(ARIPUC 13, 1979)

INFLUENCE OF MICROPHONE POSITION IN THE RECORDING OF SPEECH SIGNALS

Carl Ludvigsen

Abstract: Sound pressure levels of various speech sounds are measured simultaneously at different distances from the mouth. The observed values for low vowels and [s] differ especially close to the mouth from those predicted from the distance law for sound radiation. The variation of sound pressure with distance seems to depend on the speech sound in question. Some consequences hereof are pointed out. The results are compared with calculated values of the sound pressure from a sound source on a rigid sphere. Some of the observed deviations from the distance law seem attributable to the different frequency composition of the speech sounds. However, some of the observations (e.g. the difference between the variation of low and high vowels at positions close to the mouth) cannot be accounted for by the model.

1. Introduction

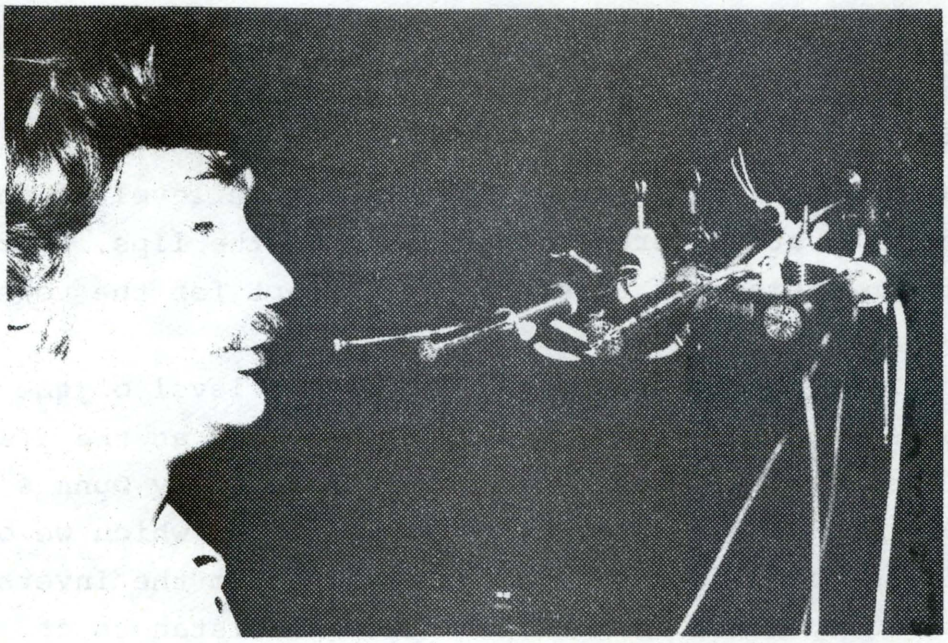
Recording of speech signals is a standard procedure in all phonetic laboratories. The microphone is typically placed at a distance of 25 - 31 cm directly in front of the mouth (see, e.g., Lehiste & Peterson 1959, Fairbanks et al. 1950, Strevens 1960, and House & Fairbanks 1953). Obviously, there are good reasons for that choice. If the microphone is placed too close to the mouth (e.g. closer than 10 cm), the expirated air will tend to generate turbulent noise, when passing the microphone; on the other hand, if the microphone is placed far from the mouth (e.g. more than 1 m away), problems with the signal-to-noise ratio may arise. The main reason for placing the microphone directly in front of the lips is that high frequency sounds are mainly radiated in this direction. Only one detailed discussion concerning this question has been published until now, namely Dunn & Farnsworth (1939).

They measure the average sound pressure at seventy-six positions, in different distances and directions from the mouth of a single speaker. They find that the variation of the intensity of speech measured on a horizontal line directly in front of the mouth is similar to that of a single source placed 0,6 cm inside the lips, i.e. the sound pressure at two points at distances L and 2L from the acoustic centre 0,6 cm inside the lips differs with 6 dB. This relationship is often called the 1/R-law. The measurements are average RMS-pressures of a 15 sec. utterance and measured in different frequency bands. The over-all average intensity conforms to the 1/R-law with a high degree of accuracy, whereas deviations are observed within some of the frequency bands.

In recording speech material in our own laboratory, we have noticed a tendency for the intensity difference between low and high vowels to be less pronounced if the microphone distance is short (e.g. 5 cm) and for the intensity of unvoiced s-sounds compared to the intensity of vowels to depend on the microphone distance as well. These observations are partly in agreement with Dunn & Farnsworth (1939), if we assume that the intensity of a vowel is mainly determined by the intensity of the first formant. However, since the measurements of Dunn & Farnsworth are average values for 15 sec. of connected speech, only a gross estimate of the variation for single speech sounds can be derived from their data. In order to throw further light on this problem, a series of measurements were carried out.

2. Measurements

A male speaker, who is a trained phonetician, sat on a specially constructed chair in the anechoic chamber of the Institute of Phonetics. The distance law (1/R-law) is complied with in this room within ± 1.0 dB in the frequency range 100 - 10.000 Hz in the space used for measurements. The subject's neck was supported by a headrest. In front of the subject's lips was placed a row of 5 microphones: 5 cm from the lips a 1/4-inch B & K condenser microphone, 10 cm from the lips a ½-inch B & K condenser microphone, and 20, 30 and 100 cm from the lips 1/1-inch B & K condenser microphones, see figs. 1 and 2.



Figures 1 and 2

Microphone set-up in the anechoic chamber. The microphones are placed at 5, 10, 20, 30, and 100 cm from the average position of the centre of the lips.

The subject read a list of nonsense syllables and isolated vowels at a constant and comfortable speaking level. Each item appeared 6 times in the list, in random order. The list consisted of the vowels:

[i, e, ɛ, æ, a, ɑ, y, ø, œ, Œ, u, o, ɔ, ʌ, ɒ]

and the nonsense syllables:

[mi, mu, ma, si, su, sa].

The signals from the five microphones were recorded simultaneously. For calibration purposes, a 200 Hz pure tone was recorded before and after the reading of the list. The maximum RMS vowel amplitudes were then registered on a B & K level recorder, type 2305. The amplitudes of the initial [m] and [s] segments were obtained by feeding the signal to an intensity meter (with a double-linear rectifier) and registering the output on a fast ink-writer (Mingograph). Identification of different segments was performed by comparing the intensity curve with an intensity curve of the high-pass filtered signal (500 Hz) and the duplex oscillogram.

3. Results

3.1 Vowels

Figs. 3-5 show the average sound pressure level of the 15 vowels measured at different distances from the lips. Each point represents an average of 6 recordings, except for the vowel [æ], which is based on 5 recordings only.

The curves show that the sound pressure level of the vowels at distances from 10 cm to 1 m decreases 5-6 dB as the distance doubles, i.e. slightly less than the 6 dB found by Dunn & Farnsworth (1939) and the theoretical value of 6 dB, which we obtain from a simple point source. The deviation from the inverse law is statistically the same for all vowels at distances of 10-100 cm from the lips.

At distances of 5-10 cm from the lips, the deviation from the inverse law is greater and is not the same for all vowels. Typically, the sound pressure for low vowels [ɒ], [Œ] and [ɑ] is only 1-2 dB higher 5 cm from the lips as compared to 10 cm from the lips, while the difference for high vowels [i], [e], [y], [ø], [u] and [o] is 3-5 dB. This difference between low and high vowels is statistically significant beyond the 0.1 per cent level.

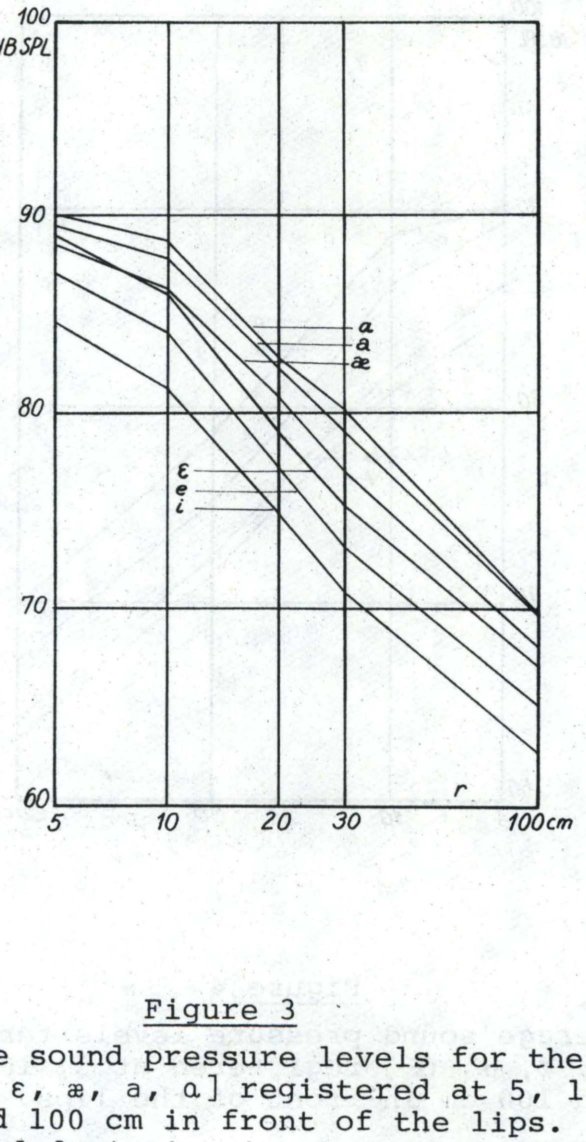


Figure 3

Average sound pressure levels for the vowels [i, e, ɛ, æ, a, ɑ] registered at 5, 10, 20, 30, and 100 cm in front of the lips. The standard deviation is close to 1 dB for all registrations.

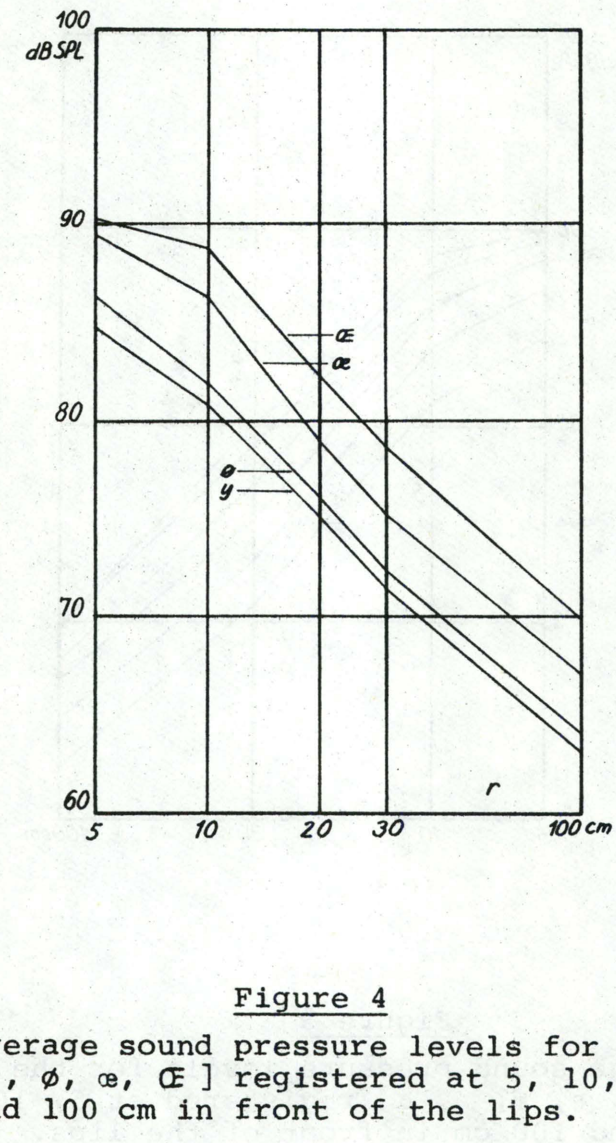


Figure 4

Average sound pressure levels for the vowels [y, ø, œ, Œ] registered at 5, 10, 20, 30, and 100 cm in front of the lips.

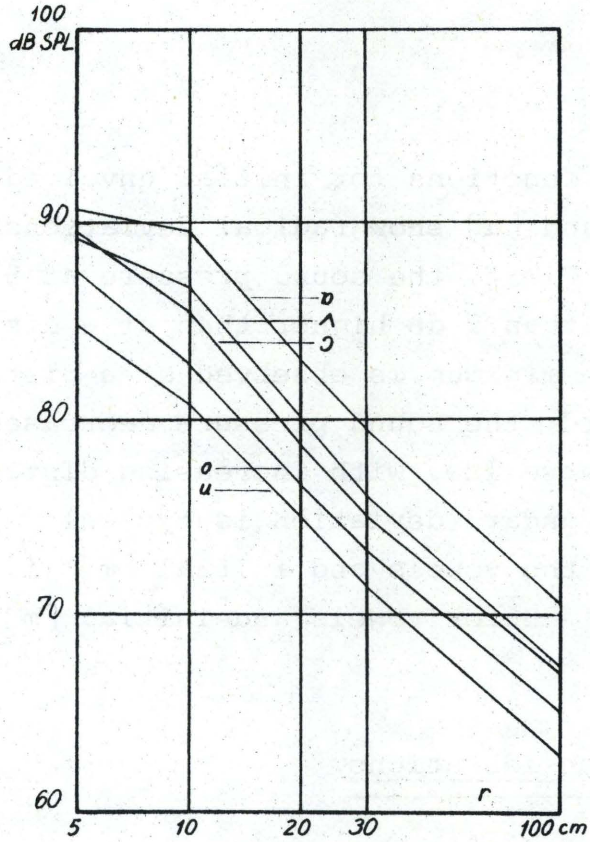


Figure 5

Average sound pressure levels for the vowels [u, o, ɔ, ʌ, ɒ] registered at 5, 10, 20, 30, and 100 cm in front of the lips.

3.2 Consonants

3.2.1 Initial [m]

The observed distance functions of initial [m] before the vowels [i], [u], and [ɑ] resemble those found for narrow vowels, except that the deviation from the inverse law seems to be smaller, cf. fig. 6.

3.2.2 Initial [s]

The distance functions for initial unvoiced [s] before the vowels [i], [u], and [ɑ] show radical deviations from the inverse law, cf. fig. 6. First, the sound pressure at 5 cm from the lips is typically less than 1 dB higher than at a distance of 10 cm. Second, a relative minimum is observed at a distance of 30 cm from the lips, and third, the sound pressure decreases more than predicted by the inverse law, with increasing distance. The unbiased estimate of the standard deviation is typically 3-5 times greater than observed for the vowels and initial [m] (i.e. 3-5 dB compared to approximately 1 dB for vowels and initial [m]).

4. Theoretical considerations

The radiation from the mouth depends on several parameters, e.g. the dimensions of the mouth orifice, the head and the body, whether the person is sitting or standing, etc. However, objects which are small compared to the wavelength or are far from the sound source and the point of observation will only have a small influence on the sound field in the point of observation. Therefore, fairly simple models such as a piston in an infinite rigid wall or a rigid sphere can be used as a basis for the calculation of the radiation from the mouth. In order to explain the observed deviations from the distance law, a model with a piston in an infinite wall is too simple, since it predicts that sound pressure varies according to the inverse law with amplitudes proportional to the frequency, provided that the dimension of the piston is small compared to the wavelength. This will be the case for piston diameters less than 4 cm in the frequency range up till approximately 2-3 kHz.

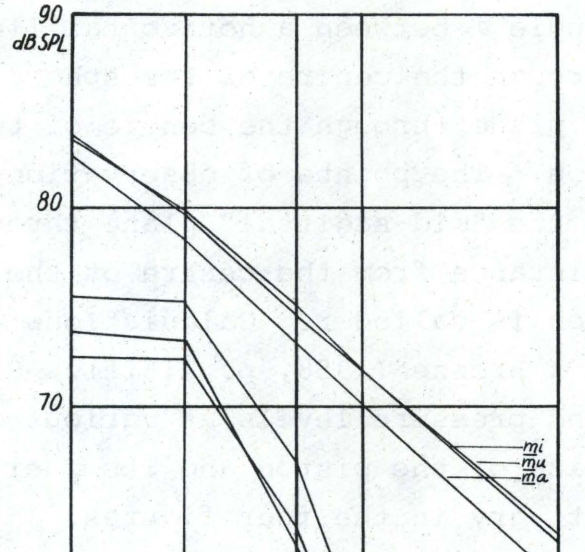


Figure 6

Average sound pressure levels for initial [s] and [m] registered at 5, 10, 20, 30, and 100 cm in front of the lips. The standard deviation for initial [m] is typically 1 dB, whereas the standard deviation for initial [s] is 3-5 dB.

The model with a piston in a rigid sphere gives less trivial results. For the calculations, a sphere with a diameter of 20 cm has been chosen as an approximation to the subject's head. The dimension of the piston is specified by the angle γ_0 between a line from the centre of the sphere to the centre of the piston, and a line from the centre of the sphere to a point on the periphery of the piston. The position of the piston on the sphere is specified by the angle φ_0 between a horizontal line in the "mid-sagittal" plane through the centre of the sphere, and a line in the "mid-sagittal" plane through the centre of the sphere and the centre of the piston. The points of observation are placed on a horizontal line in the "mid-sagittal" plane through the centre of the piston. The distance from the centre of the piston to the point of observation is called r. Calculations are performed according to Stenzel & Brosze (1958, p. 116ff). Figs. 7-10 show the calculated sound pressure levels at various distances from the piston. The size of the piston and its position on the sphere are parameters that vary in the four figures.

Fig. 7 shows the calculated sound pressures at various points in the sound field from a point source on a rigid sphere. The different curves represent different frequencies from 125 Hz to 8000 Hz. For every frequency the product of the volume displacement of the sound source and the frequency is the same arbitrary value and the ordinate scale is in dB with an arbitrary reference. On the average, the calculated sound pressure drops by 6 dB with a doubling of the distance, as predicted by the inverse law. The slope for the higher frequencies is, however, slightly less than 6 dB and slightly more than 6 dB at low frequencies.

Fig. 8 is essentially the same as fig. 7, except that the point source is replaced by a piston ($\gamma_0 = 15^\circ$) with the same volume displacement. At distances greater than 20 cm, the calculated sound pressures are almost identical for the two types of sound source. Closer to the sound source, a difference is observed. The sound pressures calculated with a piston source are somewhat smaller for a given, small, distance than when calculated on the basis of a point source. This difference is less pronounced at lower frequencies.

In fig. 9, the calculations are again based on a point source, but its position on the sphere is lowered by 30° compared to the

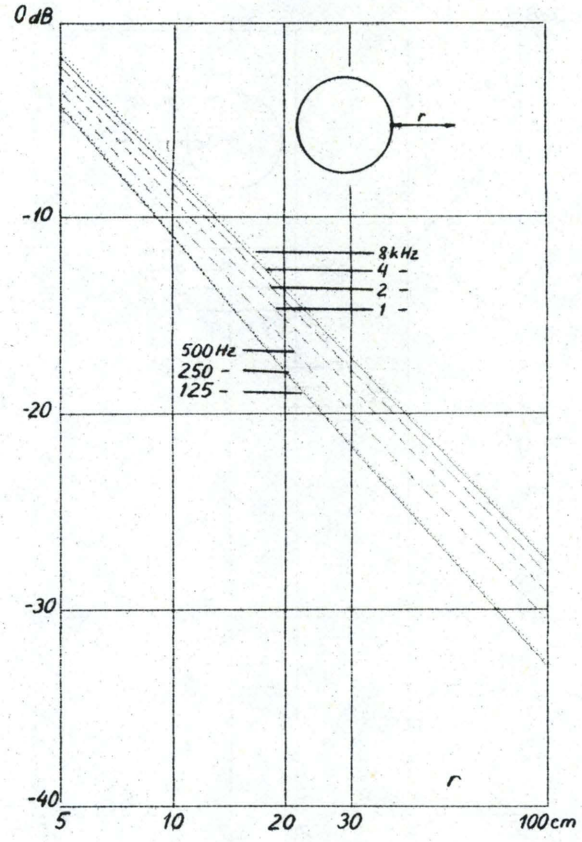
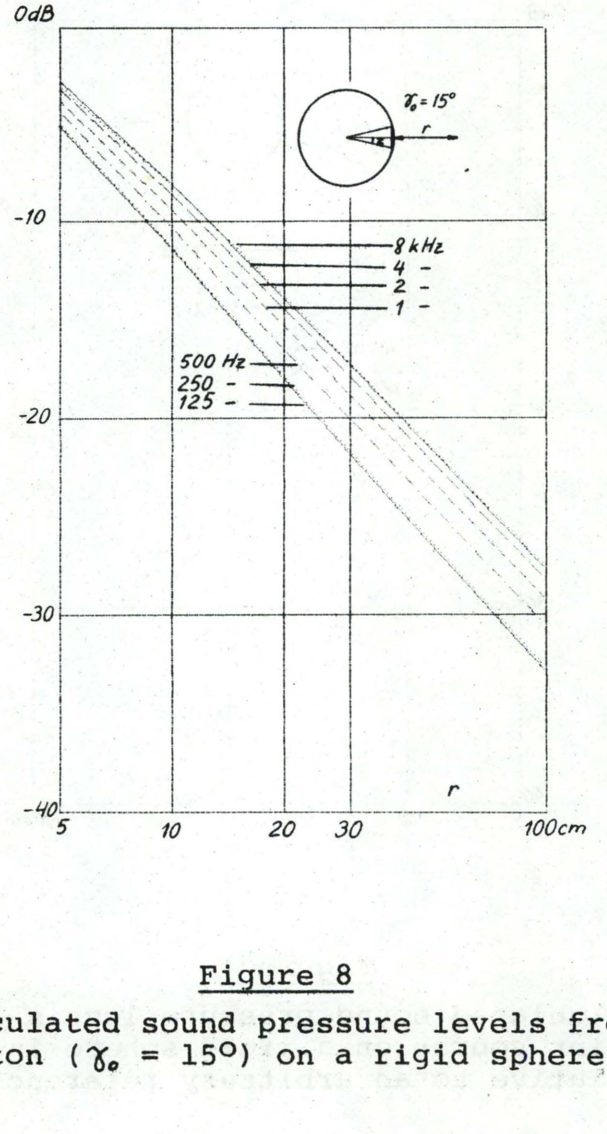


Figure 7

Calculated sound pressure levels from a point source on a rigid sphere in DB, relative to an arbitrary reference.



<u>Figure 8</u>

Calculated sound pressure levels from a piston ($\gamma_0 = 15^{\circ}$) on a rigid sphere.

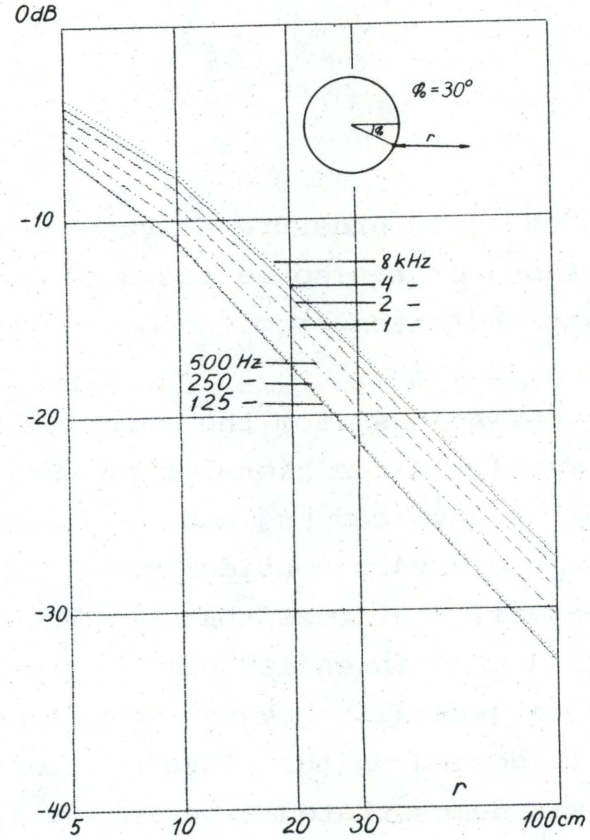


Figure 9

Calculated sound pressure levels from a point source on a rigid sphere. The point source is lowered by 30^{o}, compared with the source in fig. 7.

sound sources in figs. 7 and 8. This causes the curves to be less steep close to the sound source.

In fig. 10 the sound source is a piston ($\gamma_o = 15^o$). The centre of the piston is lowered by 30^o compared to the situation described in fig. 8. This causes a pronounced deviation from the inverse law at short distances from the sphere, especially for higher frequencies.

5. Discussion

Measurements of sound pressure at various distances from the mouth and calculations of the sound field from a source on a rigid sphere suggest that radiation from the mouth cannot be adequately described by the inverse law.

The greatest deviations from the inverse law are found close to the mouth and mainly in the higher frequency bands. Correspondingly, intensities of unvoiced [s] and, to a somewhat higher degree, of open vowels did vary considerably less with the distance from the mouth at small distances than predicted by the inverse law. This means that the intensity difference between high and low vowels, and more generally between low frequency and high frequency sounds, will depend on the distance from the microphone to the mouth. This was demonstrated experimentally by measuring the intensity of the nasal [m] and the unvoiced fricative [s] simultaneously at various distances from the mouth.

A comparison of the measured and the calculated distance functions shows that these findings can be partly accounted for by a model for the radiation, where a piston is placed on a rigid sphere as described in fig. 10. At short distances, the 4 kHz curve in fig. 10 resembles the curves for [s] in fig. 6 and the 250 Hz curve of fig. 10 resembles the curves for [m] in fig. 6. However, the observed low values of the intensity of [s] at a distance of 30 cm cannot be explained by the model and neither can the tendency for the over-all slope of [s] to be steeper than the slope of [m]. Also, the tendency for the 500 Hz curve in fig. 10 to be less steep at short distances than the 250 Hz curve is inadequate to explain the difference in slope close to the lips which were observed between low and high vowels.

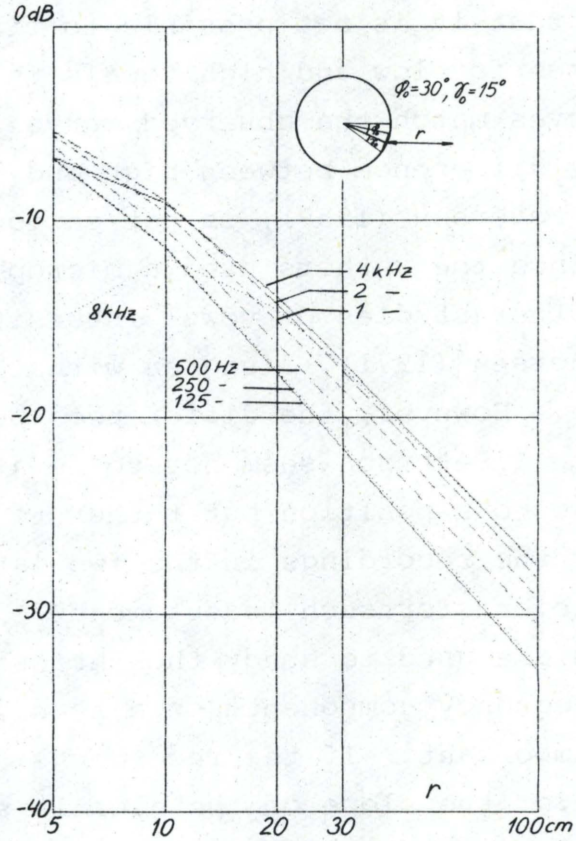


Figure 10

Calculated sound pressure levels from a piston ($\gamma_0 = 15^{\circ}$) on a rigid sphere. The piston is lowered by 30° compared with the piston in fig. 8.

If we assume that the acoustic centre for radiation of speech sounds is placed somewhat inside the mouth, the calculated slopes will be less steep at short distances from the lips. If we further assume that the position of the acoustic centre is closer to the lips for high vowels compared to low vowels, we would expect a less steep slope for low vowels than for high vowels. However, calculations show that it is not possible to find positions for the acoustic centres for low and high vowels in such a way that the calculated curves match the observed curves.

The intensity difference between high and low vowels found by e.g. Lehiste & Peterson (1959), is not seriously affected by these findings, since the authors used a microphone-to-mouth distance of 30 cm. Other studies of vowel intensity, Sharf (1966) and Ludvigsen & Thorsen (1971), disagree with the findings of Lehiste & Peterson. However, the differences between the latter studies and Lehiste & Peterson seem not to be attributable to a difference in microphone position but rather to the dimensions of the rooms used for the recordings in the two latter studies, which were insufficient for such measurements.

If recordings are used to study the intensity relation between different frequency components of a sound, the microphone position will be important. If the recordings, furthermore, are used to estimate e.g. the slope of the glottis spectrum by means of inverse filtering, appreciable differences may appear due to different microphone positions.

The microphone position is most critical at short distances from the mouth, and it seems preferable to place the microphone at a rather long distance (e.g. 1 m) from the mouth to avoid these problems.

References

Dunn, H.K. and D.W. Farnsworth 1939: "Exploration of pressure field around the human head during speech", JASA 10, p. 184-199

Fairbanks, G., A.S. House and E.L. Stevens 1950: "An experimental study of vowel intensities", JASA 22, p. 457-459

House, A.S. and G. Fairbanks 1953: "The influence of consonant environment upon the secondary acoustical characteristics of vowels", JASA 25, p. 105-113

Lehiste, I. and G.E. Peterson 1959: "Vowel amplitude and phonemic stress in American English", JASA 31, p. 428-435

Ludvigsen, C. and N. Thorsen 1971: "Comparison of sound pressure level and loudness (GF) measurements on speech sounds", ARIPUC 5, p. 163-174

Stenzel, H. and O. Brosze 1958: Leitfaden zur Berechnung von Schallvorgängen, Springer Verlag, Berlin

Sharf, D.J. 1966: "Variations in vowel intensity measurements", L&S 9, p. 250-256

Strevens, P. 1960: "Spectra of fricative noise in human speech", L&S 3, p. 32-49