Speech level measurement

1 Speech and test-signal level adjustment

For reproducible experiments concerning the effect of noise on speech transmission quality, it is important to specify the speech levels, the noise levels and the corresponding signal-to-noise ratios.

Various studies (Brady, 1965; Kryter, 1970; Berry 1971; Steeneken and Houtgast, 1978, 1986) have defined speech level measures. It was also shown that a signal-to-noise ratio variation of only 1-2 dB may have the same effect on the results as typical speaker and inter-listener variations. We therefore specified a method for measuring speech levels and noise levels which offers such a resolution. The measure should be robust for the various speech types (male/female, connected discourse/isolated words), recording conditions (background noise, frequency transfer), and should also be applicable to noise signals. We have developed such a measure (Steeneken and Houtgast, 1978, 1986) mainly for adjusting the test signal level of the STI-method to the speech level for similar conditions. We incorporated the measuring method into the existing specific hardware of the STI measuring device. The measuring method was made generally available by converting the hardware solution into a digital signal-processing algorithm. This also provided a more accurate 'true RMS' calculation algorithm as the former analogue envelope detector is omitted.

2. Speech level measuring method

A high correlation was found between the speech level and the speech intelligibility for level measures based on frequency-weighted speech signals with a reduced contribution of frequency components below approx. 250 Hz (Kryter, 1970; Steeneken and Houtgast, 1978, 1986). The standardized frequency-weighting function according to the A-filter was used for this purpose (standardized for acoustical measurements).

After filtering, the running (intensity) envelope is determined by squaring and low-pass filtering the waveform. From this envelope function the envelope distribution histogram is obtained: the RMS value can be computed from this histogram. The advantage is that the RMS value can also be obtained for values above a certain level after sampling. In order to compare levels of short speech tokens with long silent periods in between (single words) and of connected discourse, a level threshold for suppression of the silent periods is required. Hence, this threshold is applied to the envelope function of the speech signal rather than to the waveform and therefore, does not affect each zero-crossing of the speech signal. The threshold level is defined to be 14 dB below the resulting RMS level (Fig. 1). This definition is signal-related and does not strongly depend on other effects such as background noise level, shape of the envelope distribution, etc. The same principle can be applied to stationary noises but in that case the threshold function is not used.

The relation between various level measures obtained from two types of speech signals (connected discourse, and CVC words in a short carrier phrase) is given in Fig. 2. The level measures are: the 1% peak level (1% overflow criterion), the mean of the peak deflections of a sound level meter set to "fast" (dB(A) fast), the RMS values obtained with a linear detector from the envelope function (RMS_{lin}, no true rms), the RMS values obtained with a squaring detector from the envelope function (RMS_{sq}, true rms), the RMS values obtained with direct sampling and by squaring the wave-form samples (RMS_{dir}, true rms, but application of a threshold is incorrect), and the equivalent peak-level (EPL) according to Brady (1968). The last method is not applicable to noise signals. The RMS-$A_{thr,sq}$ is representative of the level measure as applied in the Esprit-SAM group. For all these measures the addition of A-filtering is indicated by "A" and the application of a threshold by (thr).

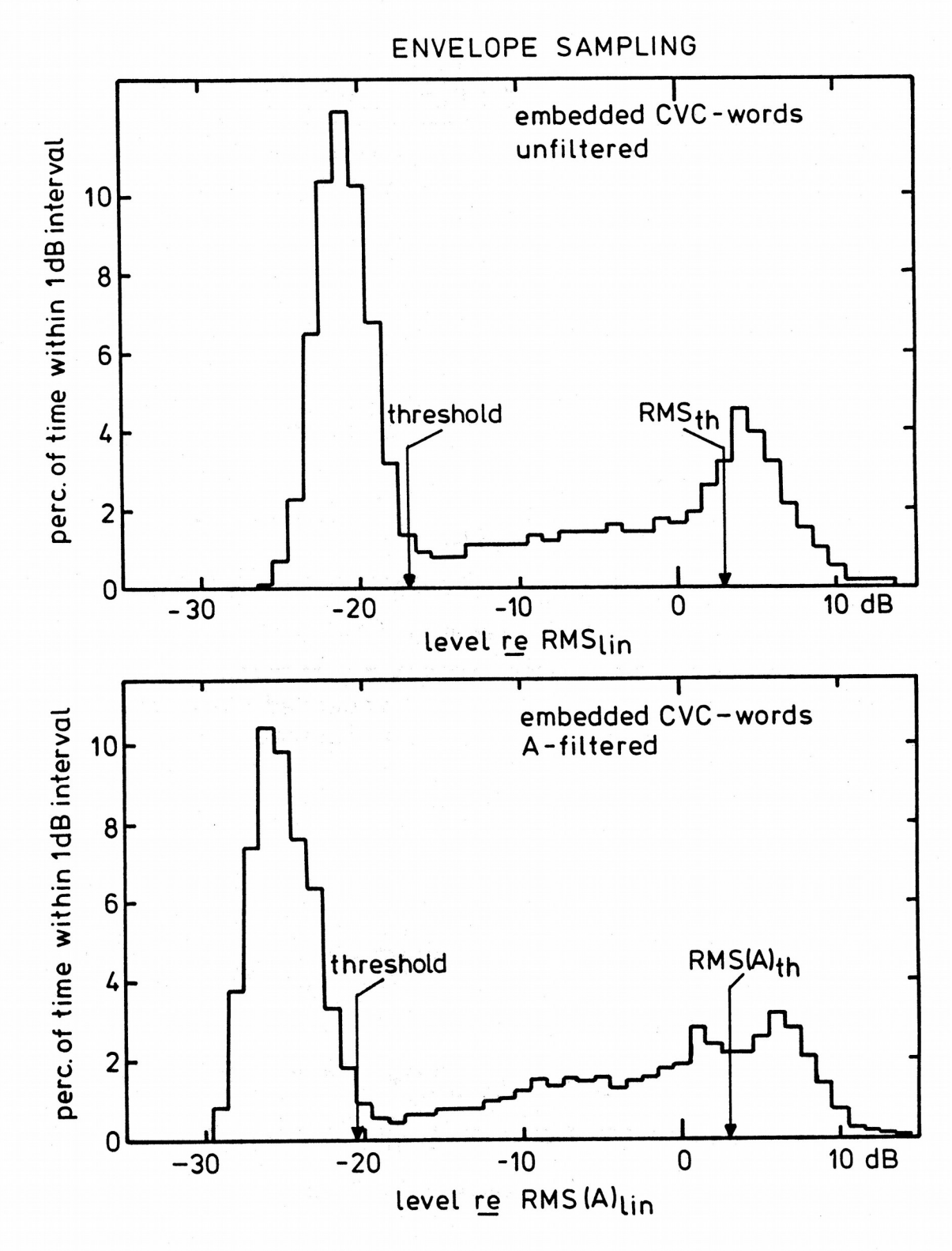

Fig. 1. Level distribution histogram of a 150-s speech fragment from embedded CVC-words obtained with envelope sampling after full-wave rectification and integration. The left arrow line represents a threshold value just above the distribution of the background noise during the silent periods.

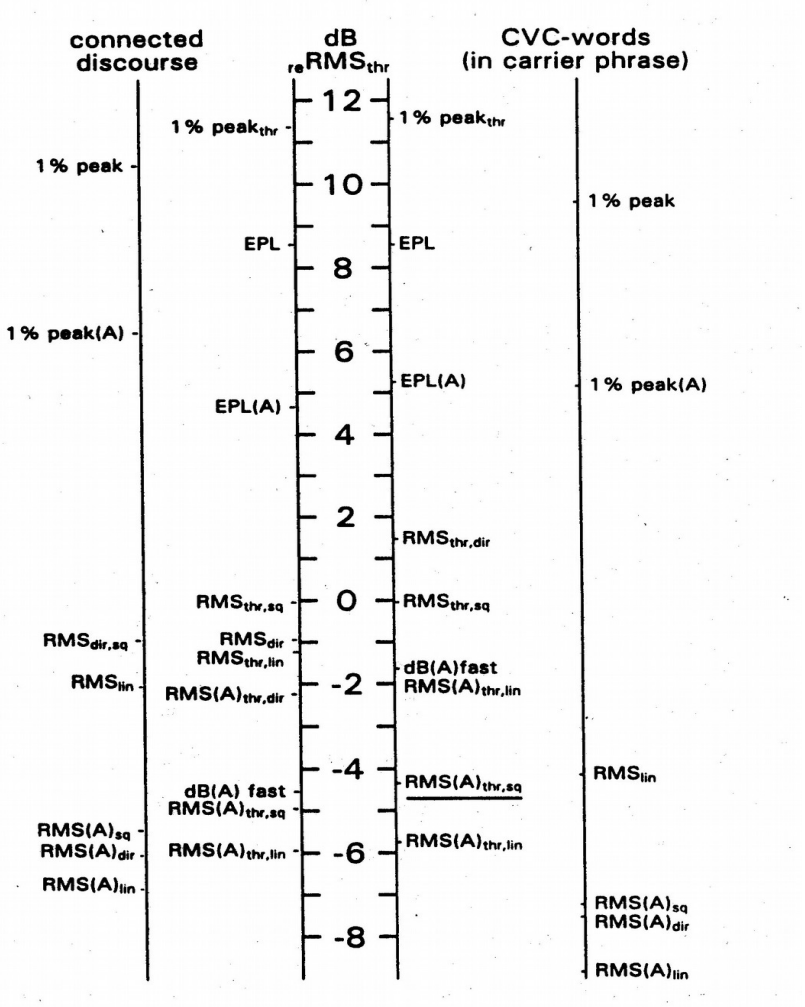

Fig. 2. Relative speech levels for different speech-level measures applied to connected discourse and embedded CVC words. The values are relative to the $RMS_{thr,sq}$ value. For STI measurements (related to CVC-word score prediction) the test- signal level must be adjusted equal to the (underlined) $RMS(A)_{thr,sq}$ value.

3. Speech level measuring procedure

A level measuring program (such as SLM) is required to calculate the RMS-level and the peak level from a digitized signal. The input is typically a data file which can be obtained by using a data-acquisition system (.wav type files).

The output of the SLM-program consists of the overall RMS-value, the RMS_{thr}-value based on samples above the selected threshold, and the peak-value. All levels are corrected for the analogue-to-digital voltage conversion range. Hence the dB values thus obtained refer to 1 μV if the analogue-to-digital converter range is specified.

For the measurement of speech levels to be related to other speech signals or noise, the A-filter switch must be on (default) and the threshold criterium must be 14 dB.

It should be noted that if the SLM-program is used for stationary signals (i.e. noise, periodic signals, etc), the use of a threshold is not relevant and the level measure without threshold correction should be used.

4. References

Berry, R.W. Speech volume measurements on telephone circuits. Proc. IEE, Vol. 118, No 2. Feb 1971.

Brady, P.T. Equivalent peak Level: Athreshold independent speech Level Measure. J. Acoust. Soc. 44, (1968) 695-699

Kryter, K.D. (1970) The effects of noise on man. Academic Press (1970).

Steeneken, H.J.M., Houtgast, T,. A comparison of some methods for measuring speech levels. Report TNO Human Factors 1986-20.