Distribution of short-term rms levels in conversational speech

Robyn M. Cox, Joseph S. Matesich, and Jeffrey N. Moore

Citation: The Journal of the Acoustical Society of America **84**, 1100 (1988); doi: 10.1121/1.396697
View online: https://doi.org/10.1121/1.396697
View Table of Contents: https://asa.scitation.org/toc/jas/84/3
Published by the Acoustical Society of America

Distribution of short-term rms levels in conversational speech

Robyn M. Cox
Department of Audiology and Speech Pathology, Memphis State University, Memphis, Tennessee 38105, and Veteran's Administration Medical Center, Memphis, Tennessee 38105

Joseph S. Matesich and Jeffrey N. Moore
Department of Audiology and Speech Pathology, Memphis State University, Memphis, Tennessee 38105

(Received 21 December 1987; accepted for publication 2 June 1988)

For 30 male and 30 female talkers, distributions of short-term rms speech levels, relative to the corresponding long-term rms levels, were determined in each of eight $\frac{1}{3}$-oct bands for six short-term measurement intervals. Consecutive, Hanning-windowed, 20-ms time records were combined to produce nominal measurement intervals ranging from 20 to 120 ms. For each measurement interval, mean distributions of short-term rms speech levels relative to long-term levels were very similar for male and female talkers, and intertalker differences were small, especially for short-term amplitudes above the median level. The distribution of short-term rms speech levels relative to long-term rms speech levels varied with measurement interval for the short-term measurements. The effect of measurement interval was least for the highest amplitude speech levels and increased as speech levels decreased. For short-term amplitudes above the median level, the effect of measurement interval was greater in higher frequency regions, whereas for short-term amplitudes below the median level, measurement interval had the greatest effect on the lower frequency bands. These data may facilitate comparisons among investigations using different measurement intervals. In addition, they have implications for amplification strategies.

PACS numbers: 43.66.Ts, 43.71.Ky, 43.72.Ar [NFV]

INTRODUCTION

In hearing aid research, speech is often specified in terms of the amplitude distribution of rms levels measured using short integration times, typically 1 s or less. Investigators have employed a variety of time constants for the short-term rms measurements. Numerous workers have relied on the classic data of Dunn and White (1940), which used an integration time of 125 ms and averaged across 11 talkers. Other investigators have used integration times ranging from 1 s to a few ms and have typically measured the distribution for the particular talker(s) used in the study.

Some of these investigations have produced apparently contradictory outcomes. For example, one study (Skinner, 1980) suggests that an appropriate amplification strategy could allow the 25% distribution level (i.e., the short-term rms level exceeded in 25% of samples) to approach the listener's discomfort thresholds, whereas another study (De Gennaro *et al.*, 1986) suggests that only the 1% distribution levels should be allowed to approach the discomfort thresholds.

However, because of differences in integration times used to determine short-term rms amplitude distributions and the different talkers for whom these distributions have been determined, it may not be appropriate to directly compare the results of diverse investigations. The measurement interval used to obtain the short-term rms speech levels is an important factor in the obtained distribution. In addition, it is possible that individual talkers may vary substantially in the range of short-term speech levels they produce. Hence, results obtained with one talker may not be generalizable to other talkers.

An additional consideration involves the relationship between short-term amplitude distributions and long-term rms levels in speech. Available data on the short-term/long-term rms ratios in speech and the effects of integration times are limited to small numbers of talkers and measurement intervals. Because of this, it is problematic to compare amplification strategies that are based on a $\frac{1}{3}$-oct band long-term rms speech spectrum (e.g., Byrne and Dillon, 1986) with those that are based on a short-term amplitude distribution (e.g., Barfod, 1972).

The investigation reported here was undertaken to generate a more complete description of the distribution of short-term rms speech levels and its relationship to the corresponding long-term rms levels. It was anticipated that these data would be useful in comparing results from diverse investigations, and, in addition, may have implications for amplification strategies. The following questions were of particular interest. (1) Are there noteworthy differences among individual talkers or between male and female talkers in the distribution of short-term rms speech levels? (2) How does the distribution of short-term rms speech levels vary as a function of measurement interval in the range 20–120 ms?

I. METHOD

A. Recording the speech samples

Two-min samples of conversationally produced continuous speech were recorded from 30 male and 30 female native talkers of American English. Recordings were made in a double-walled sound-treated room with a mean reverberation time (RT60) of 52 ms. The microphone was located 30 cm from the talker's mouth. The frequency response of the recording system was flat, ± 1 dB, from 50 Hz to 16 kHz. *Post hoc* analyses revealed that the mean integrated overall sound-pressure level of speech at 1 m was 61 dB (s.d. = 3.6 dB) and 59 dB (s.d. = 3.2 dB) for the male and female talkers, respectively. These values are almost identical to the corresponding data reported by Pearsons *et al.* (1977). The long-term rms ambient noise level ranged from 53 dB (at 250 Hz) to 35 dB (at 6.3 kHz) below the long-term rms speech level.

One minute from the middle of each speech sample and 1 min of recorded ambient noise were digitized with 12-bit resolution and a 9.0-kHz audio bandwidth.

B. Analyzing the speech samples

Long- and short-term rms levels were measured in each of eight nominally $\frac{1}{3}$-oct frequency bands, approximately centered at 0.25, 0.5, 0.8, 1.0, 1.6, 2.5, 4.0, and 6.3 kHz. One-third oct band levels were calculated from narrow-band spectrum analyses determined using a 400-line FFT analyzer (Hewlett–Packard, model 3561A).

The long-term rms $\frac{1}{3}$-oct band levels for each sample were derived from a spectrum based on one thousand 20-ms samples. Short-term levels for nominal 20-, 40-, 60-, 80-, 100-, and 120-ms intervals were computed for each of eight $\frac{1}{3}$-oct bands from rms amplitude spectra. Each rms spectrum was obtained from FFTs averaged on a power basis for the appropriate number (1–6) of contiguous data blocks. Accuracy of the measurement of short-term levels was checked using amplitude-modulated sinusoids with known peak-to-rms levels. Differences between measured and predicted short-term rms levels were < 0.25 dB.

To preserve accuracy in the spectral analyses of the speech samples, each 20-ms data block was weighted using a Hanning window prior to FFT analysis. Because the Hanning window forces the amplitude level to 0 at the beginning and end of the 20-ms time record, only the middle 7.5 ms of the record are actually observed at full amplitude. The rms level seen in this time period becomes the estimate of level for the entire 20 ms. Thus events that are completely encompassed within the initial or final 6.25 ms of a time record will not influence the measured level of that sample. In the worst case, a 12.5-ms event that happened to be perfectly centered between two contiguous data blocks would be missed entirely. Because many randomly selected speech samples were obtained from each talker with the results averaged across talkers and because sounds as brief as 12 ms are very rare in normal speech (Umeda, 1977), it was postulated that the windowing procedure would not exert a significant effect on the measured distribution levels. To evaluate this issue, the results were compared with analagous data obtained by Dunn and White (1940) using a different measurement procedure (see Sec. II).

For each talker, 48 distributions of short-term rms levels were generated (six intervals × eight bands), each comprising 500 pseudorandom samples. In each $\frac{1}{3}$-oct band, the levels exceeded in 1%, 10%, 20%, 50%, 70%, and 90% of the samples were determined. Analogous data were also obtained for the recorded ambient noise.

II. RESULTS AND DISCUSSION

Figure 1 illustrates the distribution of short-term rms levels for the 120-ms measurement interval. All data are plotted relative to long-term rms levels of speech in the corresponding $\frac{1}{3}$-oct band. The figure gives mean values for male talkers (solid line) and female talkers (dashed line). The lowest line shows the 1% values for the ambient noise in the recording room. These data may be converted to sound-pressure levels by adding the long-term rms levels of conversational speech to each $\frac{1}{3}$-oct band. In the eight $\frac{1}{3}$-oct bands used in this investigation, the mean long-term integrated levels at 1 m were: 52, 54.5, 46, 45, 42.5, 39, 37, and 35 dB SPL for male talkers and 50, 51.5, 47, 45, 40, 36.5, 35.5, and 35.5 dB SPL for female talkers. When normalized for overall level, male and female long-term average spectra were very similar in the 0.25 to 6.3-kHz range (see Cox and Moore, 1988, for more details).

Figure 1 shows that, when talkers were normalized in terms of overall level, the mean short-term/long-term rms ratios were similar for male and female talkers. This was also true for all other measurement intervals. For the 1%, 10%, 20%, and 50% amplitude distribution contours, the male and female data were almost identical. However, for the 70% and 90% contours, mean differences up to about 5 dB were sometimes found. Standard deviations were on the order of 1–2 dB for the three upper contours and 4–6 dB for the three lower contours, indicating that intertalker differences within each sex are rather small for speech amplitudes above the median level. Overall, these results suggest that intertalker differences in short-term/long-term rms ratios are probably not an important consideration in hearing aid fitting strategies. In addition, data relating hearing-aided performance to short-term/long-term rms ratios for a single

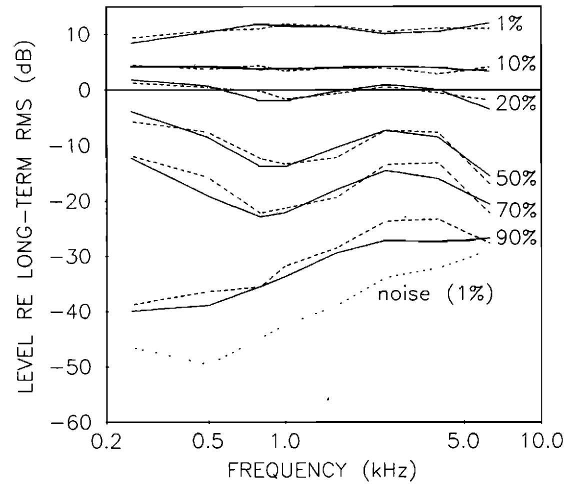

FIG. 1. Distribution of short-term rms levels relative to long-term rms level in eight $\frac{1}{3}$-oct bands (levels labeled "1%" were exceeded in 1% of samples, etc.). Measurement interval was 120 ms. Data are plotted at the $\frac{1}{3}$-oct band center frequency, and adjacent bands are joined by straight lines. Solid lines show mean data for male talkers ($N = 30$). Dashed lines show mean data for female talkers ($N = 30$). The lowest line shows the rms level of ambient noise exceeded in 1% of samples, plotted relative to the long-term rms speech level.

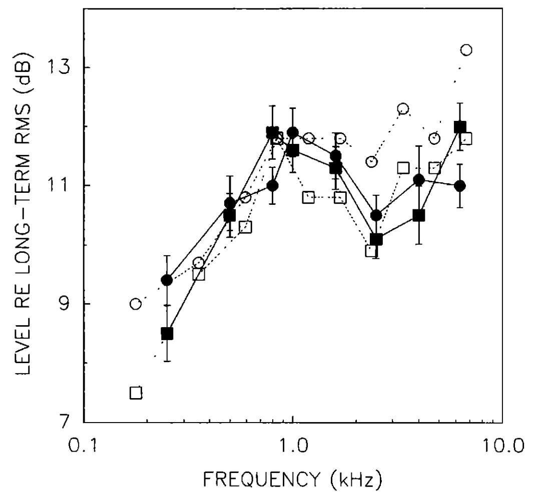

FIG. 2. Comparison of results of the present study with corresponding data reported by Dunn and White (1940). Data shown from the present study are mean rms levels, measured in 120-ms intervals, that were exceeded in 1% of samples for male talkers (closed squares, solid line) and female talkers (closed circles, solid line). Values determined from the Dunn and White data are mean rms levels, measured in 125-ms intervals, exceeded in 1% of samples for male talkers (open squares, dotted line) and female talkers (open circles, dotted line). All data are plotted relative to the corresponding long-term rms speech levels.

talker can probably validly be generalized to other talkers. However, it should be noted that, for low-amplitude distribution levels, mean differences of several decibels may be found between male and female talkers and intertalker differences may be significant.

Figure 2 shows a comparison between some of the data reported by Dunn and White (1940) and the corresponding data obtained in the present investigation. For each study, the data depicted are the short-term/long-term rms ratios for levels exceeded 1% of the time in 120-ms intervals (this study) or 125-ms intervals (Dunn and White). The Dunn and White data have been estimated by comparison of their Figs. 3 and 10 (males, $N = 6$) and Figs. 4 and 10 (females, $N = 5$). These data form the basis for the often-quoted 12-dB difference between long-term rms and 1% short-term rms levels. Data from the present study give mean values for male and female talkers and the 95% confidence intervals. Male talker data from the present study are very similar at all frequencies to those reported by Dunn and White. For female talkers, the two studies are in excellent agreement through 2500 Hz. Above this frequency, the Dunn and White 1% levels are 1–2 dB higher than the corresponding levels obtained in the present study. The excellent overall agreement between the two studies suggests that the use of the Hanning window in data collection did not significantly affect the obtained distribution levels. Because the high-frequency disparity is present for the female talkers only, it seems unlikely that this difference in results is due to the measurement procedure. However, this issue should be kept in mind if these data are compared to other data obtained using a different measurement procedure.

Figure 3(a)–(e) illustrates the effect of short-term measurement interval on the short-term/long-term rms ratios. Male and female talker data are averaged in these figures. Each figure illustrates short-term/long-term rms ratios for the five shorter intervals expressed relative to the corresponding ratios for the 120-ms interval. Positive and negative data values indicate that the short-term amplitudes for the interval in question were higher and lower, respectively, than the corresponding amplitudes measured for the 120-ms interval. Data are not reported for the 90% amplitude contour because this contour could not clearly be distinguished from the ambient noise for the shortest measurement intervals.

Figure 3(a) shows that measurement interval had a small but systematic effect on the short-term/long-term rms ratio for the highest amplitude speech events: The 1% level increased by 0.5–1.5 dB as measurement intervals decreased from 120 to 20 ms. In general, the effect was greater for higher frequency bands. The 1% level contour has been used in several hearing aid related applications, including standardized hearing aid gain measurement and prescription of hearing aid saturation level. The data shown in Fig. 3(a) indicate that measurement interval (in the 20 to 120 ms range) is probably not an important consideration in these applications.

The pattern of changes for the 10% contour [Fig. 3(b)] is similar to that seen for the 20% contour [Fig. 3(c)]. Both figures indicate that short-term amplitudes decreased as measurement interval decreased and that the effect was greatest in the 6.3-kHz band. It is somewhat surprising to note, however, that the effect of measurement interval was greater for the midfrequency bands (0.8–1.6 kHz) than in the 2.5- to 4.0-kHz region.

Figure 3(b) and (c) indicates that the effect of measurement interval on the distribution of speech levels above the median level is small but not inconsequential. For example, Fig. 1 shows that, when speech is measured in 120-ms intervals, roughly 20% of speech samples in all frequency bands are above the long-term rms levels. Thus one might conclude that, if aided thresholds are made equal to long-term rms speech levels in each $\frac{1}{3}$-oct band, speech would be audible about 20% of the time. Figure 3(c), however, shows that, when speech is measured in 20-ms intervals, the 20% contour is considerably lower than that seen for 120-ms measurements, especially in the important midfrequency bands. As a result, only 10%–15% of samples are above the corresponding long-term rms level at most measured frequencies. Overall, these data indicate that, if an amplification strategy attempts to shape and amplify the speech spectrum so that a specific proportion of speech is audible (see, for example, DeGennaro *et al.*, 1986), the prescribed frequency/gain function will depend, to some extent, on the measurement interval used in speech analysis.

Figure 3(d) and (e) reveals a substantial drop in short-term rms levels for the 50% and 70% contours, respectively, as measurement interval decreased. These figures indicate a reversal of the trend seen in Fig. 3(a)–(c) in that the effects of measurement interval on the short-term/long-term rms ratio were greater in the lower frequency bands.

As expected, lengthening the measurement interval resulted in a more narrow range of short-term speech ampli-

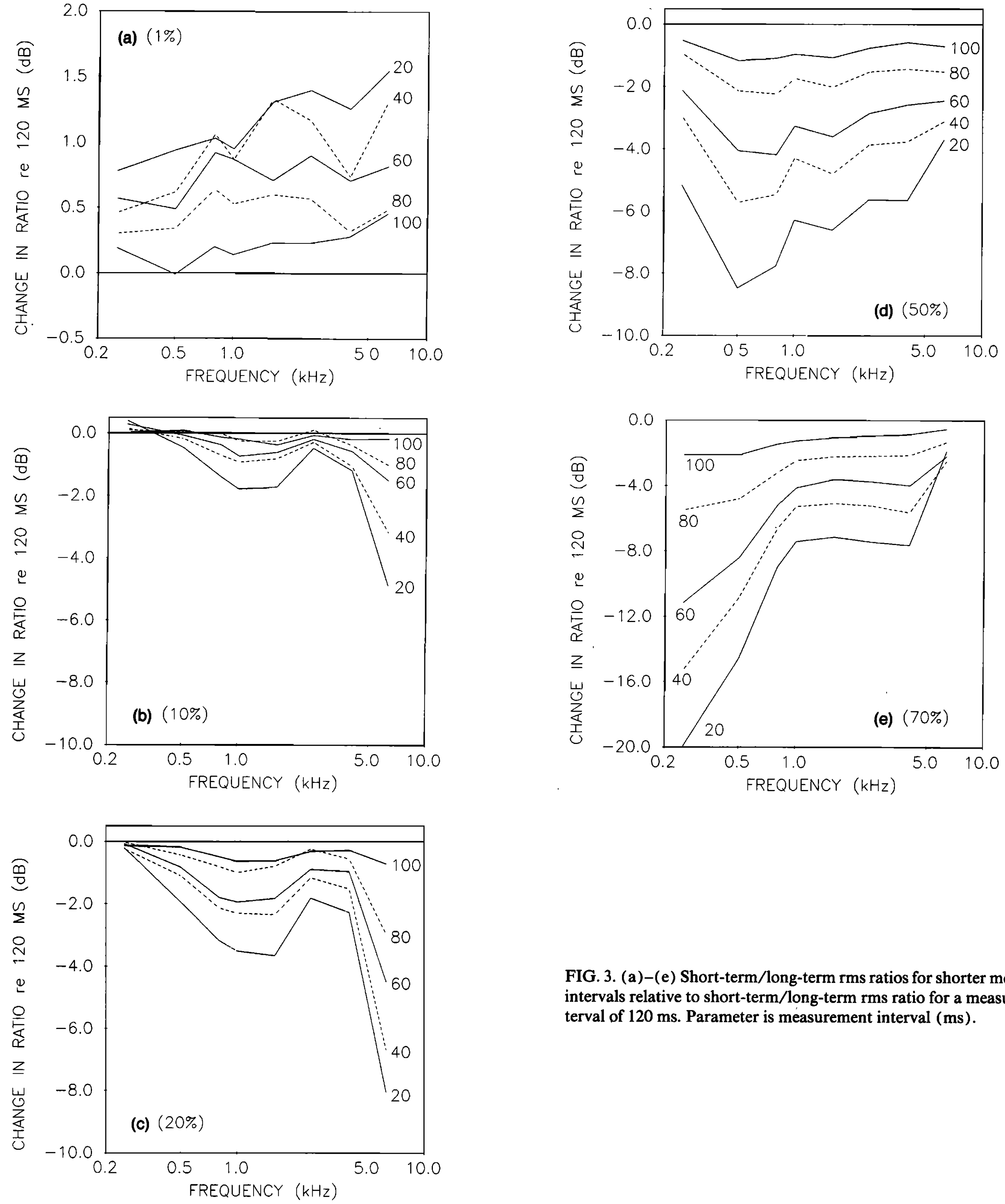

FIG. 3. (a)–(e) Short-term/long-term rms ratios for shorter measurement intervals relative to short-term/long-term rms ratio for a measurement interval of 120 ms. Parameter is measurement interval (ms).

tude levels. When short-term levels were measured in 20-ms intervals, the 1%–70% range of amplitude levels was 42.5 dB at 250 Hz and 34 dB at 4 kHz. When the measurement interval was increased to 120 ms, the corresponding ranges were 21 and 25.5 dB, respectively. This narrowed range of short-term amplitude levels was almost entirely due to a rise in the levels of contours at and below 10% for longer measurement intervals.

ACKNOWLEDGMENTS

This letter was supported in part by VA Rehabilitation Research and Development funds and also in part by the Center for Research and Innovative Services for the Communicatively Impaired, Memphis State University, Memphis, Tennessee.

Barfod, J. (1972). "Investigations on the Optimum Corrective Frequency Response for High Tone Hearing Loss," The Acoustics Laboratory, Technical University of Denmark, City, Denmark, Rep. No. 4.

Byrne, D., and Dillon, H. (1986). "The National Acoustic Laboratories' (NAL) New Procedure for Selecting the Gain and Frequency Response of a Hearing Aid," Ear Hear. 7, 257–265.

Cox, R. M., and Moore, J. N. (1988). "Composite Speech Spectrum for Hearing Aid Gain Prescriptions," J. Speech Hear. Res. 31, 102–107.

DeGennaro, S., Braida, L. D., and Durlach, N.I. (1986). "Multichannel Syllabic Compression for Severly Impaired Listeners," J. Rehabil. Res. Dev. 23, 17–24.

Dunn, H. K., and White, S. D. (1940). "Statistical Measurements on Conversational Speech," J. Acoust. Soc. Am. 11, 278–288.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (1977). "Speech Levels in Various Noise Environments," EPA Rep. No. 600/1-77-025, Environmental Protection Agency, Washington, DC.

Umeda, N. (1977). "Consonant Duration in American English," J. Acoust. Soc. Am. 61, 846–858.

Skinner, M. W. (1980). "Speech Intelligibility in Noise-Induced Hearing Loss: Effects of High-Frequency Compensation," J. Acoust. Soc. Am. 67, 306–317.

A note on reflection and transmission of waves at a boundary between two viscoelastic media

H. W. Jones and H. W. Kwan
Engineering Physics Department, Technical University of Nova Scotia, Halifax, Nova Scotia B3J 2X4, Canada

E. M. Yeatman
Electronic and Electrical Engineering Department, Imperial College, London, United Kingdom

(Received 2 November 1987; accepted for publication 1 June 1988)

This letter notes errors in the previous treatment of the problem of the reflection and transmission of acoustical waves at a boundary between two viscoelastic media. The validity of the corrections that are given is confirmed by the recovery of the solid/solid interface equations of Kolsky and by some experimental evidence.

PACS numbers: 43.35.Mr, 43.20.Fn

INTRODUCTION

H. F. Cooper[1] published an analysis of the situation that arises when an acoustical wave impinges at an angle to the normal on the plane defined by the boundary between two semi-infinite layers of viscoelastic material. The publication contains errors in the equations that are presented, and it is the purpose of this letter to set the record straight. It should be noted that the errors do not arise from the physical treatment of the problem but might well be due to printing or a similar cause. It appeared to us to be necessary in listing the correct equations to give some demonstration of their validity. This we have done in two ways: first, by the recovery of the equations for the solid/solid boundary (cf. Kolsky[2]) and, second, by recourse to experimental data. We hope that these demonstrations are sufficiently convincing.

I. STATEMENT OF THE PROBLEM

Figure 1 shows the geometry of the interface with the incident, reflected, transmitted, and mode converted sound waves. Cooper obtained equations relating the relative amplitudes of the various components of the acoustical velocity, displacement, and stress at the boundary and conserved these in the usual way. From this treatment he derived the following equations:

$$\begin{pmatrix} u_{lm} \\ v_{lm} \\ \sigma_{xlm} \\ \sigma_{ylm} \\ \tau_{lm} \end{pmatrix} = \begin{pmatrix} ik_{m1}\sin\zeta_{lm1} & ik_{m2}\epsilon_m\cos\zeta_{lm2} \\ \epsilon_m ik_{m1}\cos\zeta_{lm1} & -ik_{m2}\sin\zeta_{lm2} \\ -k_{m1}^2(\lambda_m+2\mu_m\sin^2\zeta_{lm1}) & -\mu_m k_{m2}^2\epsilon_m\sin 2\zeta_{lm2} \\ -k_{m1}^2(\lambda_m+2\mu_m\cos^2\zeta_{lm1}) & \mu_m k_{m2}^2\epsilon_m\sin 2\zeta_{lm2} \\ -\epsilon_m\mu_m k_{m1}^2\sin 2\zeta_{lm1} & -\mu_m k_{m2}^2\cos 2\zeta_{lm2} \end{pmatrix} \begin{pmatrix} \psi_{lm1} \\ \\ \psi_{lm2} \end{pmatrix}$$

$$+\delta_{m1}\psi_l \begin{pmatrix} i(\delta_{l1}k_{11}\sin\theta_1-\delta_{l2}k_{12}\cos\theta_2) \\ -i(\delta_{l1}k_{11}\cos\theta_1+\delta_{l2}k_{12}\sin\theta_2) \\ -\delta_{l1}k_{11}^2(\lambda_1+2\mu_1\sin^2\theta_1)+\delta_{l2}\mu_1k_{12}^2\sin 2\theta_2 \\ -\delta_{l1}k_{11}^2(\lambda_1+2\mu_1\cos^2\theta_1)-\delta_{l2}\mu_1k_{12}^2\sin 2\theta_2 \\ \delta_{l1}\mu_1k_{11}^2\sin 2\theta_1-\delta_{l2}\mu_1k_{12}^2\cos 2\theta_2 \end{pmatrix}, \quad (1)$$