

ENERGY MEASUREMENTS OF SPEECH SOUNDS

Carl Ludvigsen

1. Introduction

Recording of so-called intensity curves frequently forms a part of the routine procedure for processing speech material collected in a phonetic investigation. Devices used for recording of intensity curves are described by Fant (1958 p. 57).

The interpretation of such curves is generally not an easy task as is shown by Fant's characterization of the intensity meter as "... a device that produces an electrical voltage which represents but is not necessarily proportional to the intensity of the speech wave" (1958 p. 57).

The aim of this paper is twofold: to discuss very briefly some aspects of the registration of acoustic energy, and to describe a device constructed at the Institute of Phonetics for measurements of the energy of sound segments. Segments measured with this device may be very short, i.e. of a duration of a few milliseconds.

2. Acoustic measurements

Peterson and McKinney (1961 p. 81) argue that "the measurement of speech power level is relevant to the study of linguistic stress." However, very little attention has been given to the problem of which among several possible measures of speech power level is of major relevance. The following section is devoted to some loose considerations of this question.

If we want to study the acoustics of speech sounds with reference to their properties as stimuli to the ear the measurement of speech power introduced by Fletcher (1953 p. 68) does not seem to be ideal: Fletcher's speech power measures are measures of the total speech energy radiated while a person is talking, whereas here we are only interested in that part of the energy which is active in stimulating the listener. Due to the shadowing effect of the talker's head the two measures are not proportional. Furthermore, it is very difficult to perform the practical measurements.

Another frequently employed measure is obtained by placing a microphone at a distance of e.g. one meter from the talker and feeding the electrical signal from the microphone into an "intensity meter" or a sound level meter. Several problems are involved in this registration procedure: Firstly, does the electrical signal from the microphone represent the sound pressure as it would have been if no microphone were present? This is only the case if the sound field is simple, e.g. a frontal free field, and the microphone is modified to have a corresponding free field response. Secondly, is the sound uniquely described by the electrical signal from the microphone if the sound is regarded as a stimulus to a listener placed in the sound field? This is true only with regard to one type of sound field, cf. ISO Recommendation 454.

The conclusion seems to be that the sound stimulus should be registered at the listener's ear drum or, more conveniently, from a pressure response microphone placed at the end of the ear channel of a dummy head. A recording of the sound pressure at this place will presumably be the best physical description of the stimulus.

3. Processing of the speech signal

The material for further processing is supposed to

be an electrical signal proportional to the instantaneous sound pressure recorded as described above. In order to obtain a measure of the power of the signal some kind of rectification and a smoothing process have to be performed.

3.1. The rectifier

If we want a DC-voltage proportional to the instantaneous power of the microphone signal we must use a square law rectifier (SLR). It is known from the theorem of Parseval that for a periodic¹ signal the DC-value of the squared signal will be proportional to the sum of the mean powers from each of the frequency components of the original signal. Furthermore, we know that this DC-component of the squared signal is independent of the phase properties of the input signal.

The full wave linear rectifier (FWLR), on the other hand, gives an output signal whose DC-value is proportional to the square root of the mean power² of the input signal only for very simple input signals. When the signal is not very simple we cannot generally predict the deviation from the correct value. Furthermore, the output from a linear rectifier is sensitive to phase changes of the input signal.

To demonstrate differences and similarities between the two rectifiers a few examples will be examined. The DC-component emitting from a true RMS-detector consisting of a square law rectifier, a low-pass filter, and a square root device is compared with the DC-component emitting from a mean value detector, i.e. a full wave rectifier followed by a low-pass filter.

-
- 1) Similar statements can be made for aperiodic signals.
 - 2) i.e. the RMS value of the signal.

(a) Pure tone with amplitude A:

$$\text{RMS-detector:} \quad A/\sqrt{2}$$

$$\text{Mean value detector:} \quad 2A/\pi$$

(b) Narrow band noise with mean power P_m :³

$$\text{RMS-detector:} \quad \sqrt{P_m}$$

$$\text{Mean value detector:} \quad \sqrt{2P_m/\pi}$$

(c) Amplitude modulated pure tone:

Mean amplitude: A, Index of modulation: m

$$\text{RMS-detector:} \quad \sqrt{A^2/2 + m^2 A^2/4}$$

$$\text{Mean value detector:} \quad 2A/\pi$$

Now, suppose that we tested two intensity meters: One with a full wave linear rectifier (FWLR) followed by a low-pass filter and the other with a square law rectifier (SLR) followed by a low-pass filter and a square root circuit, i.e. a true RMS-detector. If we adjusted the two instruments to give the same reading for a pure tone input signal, then we would not obtain identical readings on the two instruments for any of the other two signals mentioned above. For the amplitude modulated tone we could, by varying the degree of modulation (m), obtain a varying reading on the SLR, but a constant reading on the FWLR.

None of the examples mentioned were speech sounds, and experiments (cf. Peterson and McKinney 1961 and Fant 1958 p. 57) have shown that the differences occur-

3) Cf. Rice (1944).

ring when speech sounds are considered are moderate. However, with the linear rectifier it is somewhat dubious what we actually measure, and conclusions drawn from intensity curves should perhaps be of a qualitative rather than a quantitative nature. Furthermore, the use of an FWLR will complicate a statistical treatment of the data from intensity meters since differences in the means of two variables may to a certain degree be due to unsatisfactory rectification.

Nevertheless, many rectifiers used in intensity meters are of the linear type. The reason obviously is that an SLR is more difficult to construct if it is to have a sufficiently large dynamic range. However, as technology has advanced in the last years we are now able to construct SLR's which satisfy our demands and which are not excessively expensive.

3.2. Integration

If we use an intensity meter with an SLR followed by a smoothing low-pass filter it is no simple task to measure the energy of a single sound segment. Even if the segment is isolated the interpretation of the output is not a straightforward matter. If the duration of the segment is long compared to the transient time⁴ of the filter the mean of the output signal equals the average power of the segment provided that the AC-part of the output signal is small compared to the mean, i.e. to the DC-component of the output signal. If, however, the duration of the segment approximates the transient time of the filter the output signal will be highly dependent on the shape of the impulse response of the filter. In this case the set-up does not

4) Often called "the integration time" of the filter, cf. Fant (1959 p. 9).

give useful measurements. If the duration of the segment is short compared to the transient time then the peak of the output signal will correspond to the energy of the input signal. When the set-up is calibrated for this type of measurement and the duration of the signal is known we can calculate the average power.

From these considerations we may conclude that if the energy of transients such as stop bursts is measured by the output of an ordinary intensity meter the results are subject to several uncertainties.

To overcome the difficulties involved in energy measurements of very short segments of speech one may replace the low-pass filter by a true integrator. The integration should start at the beginning of the sound segment in question and stop at the end of the segment. The final output voltage from the integrator would then be proportional to the average power of the segment, irrespective of its duration.

4. Construction of a square law rectifier and a true integrator

Fig. 1 shows a simplified diagram of the device.

4.1. Principle of the square law rectifier

The characteristic of the rectifier is a piecewise linear approximation to a parabola. The input signal will be fed into a voltage divider via an FWLR. The voltage levels on the voltage divider will trigger the four transistors on and off. The emitter resistors are selected so that the output current is approximately proportional to the square of the instantaneous input voltage to the squaring circuit.

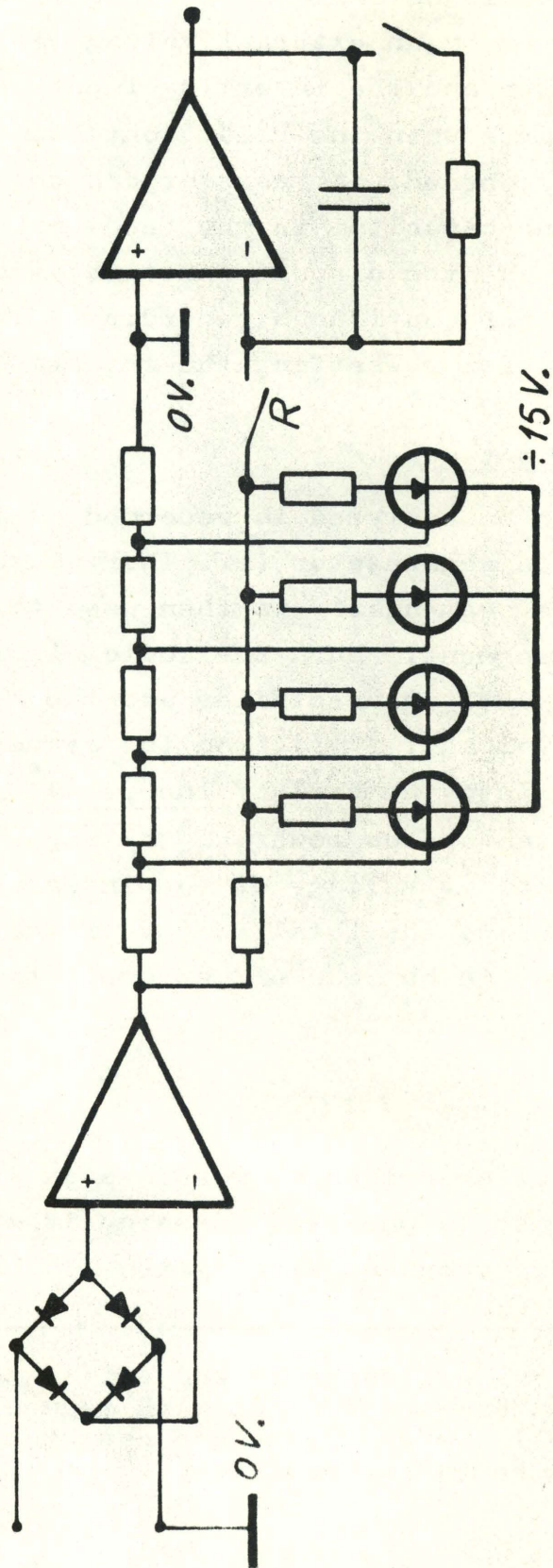



fig.1

4.2. The integrator

The integrator is an ordinary summing integrator. A relay, R, triggered by an external voltage is inserted between the rectifier and the inverting input terminal. The integration starts when the relay contacts close and stops when they are opened. A resistor can be coupled in parallel with the capacitor in the feed-back loop of the integrator to make the circuit function as a simple RC low-pass filter for continuous recording. This resistor is also used for re-setting the integrator.

4.3. Experimental set-up

The signal to be analyzed is recorded on the loop tape-recorder of the segmentator (cf. Thorvaldsen 1970). The "window" of the segmentator is then used to isolate the desired sound segment. When the audio signal from the segmentator is fed into the rectifier and the "gating signal" (i.e. a step voltage: ) from the segmentator is simultaneously fed into the coil of the relay⁵ an integration of the instantaneous power of the segment takes place. The final output voltage of the integrator will then be proportional to the total energy of the sound segment. This voltage can be read off an amplifier DC-voltmeter.

5. Final remarks

The instrument described is meant as an experimental design only. For use in phonetic research it must be incorporated in a more comprehensive system. The addition

5) This synchronous switching on and off of the input to the integrator is essential with segments of very low intensity in order to exclude contributions from background noise on the tape loop.

of a compressor as well as a more sophisticated low-pass filter for continuous recordings will be an obvious improvement.

References

- Fant, Gunnar 1958: "Modern Instruments and Methods for Acoustic Studies of Speech",
Acta Polytechnica Scandinavica.
- " 1959: "Acoustic Analysis and Synthesis of Speech with Application to Swedish",
Ericsson Technics No. 1.
- Fletcher, Harvey 1953: Speech and Hearing in Communication.
D. Van Nostrand Company, Inc.
- ISO Recommendation 454: Relation between the loudness of narrow bands of noise in diffuse-field and in a frontally incident free-field.
- Peterson, Gordon E. and McKinney, Norris P. 1961: "The Measurement of Speech Power",
Phonetica Vol. 7.
- Rice, S. O. 1945: "Mathematical Analysis of Random Noise",
Bell System Technical Journal (January 1945).
- Thorvaldsen, Poul 1969: "Construction of a Segmentator",
ARIPUC 4, p. 8-14.