Noise Pollution
Edited by A. Lara Sáenz and R. W. B. Stephens

CHAPTER 9
Noise Interference with Oral Communications

T. Tarnóczy

9.1 COMMUNICATION BY SPEECH

Speech is the most important means of communication for human beings. Its functioning is linked with a bilateral human capability, that of the ability to form and to understand speech. Basic conditions of this are the intention to communicate, the system of signals used are according to a common agreement, the emitting and receiving organs being in perfect condition and there is the ability to comprehend. The chain of speech communication is a complicated system where brain, nerves, speech organs and hearing are all participating (Figure 9.1). Feedback serving for the checking of transmission, namely, the fact that the speaker also hears his own voice, is not a basic condition but improves the speed and safety of communication.

Communication by sounds is known also in the world of animals, but this differs from human speech, because with animals each information is given by a separate signal. (Thus, of course, it is inept for information on thoughts). Human speech is a system of signals built up on 35–40 elements, where the meaning relies on various forms of connections between the elements. The possibility is given, not only to form a new speech sound compound for any meaning, but also for the construction of a grammatical system in order to develop a literary language semantically perfectly aligned.

Speech is perfectly understandable — under appropriate acoustic circumstances — for healthy people mutually knowing the system of acoustical signals agreed upon, i.e. the language. By acoustical circumstances are meant that the distance between speaker and hearer and the environmental background noise are suitable for the speech loudness. 'Intelligibility' depends on these factors and it may even be lost. Therefore, the notion of intelligibility has a great importance in the evaluation of speech communication. With noise present, both the speaker and the hearer are faced with a difficult task: the former is disturbed by noise in thinking and the formation of speech, whilst the latter in

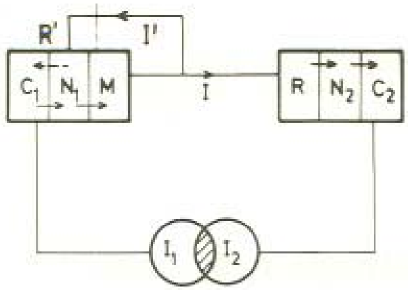

Figure 9.1 Diagram of the communication chain: I = information, $C_{1,2}$ = cerebral function, $N_{1,2}$ = nervous connections, M = motoric emission, R = reception, I', R' = feedback information for the emitter, $I_{1,2}$ = information store of both emitter and receiver

hearing the series of signals and the cerebral evaluation of the information content. It is, first of all, the phenomenon of 'masking' that makes communication difficult. (Zwicker, Chapter 4, this volume).

Brain-work is, as a matter of fact, the decoding of the acoustical material heard, and it will become more and more difficult according to what share and what part of the information is made unintelligible. Therefore, the harmful effect of speech-noise interference consists not only in the fact that the information is not understood, but also that the establishment of communication requires great efforts both on the side of emission, e.g. by shouting and on that of reception, i.e. combinative thinking.

That is why Robinson (1970) ranks speech interference among primary human effects of noise. In the case of speech interference, our brain deals not directly with the determination of some property of noise, but with the decoding of information material distorted indirectly by the masking effect of noise. It is also true that speech interference is not a secondary effect of such character as that of impaired hearing, nerval or organic diseases. This chain of thought necessitates the insertion of a third kind of noise effect between the two extreme types, the sensational judgement of the quantitative data of noise and the development of harmful effects of noise. To this inter-state, could be ranked also the sleep-disturbing effect of noise, whose further consequences are tiredness, reduced mental alertness, nervousness, etc. It is characteristic of these categories (speech interference, sleep interference) that they exist only when noise is present and with its elimination the phenomenon also disappears. This is not the case with impaired hearing, nerval taint or organic diseases which remain after the removal of noise. These are the real secondary noise effects.*

Sound-forming is usually of a not-too-differentiated quality, but it has variable components. Such is first of all the frequency band used. Within one species the emitted band of sound and the sensibility range of the receiver organ are naturally in harmony. But the sound will not necessarily be perceived by members of another species. It is very interesting that among mammalia examined until now, man is perhaps the most sensitive in the range of low-frequency sounds, while fully insensitive to those of a very high pitch. The frequency of the highest sound heard by man is 16–18 kHz, while the ears of a chimpanzee are sensitive up to 22 kHz, those of a dog go up to 38 kHz, of a washbear up to 50 kHz, of a cat up to 75 kHz and those of a bat up to 120 kHz. Non-mammals are usually insensitive to high-pitched sounds. This is surprising because previously we had though that crickets as well as birds were sensitive to very high-pitched sounds. In reality, the upper limit of the hearing of birds is around 8–12 kHz according to experiments of Tembrock and colleagues (Tembrock, 1959).

Variation is possible first of all in pitch and time but it seems much less probable that the quality of sound would have an information-carrying role. However, for man even this factor results in unlimited combination possibilities, i.e. the multiplication of meaning content.

9.2 MECHANISM OF SPEECH FORMATION

9.2.1 Vibration of Vocal Chords

The primary source of the formation of speech sounds if the glottis (Figure 9.2). The closed vocal chords are made to vibrate by the flow of air flowing out of lungs. The fundamental tones generated in this way are transformed into speech sounds by various resonator cavities, finally the sound will be radiated into the environmental airspace through the oral aperture and/or nostrils.

Beneath the vocal chords one single large cavity — the chest — is to be found. This cavity includes the energy source of the sound formation system —

* For acoustical communication, several examples may be found also in the world of animals. The primary reason for this is the suitability of biological sound for communication purposes. This type of sound may be easily and rapidly formed, it contains a relatively wide variation of information possibilities and has also adequate propagation qualities even in the presence of natural obstacles (bush, wood, forest). An important characteristic is that acoustic signals may be 'coded', that is animals belonging to other species do not know what information is contained in the given signal. Animal sound has therefore the features to be a carrier of a most important chain of communication. In the course of investigations it turns out that most animals have a relatively abundant acoustical vocabulary. At least a dozen animals are known that are able to issue 18–25 various phonetic signals, e.g. finch, dolphin, roaring monkey, etc.

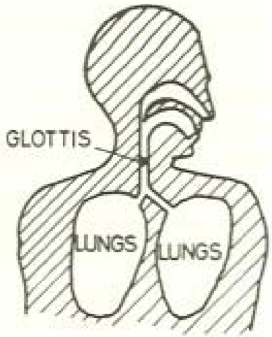

Figure 9.2 Schematic view of speech organs

the lungs. Since there are several soft and thus largely damping substances in the cavity, it is first of all the trachea that has the role of a resonator. Since this resonator has a rather low self-resonance, it does not influence the sound quality of speech sounds but may characterize individual timbre. In case of a normal speech sound the pressure of air flowing out of lungs corresponds to that of a water column of about 4 cm. When shouting very loudly and with high pitch it may even reach the pressure of a 20-cm-high water column. The energy of the streaming air changes partly into sound energy, but the efficiency of transformation is very low, only of the order of 0.1%.

Vocal chords block up the way of air like an elastic-tightened membrane and form the sound source of speech and of the singing voice. It comprises a pair of folds whose longitudinal tension, setting and gap-size may be changed depending on our will. Muscles are partly imbedded into the vocal chords and partly motivate the cartilages placed around them. Vocal chords adhere to the cartilages and thus the exact setting and control is made by means of these. Cricoid is the basic cartilage of the larynx. Above it is the thyroid cartilage and which can tighten the front wall of the larynx on the inner part of which are situated the beginnings of the vocal chords. Arytenoid cartilages are placed on the backside of the cricoid and carry the ends of the vocal chords which they activate. The cartilage covering the larynx has a protective role and has nothing to do with sound formation. In Figures 9.3(a) and 9.3(b), the cross section of larynx and the rough layout of vocal chords are presented. The movement of vocal chords may be best explained on the cross section, but experimentally it may be examined precisely in the two other major directions. Pseudo-vocal chords to be seen in the figure do not participate in sound formation under normal conditions, but in pathological cases and with operative help they may take over — imperfectly — the voice-forming role of vocal chords.

The length of vocal chords is 20–25 mm. Their movement takes place in such a way that the air-flow pouring out of lungs knocks against the obstacle raised by closed glottis. If surplus pressure exceeds the compressing strength of vocal chords the flow of air breaks through the closure. In this way the surplus pressure will immediately diminish and resulting from their elasticity vocal

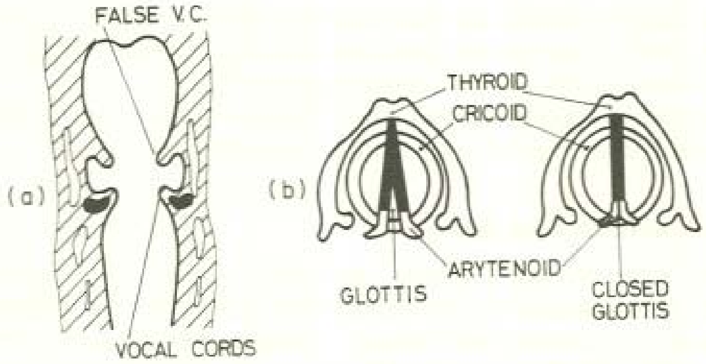

Figure 9.3 (a) Vertical cross-section of larynx; (b) horizontal sections in the height of the glottis in the open and closed states

chords are closed again. Air flowing out is continuously replaced from lungs, therefore after a certain time it will reach the surplus pressure required for a break through again, and the movement goes on. The elasticity data of vocal chords and the flow data of the air current are naturally in harmony. The relationship between flow velocity and pressure in the contraction is such that at the narrowest part of larynx (glottis) the air pressure will be least. This physical relationship partly governs the closure of the vocal chords.

Vocal chords do not form a system opening and closing in plane, but move away and upward, then with an elliptical movement they collide further down and, when closing the edges are pressed together or may even be placed one above the other. Closure does not take place at once along the glottis, either, but the glottis is gradually closed from the front to the back and thus the entire movement has a character of snake-like movement in space that may be observed by means of a stroboscope. After closing, vocal chords strive for the repetition of the process even by themselves — because of their elasticity — and therefore, they start again towards the opening upwards. The well-timed increase of pressure of air coming from lungs assists this movement (van den Berg, 1958).

Speech sounds are basically generated through the movement of vocal chords, but more detailed data of the process require examination. Such are the form of vibration, the relationship between the duration of opening and closing of the chords as well as the harmonic content of sound thus obtained. The time-pattern of the glottis opening is a function of typical triangular form where the initial stage is a steeply increasing opening. The duration of opening and closing state may be examined either by the stroboscope method or by the waveform in time of the sound obtained. According to data obtained from oscillograms the opening quotient amounts to 0.2 at the normal pitch of speech sound and may increase up to a value of 0.7 with a two-octave increase of pitch. As against this, according to the stroboscopic method the opening quotient hardly depends on the frequency and its value is around 0.7 (Timcke, 1956). Experiments show that the absolute value of the opening time is constant: 2–2.5 ms (Joos, 1948; Tarnoczy, 1951).

It is difficult to make a conclusion here, because of the problem of defining exactly what we mean by an acoustically-open state.

9.2.2 Differentiated Phonation

The sound generated by the glottis — the voice — is one of the raw materials of speech. Its oscillogram has the characteristic saw-tooth form similar to that of mechanical or electrical self-induced vibrations. The harmonic content of the voice decreases by about 12 dB/octave. This raw material is transformed into speech sounds with a characteristic timbre by the resonance effect of cavities above the glottis.

However, there are also several other methods of forming speech sounds. In the cavities above the glottis, closures and narrowings may be established and thus various kinds of noise may be created. Narrow gaps may be formed between lips, lips and teeth, tongue and teeth as well as between tongue and various parts of the palate. Voiceless fricatives (f, s, ʃ, etc.) are formed in this way. If also the voice is excited, the corresponding voiced-pairs (v, z, ʒ) will be heard. With the sudden bursting of corresponding closure stops p, t, k, b, d, g, are obtained in a voiced-form. The rapid consecutive application of the closure- and gap-forming methods will result in new sounds, they are the so-called affricates, e.g. ts, tʃ. Also the timbre of noise sounds is influenced by the cavities above the glottis, but the real resonance effect is exercised first of all in the formation of vowels and semi-vowels (m, n, l, r, etc.).

Fullest information is known about the physical structure of vowels. Vowel-forming cavities transform the voiced sounds in their harmonic content as a result of their resonances (Figure 9.4). The regions of resonance amplification shown in the figure are called 'formants'. Helmholtz suggested a filter type of theory to explain the generation of vowels but this is replaced by a more modern 'tube theory' (Fant, 1960). Figure 9.2 shows a tube with varying cross-section leading from the glottis to the opening of the mouth, its length being about 17 cm. If the vowel-forming tube is considered as a system of one-quarter of a wavelength long, its fundamental tone will be around 500 Hz, that may, of course, be modified by the geometry of the tube (tongue, opening of the mouth). Indeed, the first formants are to be found between 200–1,000 Hz, and further resonances (formants 2, 3, etc.) may be found around 1,500 Hz, 2,500 Hz, etc. also with very wide modification possibilities.

The first two formants show the acoustical character of vowels rather well. The final acoustical form of speech sounds develops after radiating through the apertures of resonator cavities. The radiation resistance of the oral cavity

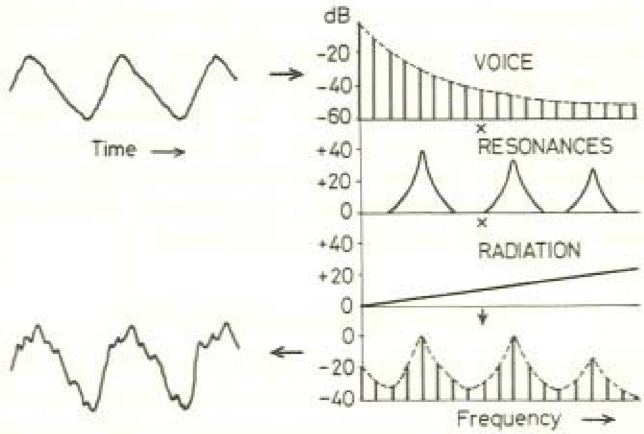

Figure 9.4 Evolution of vowel sounds. Upper left: cord tone, right: Fourier-spectrum of cord tone and modifying effects caused by transfer functions, *below right*: final spectrum, *left*: final oscillogram

raises the upper range of the spectrum by 6 dB/octave. Thus the hearing character of the upper formants will be stronger.

Therefore, the development of the acoustical character of vowels is a complicated process. Figure 9.4 gives an outline of the individual phases of the process. According to the system theory, resonance curves modify the original series of harmonics. The effect of radiation resistance already mentioned is superposed on this to give the final acoustical form. However, this may only be measured on the axis of the radiation. Laterally the radiation diagram is frequency dependent because of interferences and the shadowing effect of the head, respectively. The usual frequency spectrum of speech sounds is distorted laterally and at the back, thus making intelligibility more difficult.

Much less is known about the physical character of consonants but the formation of so-called semi-vowels is identical with that of vowels. In acoustical quality also elements of noise character may be observed and the formants of surrounding vowels also strongly influence the formation of the consonant character in time. The character of the consonant in continuous noise develops over a wide frequency band, but it may be quite well recognized and identified from analysis patterns. The most complicated mechanical recognition task is caused by stops. With them attempts are made to achieve some result by the 'locus' notion for the time being. Locus is a frequency place to which second — eventually third — formants of surrounding vowels are directed when transforming the cavities (Delattre, 1955).

With an appropriate formation of cavities a theoretically infinite number of vowel kinds may be created, but within one language usually only 5–15 vowel qualities are really used. Even if not exactly, but a nearly similar consideration holds also for consonants. From an information and theoretical viewpoint it is very important that out of the continuously changeable possibilities in very large numbers of discrete elements. This phenomenon is, of course, connected also distinguishing tone qualities. Most languages build up their vocabulary usually from 35–45 speech sounds.

9.3 SPEECH AS INFORMATION

The information content of a phonetic signal is around 5–5.5 bits. Since the speed rate is about 10–12 signals per second, the information capacity of speech is 50–60 bit/sec. The information content of a short sentence reaches 500 bit units. If our brain worked like a computer both the sending and the receiving intellect ought to make at least $2^{500} = 10^{150}$ decisions during this time for coding and decoding of the information, respectively. This data indicates how much faster the brain performs its evaluating work and how much more effeciently, than computers.

The question will be made even more complicated if we try to determine the information content of speech sounds analysed from the physical side by artifical recognition. A physical analysis of speech sound will not result in the determination of a single quality, but the analysis is extended to the pitch of sound, duration, intonation and resonance data of 3–4 formants, etc. The quality of sound should be determined on the basis of these permanently changed 8–10 data.

The solution of the problem has not been possible with our contemporary technical possibilities. However, by means of synthetization and transformation into acoustical signals of the same information data a well-understandable artificial speech may be generated. This fact points to the special activity of the brain in decoding speech information. Because of the overlapping of formant places dependent on pronunciation, decisions made on the basis of formants are not always ambiguous. Additionally, the individual timbre, the connection of the sound in question with other sounds and the emotional content of the text give a lot of extra information to the acoustical signal. Physical analysis is not able to distinguish between elements of the recorded information 'parcel'. It is an elementary observation, for example, that the loudness of speech may alter the formant structures more than would be characterized by the difference between two vowels. Physical analysis is not able to separate from the 'main information' the disturbing 'additional information' which sometimes necessarily seems more significant in characterizing the real quality, and thus, a simple analysis is not suitable for the mechanical recognition of quality. This 'parcel' character of the physical data of speech sounds is one of the important basic principles of research (Tarnoczy, 1965).

It results from the foregoing that the brain is likely to take considerably more data into consideration than is physically available when making a decision concerning quality. Besides, the brain makes not only a short-time analysis within a given sound, but permanently considers also relations with the preceding and the following sounds. What is more, it even compares the material perceived with its own lingual and intellectual vocabulary and corrects afterwards the eventually wrongly identified signal qualities. Thus, the work of the brain is enlarged further.

An interesting technical idea is that the automatic recognition and identification of quantized signals may be technically solvable. However, arising through the failure of the automatic recognition of speech sounds as signals, with the present technical level that is not possible for continuous signals. Speech itself is not a succession of information signals, but their total confluence out of which our signal recognition systems are unable to select discrete qualities. First of all, two things have yet to be solved: one is the segmentation of individual signals, while the other is their identification with elements of the quantized quality system. However, identification is not completely possible because of the suppressing effect of additional information, even if the possibility of correlation analysis is included in the solution.

Therefore, physical data of human communication have developed according to the abilities of man. The speed of formation and of understanding is about the same, and this determines the capacity of elementary information material and the information capacity of speech sound. It is obvious that the most open system types (Morse signals) could not correspond to human needs because of the time-duration of their decoding, while the closest ones (picture-writing) because of their absurdly large memory store. Therefore, speech elements had to develop by number and quality for psycho-physiological reasons.

The carrier of the information material of speech communication is always some series of physical signals. But, this is subject to the interference of environmental physical and biophysical phenomena in the course of their spreading, transformation, perception and even understanding. Only quantitative data of the information material transmitted may be measured, but not an evaluation of its contents. Yet, a relatively small phonetical change may involve a considerable deviation of its contents. Therefore, the stability of signals is a decisive problem of communication.

The informative effect of signals is not unambiguously influenced by various distortions. For example, the resonances of the oral cavity are, practically, distortions in the development of speech sounds, or in the dynamical compression of hearing in understanding. All this promotes the adequate development and reception of signals carrying information. As against this, interference by noise is always effective in the direction of the reduction of information content.

The signal-to-noise ratio (difference between signal and noise levels) is one of the most fundamental parameters in the efficiency of communication. In the understanding of speech sounds the judgement of so-called distinctive features (difference thresholds) has an important part and this judgement is made difficult by the masking effect of noise. It is a general rule of nature that the intelligibility of a series of signals masked by noise may be regained by increasing the redundancy of the carrier signals. Redundance may, for example, be increased by the multiplication of distinctive features of the individual elements, by the numerical increase of the series of signals or by the repetition of signal processes (phonetical or verbal redundance).

The next redundance possibility is in the length of the sound-signal series designating meanings and grammatical categories. The entire vocabulary of a language could be made up of sound relations of two, three and four elements. All combinations (e.g. four identical consonants one after the other) are, however, usually not made use of by languages in order to ensure intelligibility, i.e. information. Instead of this, longer words are formed. Longer signal series are less sensitive to noise, because the loss of one or another information element may be easily corrected in the brain and in this way the time for information is increased. Shortness in the length of signal increases the amount of information per unit time, but also increases the sensitivity to noise. Speech has developed in such a way that these two viewpoints are in the balance.

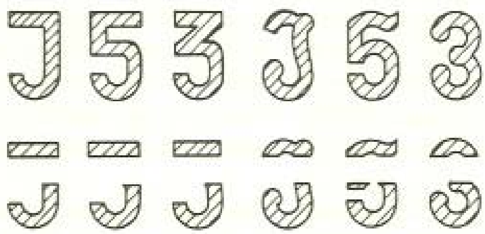

Figure 9.5 Example shows the conserving effect of redundancy. The distortion in the second row cannot eliminate the meaning of the redundant form

In Figure 9.5 the informational effect of two signal series are presented without and with noise interference, respectively. Interference has hardly any harmful effect on the information content of redundant series, while a perfectly informative signal series (containing no insignificant auxiliary signals) will become fully indecipherable (Tarnoczy, 1965).

With a large basic noise content, therefore, redundance should be increased in speech communication. For example, space language vocabularies do not allow the 'yes–no' version, but instead, the use of 'affirmative–negative' is obligatory (Webster, 1965).

9.4 ACOUSTICAL DATA OF SPEECH AND INTELLIGIBILITY

Resulting from the particular feature of speech sounds, acoustical performance and spectrum are permanently changing in the course of speech. In order to be able to determine acoustical data for the speech itself certain statistical considerations have to be made. It may be stated concerning a longer speech whether it was too still, or of normal intensity or too loud. The similar procedure may be followed also in the course of measurements. The average sound pressure level of normal speech is about 72–76 dB at a distance of 30 cm from the speaker. Measuring data under various circumstances may be found in Table 9.1. The table contains the average of the Hungarian speech of 18 men, measurements were made by the so-called speech-choir method (Tarnoczy, 1970, 1971).

Table 9.1 Average sound pressure level (dB) of man's speech at 30 cm distance from head

Speech	Front	Side	Behind
Murmured	63	—	—
Still	67	62	57
Normal	74	70	64
Loud	81	76	72
Shouted	86	—	—

Beside the average sound levels, the form of the average spectrum of speech has also to be known. In Figure 9.6 an average of eight European languages was indicated for man's and woman's voice. Data are plotted in spectrum level (energy level falling to 1 Hz theoretical bandwidth) and with a fluctuation possibility of ± 3 dB are valid for the English, German, Swedish, Russian, Italian, Hungarian (Tarnoczy, 1971) as well as Spanish (Banuls-Terol, 1965) and French (Tarnoczy, 1975) languages. In Figure 9.6 the zero level is the average (long-time) sound pressure level of speech.

From Figure 9.6 speech levels falling to octave bands of 250–4,000 Hz medium frequencies may easily be converted. These data are needed if we wish to calculate the intelligibility of speech in advance for noises of various spectral compositions.

Speech sounds are of various intensity and structure, therefore noise does not equally mask them, but the more intensive the noise the more it will mask. According to traditional definition, intelligibility is the numerical quotient of

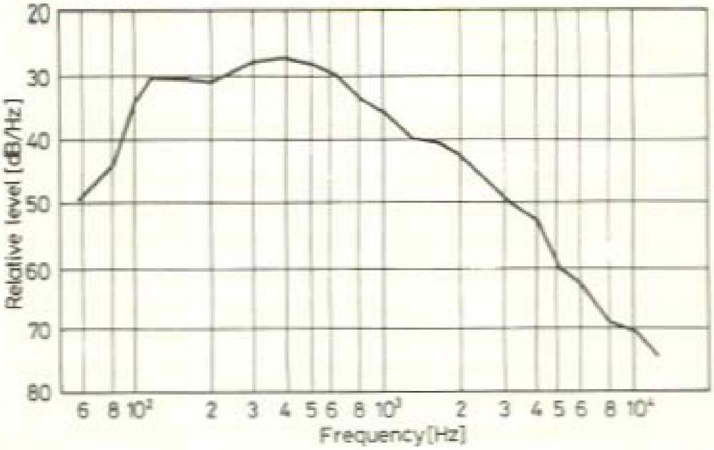

Figure 9.6 Average speech spectrum of European languages. All data fall within ±3 dB band of the curve plotted

the understood elements over all the speech elements (sound, syllable, word) communicated. However, intelligibility percentages determined on the basis of fluent speech are quite different from the values referred to individual syllables or separate (meaningful or meaningless) words (Fletcher, 1954). Namely, in the identification of meaningful texts very great deviations may be stated depending on the practice, individual abilities and intelligence as well as on the text. For example, with a 50% syllable intelligibility, speech intelligibility may reach 90%. Therefore, almost exclusively syllable intelligibility is used for investigations.

For direct measuring several procedures are known. Their common feature is that they they to find an answer to the intelligibility of mainly monosyllabic and disyllabic words with identical, or similar, voice structure as the language investigated. It is not absolutely necessary that the individual words have meaning. But, if so, then mistakable words are selected where, for example, changing one phoneme (speech sound occurring in the language in question) cause changes also in the meaning. For example, boon–coon–loon–moon–noon–soon.

The intelligibility percentage determined may be used for indicating the suitability for work of the acoustical surrounding. However, because of the clumsiness of subjective measuring some computation methods have developed also in the course of time.

One of them is the estimation method using the total noise level. Following from the data of Table 9.1 the masking effect of noise may be somewhat compensated by an increase of sound intensity. An important consideration is the relationship between spectra of speech and interfering noise. For example, 'white noise' (the energy density of spectrum is constant as a function of

Table 9.2 Percentage intelligibility stability of stops for the English and Hungarian languages

Sound	English	Hungarian
p	50.5	46.5
t	76.5	87.5
k	50	78.5
b	75	75
d	73	68
g	59	66.5

frequency) masks stops according to Table 9.2. The basic assumption of the data of the table is that the sound pressure level of speech and noise should be identical, i.e. the signal to noise ratio is 0 dB. In such cases the intelligibility of the individual sounds is called intelligibility stability (Miller and Nicely, 1955)

Under similar circumstances the intelligibility of vowels may reach even 98%. Therefore the exact knowledge of specral relations is needed. Fletcher (1954) gave possibilities for other solutions by introducing the term of 'articulation index'. The essence of articulation index lies in that the intelligibility, referring to the entire frequency range, is made up of partial intelligibilities achieved in the individual frequency bands. On the basis of this the action of speech interference level (SIL) (Beranek, 1960) and its modified forms were introduced.

Table 9.3 Admissible PSIL-values in dB mean values as a function of speech distance and sound intensity

Distance (m)	Normal	Raised voice	Loud	Shouted
0.3	68	74	80	86
0.6	62	68	74	80
0.9	58	64	70	76
1.2	56	62	68	74
1.5	54	60	66	72
1.8	52	58	64	70
3.6	46	52	58	64

The preferred speech interference level (PSIL) is the simple arithmetical mean value of the noise level to be measured in octave bands with 500, 1,000 and 2,000 Hz medium frequencies, respectively. As a function of this Table 9.3, according to Webster's personal communication (Burns, 1973), gives the admissible distance and sound intensity to be used for intelligibility in speech communication. The data are 3 dB higher than Beranek's figures for lack of influence by lip-reading.

The computations presented did not take spectral deviations of speech and interfering noise into consideration. The spectrum of speech may be assumed as given according to Figure 9.6. Of course, the spectral forms of speech with intensities deviating from the normal one are also known (Tarnoczy, 1971). The spectrum of the given noise should be compared to them. But, this is not enough to forecast intelligibility, since also partial intelligibility percentanges falling in the individual octave bands have to be known. Reference data are presented in Table 9.4 where partial values falling to the individual octave bands may be found for English (Fletcher, 1954), Russian (Jofe, 1954) and Hungarian (Tarnoczy, 1974) languages. The first two are the results from conversion, while the third is from direct measurement.

Table 9.4 Partial intelligibility percentages (X) measured in octave bands, without background noise

Band middle (Hz):	125	250	500	1k	2k	4k	8k
English	3	15	29	28	17	8	–
Russian	1	6	23	32	26	10	2
Hungarian	2	13	18	22	22	20	3
Mean values:	2	11	23	27	22	13	2

It follows from the table that the octave band with 4,000 Hz medium frequency largely contributes to the intelligibility. Therefore, level values of noise components should be determined in five octave bands. Furthermore, according to an idea of D. E. Broadbent (see Burns, 1973) we may agree that if some component of noise level is at least 30 dB below the speech level, then it has no effect on intelligibility, while if it is at least 20 dB above it, then understanding will be made quite impossible. Transitional cases are handled proportionally. Level data of the individual bands are weighted by partial percentages of intelligibility, and the partial results are then added. Thus the final intelligibility percentage will be obtained in the given noise.

In Figure 9.7 beside octave-band data of the average energy spectrum of speech of normal intensity, octave levels of an imagined noise source were also plotted. If the computation mentioned is made according to the summation of

$$\frac{(S-N)_{250}+20}{50}X_{250}+\frac{(S-N)_{500}+20}{50}X_{500}+\ldots$$

then an intelligibility of 51.4% will be obtained for the English language and that of 54.7% for the Hungarian one. For the verification of the computation use data of Table 9.4 and Figure 9.7.

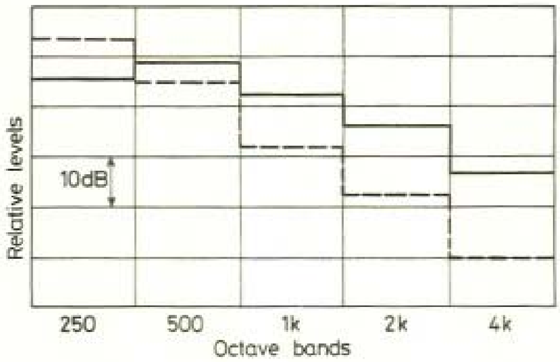

Figure 9.7 Diagram for computation of interfering effect of a given noise with speech, i.e. of intelligibility of syllables in a given language. Straight lines: speech spectrum, dashed lines: noise spectrum. See details in the text

9.5 REFERENCES

Banuls-Terol, V. (1965). *Proc. 7th Intern. Congr. on Acoust.*, Budapest, Vol. 3, 293–296.

Beranek, L. L. (1960). *Noise Reduction*, New York, McGraw-Hill.

Delattre, P. C., Lieberman, A. M. and Cooper, F. S. (1955). *J. Acoust. Soc. Amer.*, **27**, 769–773.

Fant, G. (1960). *Theory of Speech Production*, Mouton, 'S-Gravenhage.

Fletcher, H. (1954). *Speech and Hearing in Communication*, Van Nostrand, New York.

Jofe, V. K. and Janpolsky, A. A. (1954). *Tables and Graphics for Computation in Electroacoustics* (in Russian). Moscow-Leningrad.

Joos, M. (1948). Acoustic phonetics in *Suppl. to Language*, Baltimore.

Miller, G. A. and Nicely, P. E. (1955). *J. Acoust. Soc. Amer.*, **27**, 338–352.

Robinson, D. (1970). *Aeronautical Research Control*, C. P. 1112, London, HMSO.

Tarnoczy, T. (1951). *J. Acoust. Soc. Amer.*, **23**, 42–44.

Tarnoczy, T. (1965). *Rap. du 5e Congr.*, Internat. D'Acoust. Liege, Vol. 2, 371–387.

Tarnoczy, T. (1970). *Acustica*, **23**, 173–188.

Tarnoczy, T. (1971). *Acustica*, **24**, 57–74.

Tarnoczy, T. (1974). *Kep es Hangtechnika*, **20**, 97–105 (in Hungarian).

Tarnoczy, T. (1975). *Acoustique Telecomm*, Colloque FASE75, Paris, 123–132.

Tembrock, G. (1959). *Tierstimmen*, Wittenberg.

Timcke, R. (1956). *Z. Laryngologie*, Stuttgart **35**, 331–335.

van den Berg, Jw. (1958). *J. Speech and Hearing Res.*, **1**, 227–244.

Webster, J. C. (1965). *J. Acoust. Soc. Amer.*, **37**, 692–699.

Zwicker, E. (1985). Psychophysics in hearing (Chapter 4 in this book).