BNAM 2021
Baltic-Nordic Acoustics Meeting
3rd-5th of May, Oslo, Norway

Effects of Intermittent Noise on Real Ear Measurements in Hearing Aid Fitting

Palle Rye, Sreeram Khaitali Narayanan and Dorte Hammershøi
Department of Electronic Systems, Aalborg Universitet, Fredrik Bajers Vej 7B, DK-9220 Aalborg, Denmark, par@es.aau.dk

When fitting a hearing aid, real ear measurements (REM) can prove a valuable tool for adjusting the gain of the hearing aid. In REM a probe microphone is inserted into the ear canal carefully placing the probe microphone tip a few millimeters from the tympanic membrane. Typically a reference microphone is present outside the ear canal near the pinna. Proper calibration before the REM measurement ensures a flat response of the probe microphone when placing the tip at the same positions as the reference microphone.
The real ear unaided gain (REUG) is measured as the level difference between the reference microphone and the probe microphone placed near the tympanic membrane. Inserting the hearing aid the real ear aided gain (REAG) can be measured similarly and the real ear insertion gain (REIG) is calculated as the difference between the REAG and the REUG. Fitting a hearing aid typically involves matching the hearing aid gain to match a REIG prescription target derived by the patients audiogram and occasionally other diagnostic data. To ensure matching the REIG to the prescription target care must be taken to ensure accuracy in the REM measurements.
Inspecting a database of thousands of clinically recorded REM a handful of anomalies were detected and the present study seek to reproduce the observed effects under controlled laboratory conditions. The study involves measuring REM using manikins as well as human subjects. A semi-automated test setup was developed to synchronise the measurement equipment and the noise disturbance in order to investigate the effects of different types of intermittent noise on REM. Preliminary results indicate that intermittent noise burst even shorter than 1s in duration can cause detrimental effects on the real ear measurement.

1 Introduction

In treating a disorder, it is important to administer the appropriate dose. Too small a dose is likely ineffective, while too large a dose could be excessively uncomfortable or even damaging. Deciding the appropriate dose typically relies on the characterization of the individual need together with a validated prescription. The merits of verifying that the dose administered follow the selected prescription seem indisputable. Treating hearing disorders with hearing aids should be no different in that regard unless it might be argued that some modern hearing aids are more like a consumer device than clinical treatment. In clinical treatment, there are plenty of prescription target gain paradigms to choose between, ranging from the widely accepted NAL-NL2 [1], the DSL v5.0 [2], hearing-aid manufacturer proprietary targets, or present-day experimental paradigms, e.g., as proposed by Sanchez-Lopez et al. [3].

To verify how close the actual gain is to the prescribed target Real Ear Measurement (REM) using probe tube microphone is recommended by several organizations, e.g., the American Academy of Audiology, the British Society of Audiology, and British Academy of Audiology [4]. Further compelling arguments for using REM in the progress toward an evidence-based approach to hearing aid fitting are given in [5].

The Better hEAring Rehabilitation (BEAR) project aims to improve hearing aid treatment in Denmark. Progressing toward evidence-based approaches is one of the mainstays of the project. In the first part of the project, both subjective and objective data were collected from nearly 2000 patients reflecting current clinical practice in hearing aid treatment in Denmark, as described by Wolff [6]. The database contains extensive data from a diverse group of patients, both new and experienced users. During the analysis of data, different types of anomalies were detected in both REUG and REIG.

The first and most prominent anomaly detected was caused by a subset of unilateral and CROS fittings not marked as such. The untreated ear was measured to avoid confusion over missing data, resulting in near 0 dB REIG and, consequently, extreme deviations to the NAL-NL2 fitting target prescription. Correction of the missing markings allows the measurements to be appropriately analyzed in separate studies.

With the amount of data measured, such few outliers did not significantly impact using traditional statistical methods. However, when analyzing data by clustering according to similarity, e.g., using k-means[7], even a modest number of anomalies with similar traits may result in a cluster of its own. An interesting example is shown in Figure 1, where k-means were used to cluster the clinically measured REUG into four groups. Cluster 1-3 might be explained by physical traits causing clustering according to the center frequency and shape of ear canal resonance. However, the average response in Cluster 4 exhibits a low frequency offset around 5dB that is difficult to explain from an acoustics point of view. When inspected individually, many of the Cluster 4 responses does not appear strikingly different from a reasonable REUG, and it is not straightforward to set an exact criterion when a measured REUG would need to be re-measured.

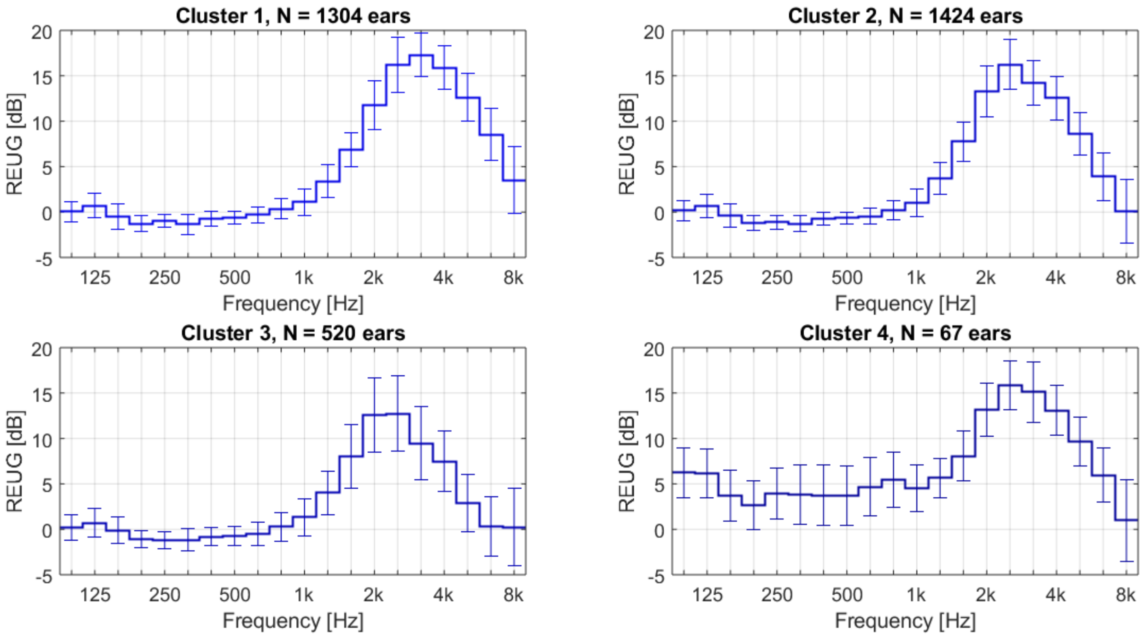

Figure 1 Mean and standard deviations of clinically measured REUG when sorted into 4 groups by similarity using k-means. Cluster 4 exhibit an unrealistic offset in the frequency region 100 Hz - 1 kHz.

A working hypothesis was that the inexplicable low frequency REUG offset could be caused by short duration disturbing noise of intermittent nature that would not be detected as interfering with the measurements. Given the current recommendation for using REM as a standard practice to provide better hearing rehabilitation, it is interesting to understand to what extent such noise could affect REM. To reproduce REUG responses similar to the Cluster 4 identified in the clinical database a set of laboratory experiments were performed.

2 Methods and materials

The laboratory measurements were performed in an anechoic room with a loudspeaker array used to generate intermittent disturbing noise. An Interacoustics Affinity 2.0 measurement system with REM module REM440 was set up to be interchangeable with any of the two identical measurement systems used for the data collection at the two separate clinical sites. A twenty seconds long ISTS[8] signal was used as the input signal to measure REUG.

Short term repeatability of the laboratory setup without added disturbing noise was evaluated by running 32 consecutive measurements without repositioning the probe microphone. The same types of measurements estimated long term repeatability, but with repositioning of the probe microphone. When using manikins, it was notably easier to reposition the probe microphone by bump-and-pull to nearly the same position between repetitions so the results may suggest a higher accuracy than for human subjects. For the manikins with ear couplers the probe was placed as deep as possible.

A Matlab script was constructed for controlling the intermittent disturbing noise parameters and synchronization between the REM measurement on Affinity and the disturbing noise playback through RME MADIfaceXT and ADI-8. The speakers used for playback of intermittent noise were part of a more extensive speaker array setup; however, a circular ø4m arrangement with 8 Genelec 8020 speakers at ear height in the horizontal plane was used to play pink noise as intermittent noise The test subject or manikin was placed in the center of the circular array with the Affinity measurement system placed on a small table in front, as shown in Figure 2 (a).

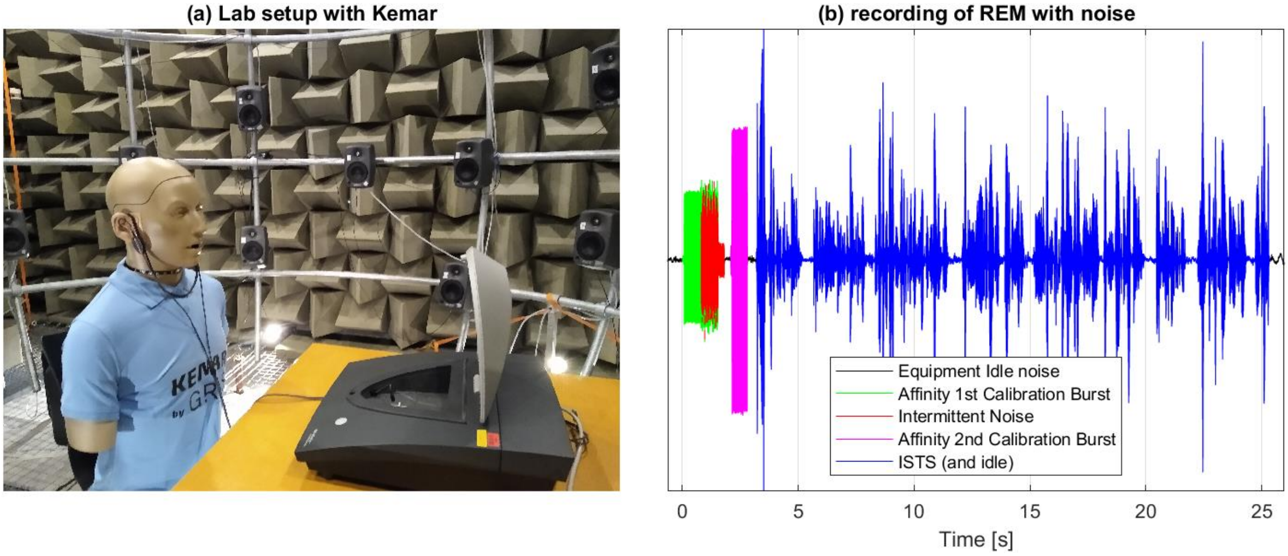

Figure 2(a) Test setup in an anechoic chamber using Kemar manikin and (b) recorded example of waveform showing typical timing from starting the REM through the first phase of Affinity calibration

In order to estimate the noise levels required to influence the REM, an initial exploration used diffuse field pink noise continuously playing throughout the measurement. Sound pressure levels below 45 dBA were found to cause less than 0.5 dB deviation, whereas levels above 65 dBA would cause the ISTS signal from Affinity to be practically muted. Although the resulting REIGs exhibited deviations exceeding 5 dB above 4 kHz, the frequency region below 4 kHz was within +/-2 dB of the noise-free REIG. Thus, it effectively uses the diffuse field pink noise to measure the transfer function between reference and probe tube microphone.

For studying the effects of intermittent noise, the experimental parameters varied were noise level, duration, the direction of incidence in the horizontal plane, and the delay between starting the REM measurement and beginning playback of the noise. A set of reference measurements were performed without noise. The probe microphones were not repositioned between measurements unless so stated.

For each parameter combination, a set of measurements for each ear were recorded and stored in a separate XML file containing both left and right ear responses and a comment describing the actual parameter settings. Varying the parameters outlined in Table 1, different series of tests were run using two human subjects and three manikins: G.R.A.S. 45BC KEMAR, Brüel & Kjær HATS Type 4128, and AHead Simulations CARL, adding to a total of 1098 measurements.

Table 1: The list of parameters varied for each measurement condition concerning the type of head used.

Head / Noise	G.R.A.S. 45BC KEMAR	Brüel & Kjær HATS Type 4128	AHead Simulations CARL	Human subject A	Human subject B
Level (dB SPL)	35, 45 , 55	35, 45 , 55	35, 45 , 55	35, 45 , 55	35, 45 , 55
Duration (ms)	250, 500, 1000	250, 500, 1000	250, 500, 1000	250, 500, 1000	250, 500, 1000
Horizontal Incidence	0°, 45°, 90°, 135°, 180°, 225°, 270°	0°, 45°, 90°, 180°	0°, 45°, 90°, 180°	0°, 45°, 90°, 180°	0°, 45°, 90°, 180°
Delay (s)	1, 1.5, 2, 3	3	0.5, 3, 4, 5	3	3

The internal frequency resolution of the measured amplitude FFT spectra was 1024 at a sampling frequency of 44.1 kHz [9]. The XML files contain 231 measured values with a spacing of 43 Hz in the frequency region approximately 86 Hz – 9976 Hz. Those endpoints are labeled 100 Hz and 10000 Hz; respectively, it is unclear whether they are relabeled points or a result of interpolation or averaging over unsaved internal FFT points. The 1024 points were reconstructed by replicating the endpoints up- and downwards. The third-octave band levels were calculated by converting the spectra from dB to linear gain values, upsampled to 2048 points using spline, taking the RMS of the FFT bins within each band, and finally converted back to dB.

The delay parameter controls the time the noise is added to the Affinity measurement cycle. Shorter delays below 3s make the noise burst coincide with the Affinity calibration bursts, and longer delays target the ISTS measurement cycle, as exemplified in Figure 2 (b).

Finally, a set of measurements featuring a single but less well-controlled experimental parameter were performed using the Kemar Manikin. The probe tube was touched lightly with an eyeglass temple tip during the ISTS part of the measurement. The purpose was to simulate a kind of structure-borne intermittent noise, which might arise from physical interaction with the probe tube during measurement.

3 Results

The short term and long term repeatability using the Affinity test system to measure the REM using manikins under anechoic conditions were below 1 dB up to 2.5 kHz, as illustrated by the typical examples in Figure 3. Under anechoic conditions using manikins, the Affinity REM measurement system's repeatability is better than 1 dB up to 2.5 kHz. Earlier studies have found the short term test-retest reliability for probe tube REUG to be between 1 dB and 1.6 dB between 1 kHz and 3 kHz [10].

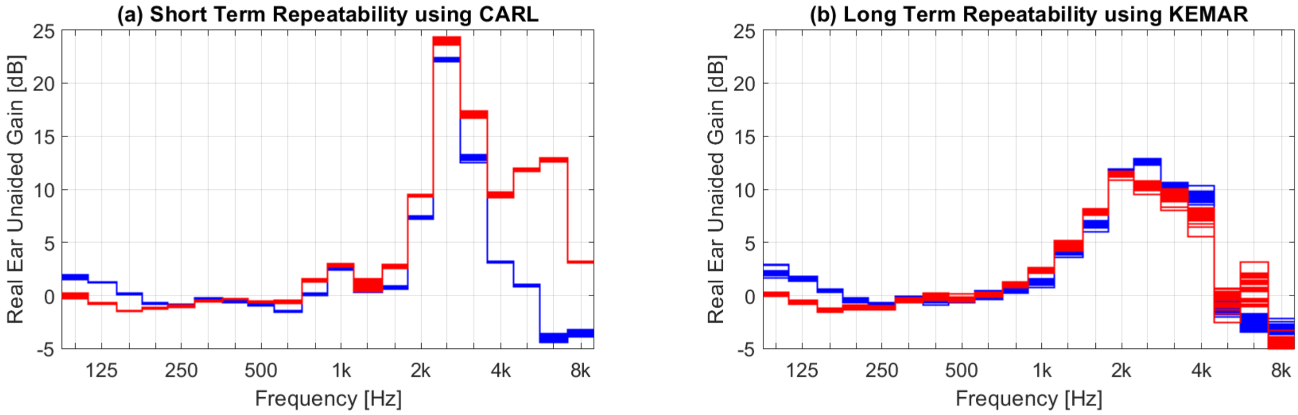

Figure 3 Typical examples showing 32 measurement repetitions for left (Blue) and right (Red) ears for (a) the Carl Manikin examining short term repeatability and (b) the KEMAR long term repeatability.

In order to compare the results of intermittent noise across different head types, the delta to the reference measurement for each head is used. Using the k-means algorithm to cluster the delta values from the complete lab dataset into four groups yields the result shown in Figure 4.

While Cluster 2 does exhibit a low-frequency offset anomaly similar to the clinical recorded REUG, it has shown nearly 20 dB higher levels at 6 to 8 kHz, which is not apparent in the clinical REUG. Upon inspection, the cluster comprises experimental parameters at high levels that coincide with the second calibration burst of the Affinity measurement system, effectively muting the ISTS, which is highly unlikely to go unnoticed in the clinical scenario.

Results from the final set of data collected in this study touching the probe tube intermittently during the ISTS phase of the measurement are shown in Figure 5. A combination of low frequency offset and only moderate disturbance at higher frequencies is observed.

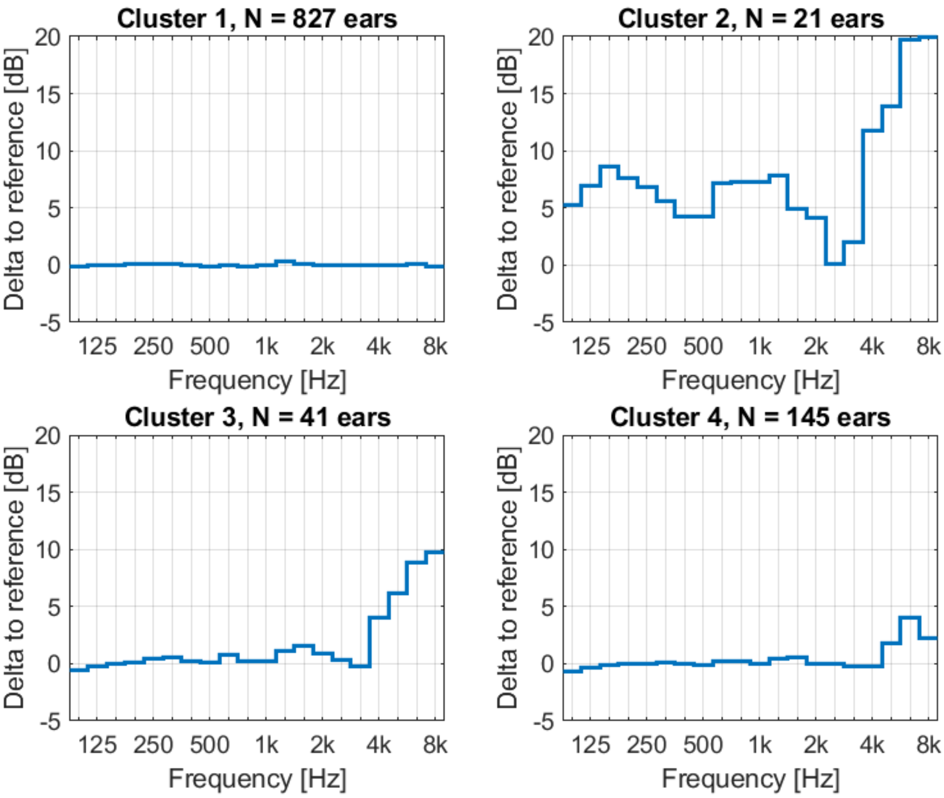

Figure 4k-means sorting of delta between noisy REUG and their individual noise-free reference REUG

The shape of the gain curve measured in the lab by intermittent touching of the probe tube follows a similar trend to that observed in cluster 4 of the clinical data. A multivariate ANOVA of the one-third octave band data showed no significant difference between clinical and lab measurements for frequency bands, 125 Hz, 160 Hz, 200 Hz, 250 Hz, 400 Hz, 1000 Hz, and 2000 Hz. It implies that the low frequency offset observed in the clinical data could have resulted from intermittent touching of the probe tube during the measurement. The touch resulting in the observed effect could have been introduced due to the movement of the head, hair or another object. The significant differences at higher frequencies was expected due to the controlled environment and homogenous head used in the lab setting, contrary to the clinical setting.

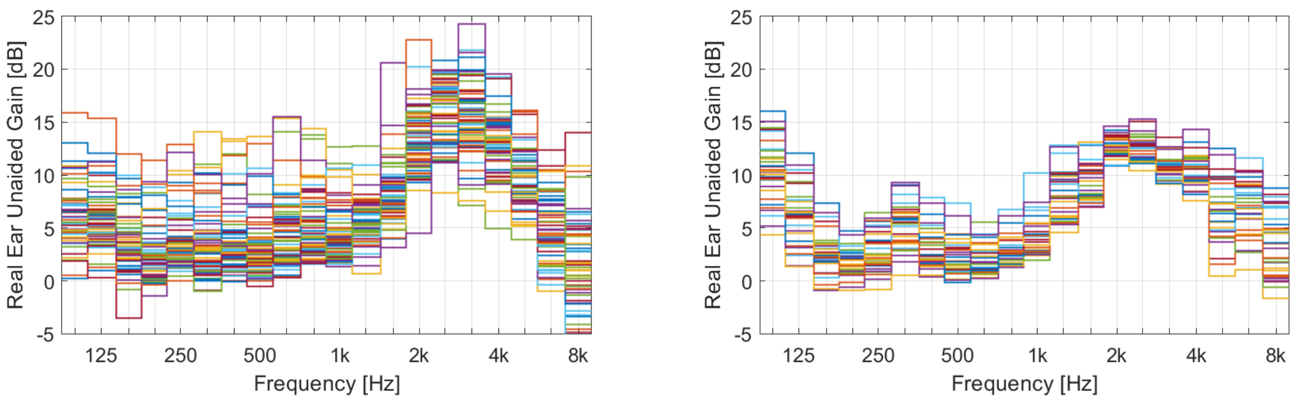

Figure 5 Comparison of Clinical data from Cluster 4 (left) with 32 repetitions with touching the probe tube during ISTS part of measurement (right).

4 Conclusion

Overall the measurement system remains remarkably robust in the intermittent noise scenarios tested in this study. Touching the probe tube during measurement was the intermittent noise scenario found to be in correspondence with the anomalies detected in the clinical data.

5 Acknowledgments

This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project), Oticon, GN Resound, Widex, and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force, Aalborg, Odense and Copenhagen University Hospitals). The funding and collaboration of all partners are sincerely acknowledged.

References

[1] G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer, "The NAL-NL2 prescription procedure," *Audiol. Res.*, vol. 1, pp. 2010–2012, 2011.

[2] S. Scollie *et al.*, "The Desired Sensation Level Multistage Input/Output Algorithm Work that Leads to the 2005 Algorithm," *Trends Amplif.*, vol. 9, no. 4, pp. 159–197, 2005.

[3] R. Sanchez-Lopez, M. Fereczkowski, S. Santurette, T. Dau, and T. Neher, "Towards Auditory Profile-Based Hearing-Aid Fitting: Fitting Rationale and Pilot Evaluation," *Audiol. Res.*, vol. 11, no. 1, pp. 10–21, 2021.

[4] British Society of Audiology and British Academy of Audiology, "Guidance on the use of real ear measurement to verify the fitting of digital signal processing hearing aids," 2007. [Online]. Available: http://www.thebsa.org.uk/docs/RecPro/REM.pdf.

[5] H. G. Mueller, "Perspective : real ear verification of hearing aid gain and output," *GMS Zeitschrift für Audiol. (Audiological Acoust.)*, vol. 2, pp. 1–11, 2020.

[6] A. Wolff, *Health-Related Quality of Life Following Hearing Aid Treatment a large Cohort study*. Aalborg Universitetsforlag, 2019.

[7] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, pp. 281–297, 1967.

[8] I. Holube, S. Fredelake, M. Vlaming, and B. Kollmeier, "Development and analysis of an international speech test signal (ISTS)," *Int. J. Audiol.*, vol. 49, no. 12, pp. 891–903, 2010.

[9] Interacoustics, "Instructions for Use – Part 1 Affinity2.0/ Equinox2.0." [Online]. Available: https://www.interacoustics.com/download/affinity/affinity-manuals/106-instructions-for-use-affinity-part1.

[10] D. B. Hawkins, E. A. Alvarez, and J. R. Houlihan, "Reliability of three types of probe tube microphone measurements," *Hear. Instruments*, vol. 42, pp. 14–16, 1991.