard vu-meter method Rabinowitz *et al.* (1992) used for calibrating the speech levels on their recordings.

More importantly, the rms level of the 16 sentences (pauses deleted) was 4.6 dB below the calibration tone.

Taking into account the +0.2 dB difference in vu-meter-to-cal-tone reading, the difference between rms and vu-meter reading found in Experiment 2 was exactly 4.8 dB, the same as in Experiment 1.

Thus the same 4.8 dB difference obtained earlier by Ludvigsen (1992) and the author using a single sentence and a male talker was obtained again with 16 sentences and a female talker. The standard deviation of the rms and vu methods, incidentally, was nearly identical across the individual sentences.

It is worth noting that a total of 30 s of speech was used in Experiment 1, and approximately 60 s of speech was used in Experiment 2 for the 16 sentences.

4. A simple COOLEDIT method for approximating vu-meter readings

In a search for a simple COOLEDIT-based method of obtaining speech levels equivalent to vu-meter readings, the author highlighted, on the COOLEDIT waveform display, 50 ms segments of the two or three highest-amplitude portions of the speech waveform in each sentence. Using the *Analyse, Statistics* option available in COOLEDIT, the *Total RMS Power* in dB re full scale was obtained for each chosen segment. The average of those 2 or 3 dB values was examined as an approximation to a vu-meter reading by looking at the rms-to-vu-meter difference for that sentence. That method, using the sentence "A rod is used to catch pink salmon," gave an rms-to-simulated-vu difference of 4.7 dB, indicating a good approximation was obtained.

5. Discussion

Both the rms and vu-meter methods provide accurate measures of speech levels. It should be possible to use one and add or subtract 4.8 dB as a good estimate of the other. Given the difference, however, it is important to report the method used to adjust the calibration tone on any recording.

It is even more important to provide information as to the method used in an experiment, to avoid continuing confusion with reports of surprisingly good SNR performance from digital signal processing of speech in broadband noise when the apparent "improvement" has been augmented by use of a rms rather than a vu measure of speech levels. To explain, both the rms and vu-meter methods will yield essentially identical values for broadband random noise, but approximately a 5 dB difference for actual speech.

Acknowledgments

Carl Ludvigsen wrote the software to simulate the ANSI standard vu-meter characteristics in 1993, and provided the graphical comparison between rms and vu readings shown in Fig. 1. He gave his permission to use Fig. 1, but reported that his vu-meter code was written in Borland Pascal and is now buried in backup tapes. Fortunately, Lobdell and Allen (2007) recently developed a new vu-meter-simulation code that meets the ANSI standard and, more importantly, is readily available. In their paper, they illustrate the practical use of this code in obtaining new information regarding the statistics of speech.

References and links

ANSI Standard S3.6-1969 (**1969**). "Specification for audiometers."

ANSI Standard S3.6-2004 (**2004**). "Specification for audiometers."

Chinn, H. A., Gannett, D. K., and Morris, R. M. (**1940**). "A new standard volume indicator and reference level," Proc. IRE **28**, 1–8.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**, 2395–2405.

Lobdell, B. E., and Allen, J. B. (**2007**). "A model of the VU (volume-unit) meter, with speech applications," J. Acoust. Soc. Am. **121**, 279–285.

Ludvigsen, C. (**1992**). "Comparison of certain measures of speech and noise level," Scand. Audiol. **21**, 23–29.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). "Relations among different measures of speech reception in subjects using a cochlear implant," J. Acoust. Soc. Am. **92**, 1869–1881.

Tillman, T. W., Carhart, R., and Wilber, L. (**1963**)."A test for speech discrimination composed of CNC monosyllabic words (N. U. auditory test No. 4)," Report No. SAM-TDR-62, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.

Tillman, T. W., and Carhart, R. (**1966**). "An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University auditory test no. 6," Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.

Z24.13-1953 (**1953**). "American National Standard, Specifications for Audiometers."