ELSEVIER

Speech Communication 34 (2001) 13–23

SPEECH COMMUNICATION

www.elsevier.nl/locate/specom

Optimized second-order gradient microphone for hands-free speech recordings in cars

Roland Aubauer *, Dieter Leckschat

Siemens AG, ICM CD TI 2, Frankenstrasse 2, 46395 Bocholt, Germany

Abstract

Hands-free telephony and automatic speech recognition in adverse acoustic conditions require noise reduction and speech enhancement methods to achieve acceptable quality and speech recognition rates. This contribution is primarily focussed on the optimization of the recording microphone. The directivity of the microphone can be used to reduce sound components outside the microphone's main axis which suppresses noise and room reverberation effectively. Considering the outer dimensions, gradient microphones possess a high directivity and good noise suppression properties. In this study, the feasibility and the noise-cancelling characteristics of an optimized second-order gradient microphone are investigated. The development of a simple sensitivity-matching principle allows the use of low-priced microphone capsules. The directivity index, characterizing the diffuse noise suppression, is nearly frequency-independent in the telephone bandwidth and is about 3 dB higher than that of conventional directional microphones used in telecommunication. The microphone is evaluated in an automobile environment with the help of a speaker-dependent single-word recognizer and is compared with lower-order gradient microphones. The noise robustness of the speech recognizer is improved by at least 4.7 dB, provided the microphone is in an optimum position. Conclusions are then drawn regarding such an optimum microphone positioning.

Zusammenfassung

Freisprechtelefonie und sprachgesteuerte Systeme gewinnen in allen Bereichen unseres Lebens an Bedeutung. Insbesondere für den Einsatz im Kraftfahrzeug werden Lösungen gesucht, die bei widrigen akustischen Bedingungen, wie z.B. hohen Störgeräuschpegeln, akustischen Mehrwegeausbreitungen und konkurrierenden Sprechern die Sprachverständlichkeit und die Spracherkennungsrate steigern. In diesem Beitrag wird ein optimiertes Gradientenmikrofon zweiter Ordnung mit vergleichbar geringen Abmessungen vorgestellt. Das realisierte Mikrofon verfügt über eine hervorragende nahezu frequenzunabhängige Richtcharakteristik. Schallanteile außerhalb der Mikrofonhauptrichtung werden stark gedämpft und Störgeräusche wirkungsvoll unterdrückt. Das die Unterdrückung von Diffusschall charakterisierende Bündelungsmaß des Mikrofons ist mit 8,8 dB um zirka 3 dB höher als das von derzeitig in Telekommunikationsanwendungen verwendeten Richtmikrofonen. Ein neuartiges und gleichzeitig einfaches automatisches Empfindlichkeitsabgleichverfahren erlaubt den Einsatz von preisgünstigen Mikrofonkapseln. Das Mikrofon wurde in einem Kraftfahrzeug untersucht und anhand der Erkennungsrate eines sprecherabhängigen Einzelwort-Spracherkenners mit Gradientenmikrofonen geringerer Ordnung verglichen. Durch den Einsatz des Mikrofons konnte die Störgeräuschrobustheit des Erkenners um mindestens 4,7 dB gesteigert werden. Es werden Schlüsse bezüglich der optimalen Anordnung von Freisprechmikrofonen in Kraftfahrzeugen gezogen.

* Corresponding author.
E-mail address: roland.aubauer@bch.siemens.de (R. Aubauer).

0167-6393/01/$ - see front matter
PII: S 0 1 6 7 - 6 3 9 3 (0 0) 0 0 0 4 3 - 1

Keywords: Automobile; Car; Directivity; Gradient microphone; Hands-free; Mel-frequency cepstral coefficients; Microphone; Multipath propagation; Noise; Quadrupole; Reverberation; Room impulse response; Room transfer function; Speech recognition; Telecommunication; Vehicle

1. Introduction

Many users demand a more natural and intuitive man-machine-interface that allows hands-free telecommunication, hands-free control and data input. Automatic speech recognition can be a helpful addition to keyboard, mouse and touch screen input devices. Applications are found in telecommunication (e.g. voice controlled name dialing, automatic information systems), in automobile and home control (control of non-safety functions in cars, household appliances, lighting installations, etc.) and equally in industry where voice data input and control functions are helpful in reducing user distraction and therefore contribute to efficiency, safety and convenience.

Today's speech recognition systems often lack robustness in adverse conditions where environmental noise and reverberation are interfering with the signal and where distortions are added during the signal transmission. Other speakers in the background competing with the user's voice challenge automatic speech recognition as well. The recognition rate in practical use is often much below the values yielded under ideal conditions in speech laboratories and is often not acceptable to the user. For command and control speech recognition applications, a recognition rate of 82% which is equal to 9 correctly recognized out of a total of 11 words is regarded as the absolute minimum requirement.

This article concentrates on the acoustic front-end of a speech recognition system and investigates the correlation between different gradient microphones and the recorded signal quality. The microphone positioning, the room acoustics and the environmental noise are taken into account. The test scenario is voice command and control in an automobile environment, one of the most promising applications for speech recognition. This assumption is confirmed by the imminent or already existing prohibition of phone calls while driving a car without a hands-free device, in many countries. Under safety aspects, hands-free telephony makes practically no sense without voice-controlled mobile phones.

In the following, the microphones, the test environment and the speech recognition system are described. Special focus is put on a new gradient microphone and on the microphone positioning in the vehicle. The driving noise and the room transfer function of the car are used to process recordings of a speech data base to simulate speech recordings with different microphones at different positions inside the driving car. The simulated speech is evaluated with a speech recognizer. The influence of the microphone directivity on the speech recognition performance is discussed and conclusions are drawn concerning an optimum microphone positioning.

2. First- and second-order gradient microphones

Microphones receiving sound only from one side of a membrane are pressure sensors. The directivity is omnidirectional as long as the microphone is small compared to the wavelength. Directivity is obtained when the sound affects both sides of the membrane so that the microphone becomes sensitive to the sound pressure difference (Zollner and Zwicker, 1993). Examples for such microphones are the dipole and the cardioid microphone. Both are first-order gradient microphones and are widely used in applications where noise suppression is necessary.

A second-order gradient microphone with higher directivity and better noise suppression is obtained when combining two cardioid microphones and subtracting their output signals from each other (Sessler and West, 1975; Olson, 1979). The principle of the so-called quadrupole microphone is based on the phase relation of the sound arriving at the microphones. The cardioid microphones are placed separately in the distance, Δx, and are oriented with their rotational axes in line. The output signals of the cardioid microphones are

$$x(t,\varphi) = \Gamma_c(\varphi)\cdot e^{j2\pi ft}, \tag{1}$$

$$y(t,\varphi) = \Gamma_c(\varphi)\cdot d\cdot e^{j2\pi f(t+(\Delta x/c)\cdot\cos\varphi)}, \tag{2}$$

where φ is the angle of sound incidence, $\Gamma_c(\varphi)$ is the directivity function of the cardioid microphone according to Table 1, c the sound velocity, f the frequency and d is a factor determining a sensitivity difference between the microphones. The difference signal, $s(t, \varphi)$, for frequencies $f \ll c/\Delta x$ is consequently (Fig. 1)

$$\begin{aligned} s(t,\varphi) &= x(t,\varphi) - y(t,\varphi) \\ &\approx \Gamma_c(\varphi)\cdot((1-d) \\ &\quad - j\cdot(d\cdot 2\pi f\cdot(\Delta x/c)\cdot\cos\varphi))\cdot e^{j2\pi ft}. \end{aligned} \tag{3}$$

The microphone difference signal, $s(t, \varphi)$, becomes

$$s(t,\varphi) \approx \Gamma_c(\varphi)\cdot\cos\varphi\cdot 2\pi f\cdot(\Delta x/c)\cdot e^{j(2\pi ft+\pi/2)}, \tag{4}$$

when both cardioid microphones are equally sensitive ($d=1$). The directivity of the quadrupole microphone, $\Gamma_q(\varphi)$ is as follows (see also Table 1):

$$\begin{aligned} \Gamma_q(\varphi) &= \Gamma_c(\varphi)\cdot\cos\varphi \\ &= 0.5\cdot(1+\cos\varphi)\cdot\cos\varphi. \end{aligned} \tag{5}$$

The directivity factor, γ, is defined as a function of the microphone directivity, $\Gamma(\varphi)$ (Zollner and Zwicker, 1993),

$$\gamma = \frac{A}{\oint_A \Gamma^2(\varphi)\,dA}; \qquad \text{e.g. } A = 4\pi r^2. \tag{6}$$

The directivity index, *di*, is calculated from the directivity factor, γ, and the logarithm

Table 1
Theoretical directivity of common gradient microphones up to the order of two[a]

	Omnidirectional	Dipole	Cardioid	Hypercardioid	Quadrupole
Directivity $\Gamma(\varphi)$	1	$\cos(\varphi)$	$0.5+0.5\cdot\cos(\varphi)$	$0.25+0.75\cdot\cos(\varphi)$	$0.5\cdot\cos(\varphi)+0.5\cdot\cos^2(\varphi)$
Directivity index *di*	0 dB	4.8 dB	4.8 dB	6 dB	8.8 dB

[a] The directivity of the quadrupole is the product of the directivities of cardioid and dipole. The directivity index gives the suppression of diffuse noise (arriving equally from all directions) in relation to the signal arriving from the microphone's main axis (0° direction).

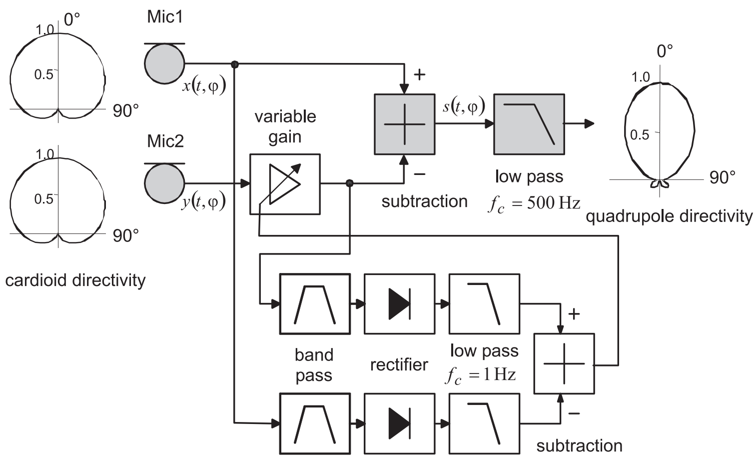

Fig. 1. Block diagram and directivity pattern of the cardioid and the quadrupole microphone. The diagram includes automatic sensitivity matching of the cardioid microphones. In gray, quadrupole microphone without automatic sensitivity control. The cutoff frequencies of the low-pass filters are denoted with f_c.

$$di = 10 \cdot \lg\gamma \text{ dB}. \quad (7)$$

The signal subtraction increases the directivity of the cardioid microphones by adding zeros in the directivity diagram at $\varphi = 90°$ and $\varphi = 270°$ sound incidence (cosine term in Eq. (4); see Fig. 1). According to Eq. (4), the frequency response below the first spectral notch at $f_{0n} = n \cdot c/\Delta x$ $(n = 1, 2, \ldots)$ is frequency-proportional and increases by 6 dB per octave. This, at the same time, is the usable frequency range, where the frequency response usually has to be equalized by first-order low-pass filtering. The cutoff frequency of the low-pass determines here the lower bound of the usable frequency range. The block diagram and the directivity diagram of the quadrupole microphone is shown in Fig. 1 whereas the frequency response is shown in Fig. 2.

In practice, sensitivity differences of the cardioid microphone capsules lower the directivity of the quadrupole microphone as a function of the frequency (see Fig. 3). With a sensitivity difference factor $d \gg 1$ or $d \ll 1$, the $(d - 1)$ term in Eq. (3) outweighs the imaginary term, $(d \cdot 2\pi f \cdot (\Delta x/c) \cdot \cos\varphi)$, and the directivity of the quadrupole is the same as that of the cardioid microphones. For a given realistic microphone sensitivity difference of 3 dB, the typical directivity

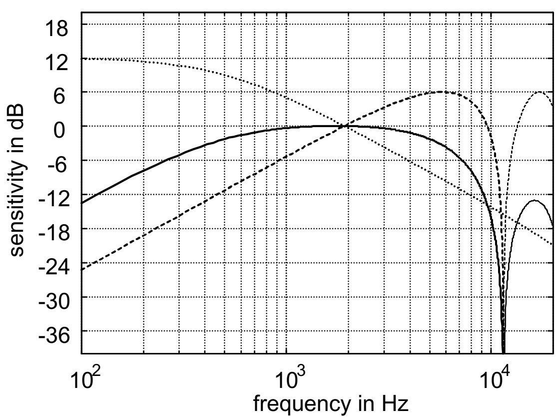

Fig. 2. Frequency response of the cardioid microphone difference signal, $s(t)$, for a microphone distance $\Delta x = 0.03$ m (dashed). Frequency response of the Butterworth low-pass filter with a cutoff frequency of 500 Hz and a + 12 dB gain (dotted). Frequency response of the resulting quadrupole microphone after low-pass filtering (line).

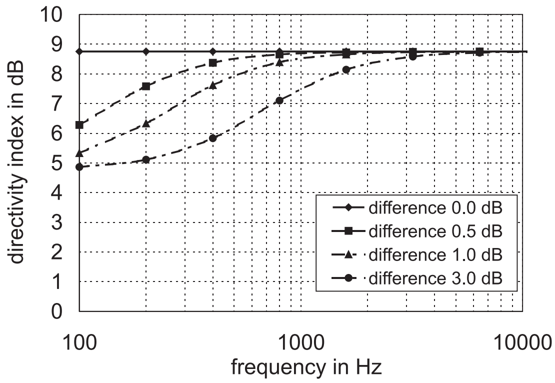

Fig. 3. Directivity index of the quadrupole microphone at various sensitivity differences of the cardioid microphone capsules as a function of the signal frequency.

of the quadrupole microphone is only developed at frequencies higher than 2000 Hz. Below 500 Hz, the directivity is close to that of the cardioid capsules, so that almost no gain in directivity and no noise suppression is achieved. This result is remarkable since noise suppression is of greatest importance in the lower speech frequency range where the distribution of typical environmental noise (e.g. automobile driving noise) has maximum levels. Precise sensitivity-matching of the cardioid microphones is therefore absolutely necessary.

Up to now, sensitivity-matching of microphone capsules is achieved either during the microphone production process or by selection or manual gain control of the microphone signals. All these methods are cost-intensive and have so far prevented the use of the quadrupole microphone on a broad market. This problem is solved by an automatic but simple sensitivity-matching principle which is based on the level differences of the microphone signals. Since the principle of the quadrupole microphone is exclusively based on the phase relation of the cardioid microphone signals, the envelope of the signals can be used for sensitivity-matching. The signal envelope is determined by signal rectification and low-pass filtering of the microphone signals (see Fig. 1, lower part). The low-pass ($f_c = 1$ Hz) determines the time-constant of the matching process. Long-time averaging here enables precise sensitivity-matching.

Considering the tolerance schemes of the cardioid capsules, it may become necessary to limit the sensitivity-matching to a certain frequency band by band-pass filtering (Fig. 1, band-pass). Typically, the band-pass covers about one octave and its center frequency is set between 300 and 500 Hz. Nevertheless, optimum results are achieved with microphone capsules whose frequency responses run about in parallel. In applications where high levels of low frequency noise dominate, e.g. speech recognition in automobiles, the band-pass can be omitted since the noise leads to an automatic sensitivity-matching in the desired frequency range. With the realized electronic circuit based on this method, the microphone sensitivity is balanced to a maximum deviation of ± 0.5 dB in the dominating frequency band, as long as the sound pressure level exceeds 65–70 dB SPL (SPL is sound pressure level re 20 μPa). Errors in the sensitivity-matching circuit have to be kept low as they result in a mismatch of the microphone capsule frequency responses that leads to a widened directivity of the quadrupole microphone. The sensitivity-matching is active while the microphone is in use and compensates for short- and long-term sensitivity drifts of the microphone capsules. The expenditure for the sensitivity-matching is low enough to offer an interesting overall cost level for the quadrupole microphone.

The realized quadrupole microphone with automatic sensitivity control has a significantly improved directivity over the full telephone bandwidth from 300 to 3400 Hz. With $di_{\mathrm{q}} \geqslant 8.0$ dB, the directivity index of the quadrupole microphone is at least by 2.0 dB higher than that of the optimum first-order gradient microphone with hypercardioid characteristic ($di_{\mathrm{hc}} = 6.0$ dB). The highly developed directivity of the quadrupole microphone produces an advantageous suppression of diffuse noise and reverberation and thus improves the quality of speech recordings. The microphone is therefore especially suited for automatic speech recognition in adverse acoustical conditions.

The present study compares four different microphones with omnidirectional, cardioid, hypercardioid and quadrupole characteristics. All microphones consist of electret microphone capsules having a diameter not larger than 10 mm. The hypercardioid microphone is hereby a special version optimized for surface boundary use.

3. Acoustical characteristics of the automobile

The acoustical characteristics of the automobile used in the test (A-Class Mercedes) are described by the room transfer function and by the driving noise at different speeds. The measurements of these features are used to simulate speech recordings of several speakers inside the car, that are evaluated by an automatic speech recognizer. This procedure has the advantage of defining constant sound recording conditions. Exactly the same speech utterances and driving noise patterns can be used for the microphone evaluation.

3.1. Room transfer functions

Four different microphone positions are used to investigate the influence of the sound propagation on the speech recognition. The selected positions on the dashboard (DA), at the A-column (AC), at the car top (CT), and at the interior light (IL) are on the passenger's side of the car and are shown in Fig. 4. The combined microphone and room transfer functions are determined via dummy head (Head Acoustics) and maximum-length sequences measurement technique for all combinations of microphones and positions (4×4). Here, the combined transfer function, $H(f)$, is the complex microphone signal spectrum, $S(f)$, over the equalized emitted signal spectrum, $P(f)$, that is measured at the mouth reference point of the dummy head where the artificial mouth is used as a sound source (ITU-T recommendation, 1996, p. 58),

$$H(f) = S(f)/P(f). \tag{8}$$

The distance artificial mouth-microphone is 0.19 m (CT), 0.28 m (IL), 0.36 m (AC) and 0.70 m (DA), respectively. Fig. 5 shows by means of examples the magnitude of the combined transfer functions of three microphones and two positions. The distinct notch patterns result from signal interference of the direct and indirect sound propagation, for example over the car's windscreen or side

Fig. 4. Microphone positioning in the A-class Mercedes. (DA) 'dashboard', (AC) 'A-column', (CT) 'car top' and (IL) 'interior light'.

windows. The notch pattern of a more directive microphone is less distinctive since sound which arrives from outside the microphone's main axis is attenuated. The reverberation time, T_{H60}, of the car can be determined from the room impulse response, $h(t)$, of the omnidirectional microphone and is about 65 ms.

It should be noted that the proximity effect increases the sensitivity of the gradient microphones at lower frequencies significantly, but only for small distances between the sound source and the microphone. The combined transfer function of the quadrupole microphone at the 'CT' position is almost perfectly flat due to this effect (see Fig. 5(e)). In the same way as frequencies are emphasized, the signal-to-noise ratio in the particular frequency band is improved.

3.2. Driving noise

The driving noise of the automobile at different speeds on an ordinary road (80 km/h) and on the autobahn (100 and 130 km/h) is alternately measured with all test microphones mounted in the selected positions. The microphones are cyclically rotated which ensures that only one microphone is at a certain position at a specific time. Simultaneously, a reference microphone close to the left ear of the dummy head records the driving noise on a second channel of the recording device. The mean reference noise level at a particular driving speed is later used to calibrate the level of the noise recordings of the test microphones, thus compensating for different wind speeds and road surfaces. Nevertheless, recordings are always made at the same road sections. The deviation from the mean reference noise level is 2 dB at its maximum.

As shown in Fig. 6 the driving noise spectrum is dominated by lower frequency components from air flow and tire noise. Below 1 kHz the noise spectrum level decreases to higher frequencies by 6 dB/octave, whereas above 1 kHz the spectrum level decreases faster by about 12 dB/octave. The spectra recorded at different speeds runs about in parallel. Therefore, noise at speeds not measured can be generated through appropriate amplification or attenuation. The A-weighted noise sound pressure level is determined by the reference microphone and ranges from 63 dB(A) SPL at 80 km/h, 67 dB(A) SPL at 100 km/h, 71 dB(A) SPL at 130 km/h to 74 dB(A) SPL at 160 km/h.

Table 2 shows the noise levels of the microphone output signals relative to the microphone sensitivity at 1 kHz. The lowest levels are always measured with the quadrupole microphone,

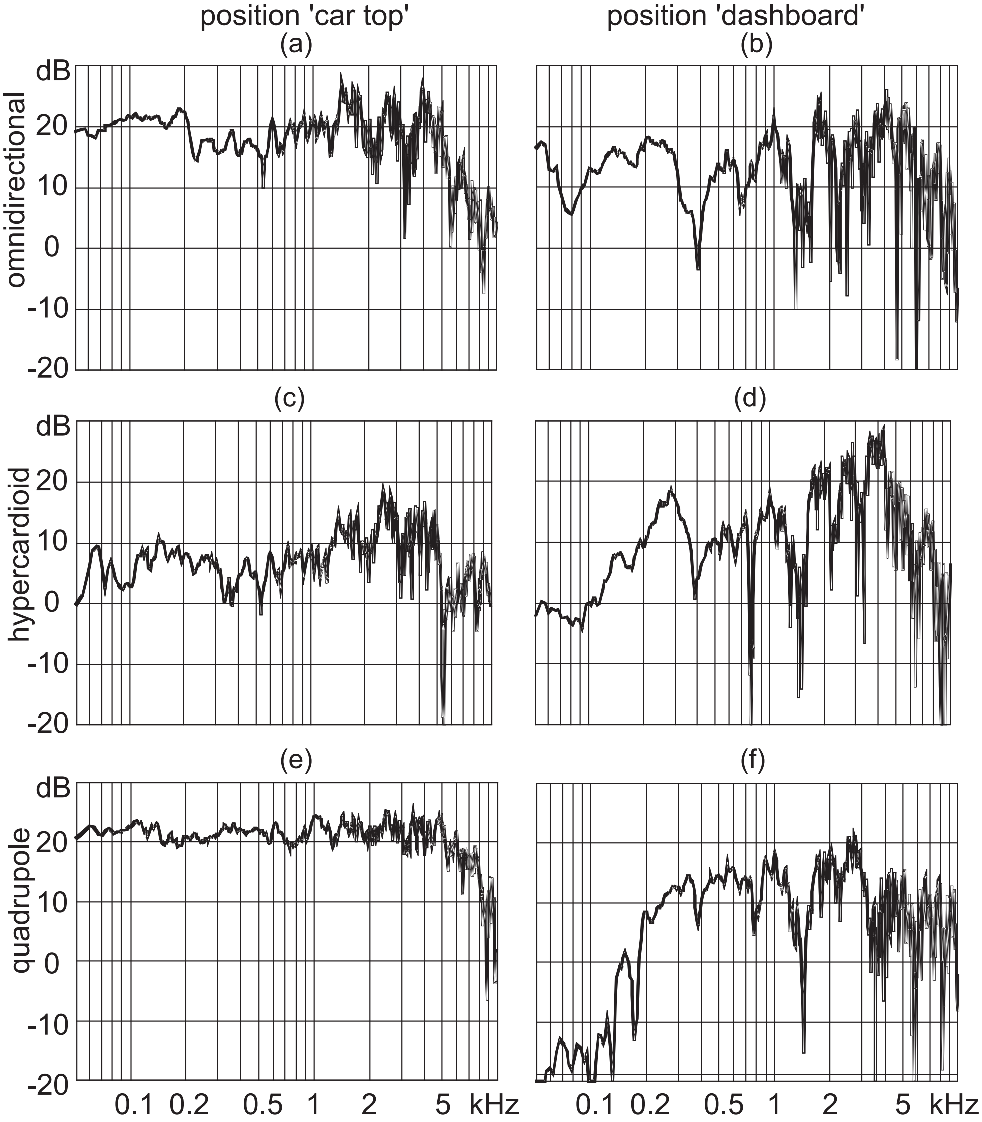

Fig. 5. Combined microphone–room transfer functions of the omnidirectional, the hypercardioid and the quadrupole microphone at the positions (CT) and (DA).

whereas measurements with the omnidirectional microphone show the highest noise levels. The microphone position ‘IL’ is particularly quiet and is therefore favorable. The ‘AC’, in contrast, is relatively loud. The quadrupole microphone attenuates noise almost frequency independent by up to 10 dB in relation to the omnidirectional microphone.

4. Speech recognition performance as a function of the microphone

The suitability of the selected microphones and their positioning for hands-free speech recordings in an automobile are evaluated by a speaker-dependent isolated-word recognizer. The recognizer includes automatic voice activity detection

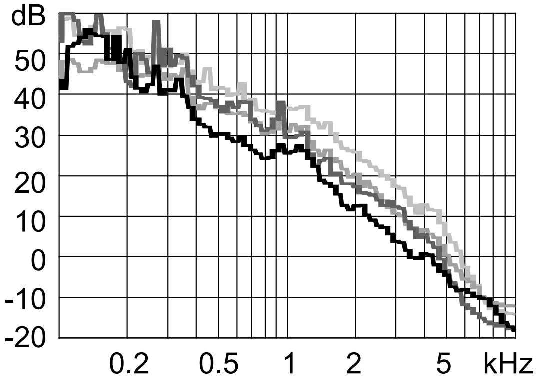

Fig. 6. Recorded driving noise spectra of different microphones (quadrupole (black line), cardioid (dark gray line), hypercardioid (gray line), omnidirectional microphone (light gray line)) mounted at the CT position at 100 km/h driving speed. The noise spectra levels in dB re $1/\sqrt{\mathrm{Hz}}$ are relative to the microphone sensitivity at 1 kHz.

applied to the noisy signal. The recognition is performed by comparing an input vector with templates of trained speech utterances in the vocabulary. The input vector consists of extracted input signal features based on Mel-frequency cepstral coefficients (MFCC). The comparison method uses the dynamic time warping (DTW) technique to enable the comparison of vector series with unequal lengths and varying durations within the series. The speech recognition system is trained using two utterances of each sample out of a 20-word vocabulary that contains the German digits 'zero' to 'nine' and German command words like 'start', 'stop', 'message', etc.

The recognition is tested using a third utterance of each word in the vocabulary. Utterances having a parameter distance to the training templates larger than a certain threshold are automatically rejected. The speech recognition rates are evaluated for three female and three male speakers.

The hands-free speech recording simulation is shown in Fig. 7. For the training of the recognizer it is assumed that the car is not in motion and therefore no driving noise is added to the speech samples that are convolved with the combined microphone–room impulse response. For recognition purposes, the driving noise of different speeds (80, 100, 130 km/h) is added after the convolution process. In addition, the standstill mode (no noise added) and speeds of 115, 150 and 170 km/h are simulated. Car noise at 115 km/h is generated by adding 2 dB to the noise recordings at 100 km/h. Car noise at 150 and 170 km/h is generated by adding 2 and 5 dB, respectively, to the noise recordings at 130 km/h. The speech samples are always filtered by a telephone bandpass before being applied to the speech recognizer. The Lombard effect is not taken into account since all speech utterances are recorded in a non-reverberant studio with a very low environmental noise level.

Table 2
A-weighted noise levels recorded with the given microphones at different positions inside the car and at different driving speeds[a]

Speed in km/h	Position	Noise level in dB(A) omnidirectional	Noise level dB(A) cardioid	Noise level dB(A) hypercardioid	Noise level dB(A) quadrupole
80	DA	65.4	61.8	63.3	57.1
	AC	66.8	62.9	64.5	59.7
	CT	65.5	60.0	66.0	60.8
	IL	64.4	59.8	60.2	55.8
100	DA	69.8	66.4	68.5	61.8
	AC	71.1	65.9	68.1	63.2
	CT	70.1	65.5	68.7	62.8
	IL	69.3	65.0	63.6	59.7
130	DA	73.8	70.5	70.8	65.0
	AC	75.7	70.0	70.8	67.7
	CT	74.1	69.6	73.9	67.5
	IL	74.6	68.4	68.7	64.6

[a] Highest noise levels at a certain position are marked dark, lower levels accordingly lighter. Noise levels are in dB(A) relative to the microphone sensitivity at 1 kHz (DA dashboard; AC A-column; CT car top; IL interior light).

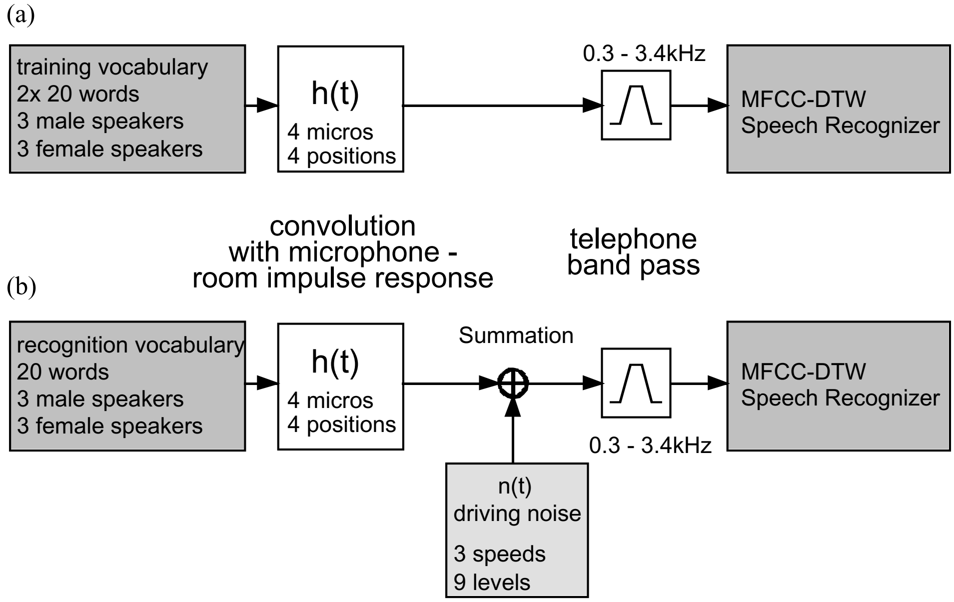

Fig. 7. Block diagram of the hands-free speech recording simulation. (a) Generation of the training speech samples taking the combined microphone–room transfer function into account. (b) Generation of the recognition speech samples taking both the combined microphone–room transfer function and the driving noise into account.

The recognition rates of the simulated microphone recordings are shown in Fig. 8 as a function of the reference car noise levels. The average speech sound pressure level at a distance of 0.5 m from the mouth is 70 dB(A), so that the mean A-weighted signal-to-noise ratio of the sound pressure would be at a virtual position, viz. 0.5 m in front of the mouth, 7 dB(A) at 80 km/h, 3 dB(A) at 100 km/h and −1 dB(A) at 130 km/h. The linear non-weighted signal-to-noise ratio is between 3 and 4 dB lower.

The word recognition rate is greater than 98% at a reference noise level of less than 50 dB(A) SPL with all microphones at all positions. At higher noise levels, the best results are always achieved with the quadrupole microphone, irrespective of its mounting position. With the quadrupole microphone, a gain in the noise robustness of the speech recognizer between 3.5 and 8.2 dB in relation to the omnidirectional microphone is achieved, and a gain between 0.9 and 4.8 dB in relation to the tested cardioid microphones. The gain here is the ratio between the reference noise levels at the point where the speech recognition rate is at 82% (see Fig. 8). The gain is always depending on the microphones' mounting position. The recognition rates of the first-order gradient microphones are relatively similar, although the cardioid microphone achieves higher recognition rates in the positions 'AC' and 'CT' than the hypercardioid microphone. It is interesting to note that compared to all other microphones the highest gain of the quadrupole is achieved at the position 'CT', the position with the closest microphone–mouth distance (MMD) of 0.19 m. The combination of directivity and proximity effect is here particularly beneficial.

In general, the MMD is the most important parameter for the microphone positioning since it influences the signal-to-noise ratio of the recorded signal directly. The highest recognition rates are achieved at the position 'CT' (MMD 0.19 m) and the lowest ones at the 'DA' (MMD 0.7 m). In between, roughly identical recognition rates are obtained at the 'AC' and at the 'IL' with a MMD of 0.36 and 0.28 m, respectively. For some microphones, better results are achieved at the 'IL'. The variability of the speech recognition performance

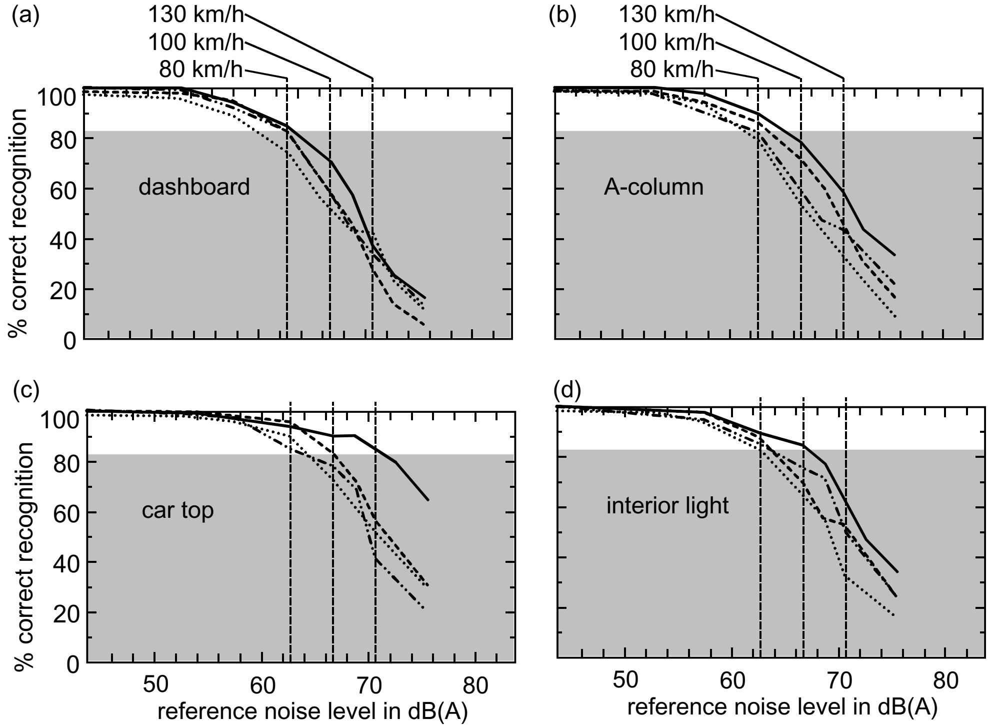

Fig. 8. Word recognition rates of simulated automobile speech recordings with various microphone types at different mounting positions (quadrupole microphone (filled line), cardioid microphone (dashed line), hypercardioid microphone (dashed and dotted line), omnidirectional microphone (dotted line).

as a function of the microphone position is expressed by an 8 dB noise level difference between the best and the worst position of the quadrupole microphone at the 82% recognition threshold.

Sufficient recognition rates higher than 82% are only achieved with the quadrupole microphone at the 'CT' position up to driving speeds of 130 km/h and at the 'IL' position up to speeds of 100 km/h. With no other microphone sufficient recognition rates are achieved at speeds faster than 100 km/h.

5. Discussion and conclusion

A second-order gradient microphone is investigated and compared with conventional omni- and unidirectional (cardioid) microphones. Basis for this investigation is the performance of an automatic speech recognizer in an automobile environment. The so-called quadrupole microphone is completed by an automatic sensitivity-matching electronic circuit through which extreme requirements can be avoided regarding the frequency response tolerance limits of ± 0.5 dB of the used microphone capsules. The directivity index of the realized quadrupole microphone is larger than 8.0 dB and therefore by at least 3.2 dB higher than that of an ideal cardioid microphone.

The study shows that the word recognition rate can be significantly improved in adverse acoustical conditions as a function of the microphone directivity. The quadrupole microphone suppresses noise by up to 5 dB better than the tested cardioid microphones and it suppresses noise even up to 10 dB better than the omnidirectional microphone. At the same time, the use of the quadrupole microphone improves the noise robustness of the tested speech recognizer by about the same factor, provided the microphone is in optimum position in the vehicle. In comparison with other noise reduction methods, e.g. the spectral subtraction method known from the field of statistical signal processing (Ephraim, 1992), the presented method adds no artifacts, like musical tones, neither to the recorded speech nor to the background signal. The presented results will therefore be generally applicable to all types of speech recognizers.

The measured driving noise levels are relatively independent from the microphone positioning (± 1.2 dB for the omnidirectional and ± 2.5 dB for the quadrupole microphone). Nevertheless, the noise level in the test car is remarkably low at the position 'IL' and notably high at the position 'AC'. Compared with that, the signal-to-noise ratio at the microphone and consequently the speech recognition rate is crucially influenced by the distance between microphone and speaker. Suitable microphone positions are therefore those which are close to the speaker's mouth. The best microphone position in this study is at the 'CT' in front of the speaker.

Nomenclature

φ	angle of sound incidence
Γ	directivity
γ	directivity factor
di	directivity index
j	imaginary unit ($\sqrt{-1}$)
π	pi

References

Ephraim, Y., 1992. Statistical-model-based speech enhancement systems. Proc. IEEE 80 (10), 1526–1555.

ITU-T Recommendation P.58, 1996. Head and torso simulator for telephonometry. International Telecommunication Unit, Telecommunication Standardization Sector, Series P: Telephone Transmission Quality, Objective Measuring Apparatus, Helsinki.

Olson, H.F., 1979. Directional microphones. An anthology of articles on microphones from the pages of the J. Audio. Eng. Soc. 1–27, (1953–1979), pp. 190–194.

Sessler, G.M., West, J.E., 1975. Second-order gradient unidirectional microphone utilizing an electret transducer. J. Acoust. Soc. Am. 58 (1), 273–278.

Zollner, M., Zwicker, E., 1993, Elektroakustik, Springer, Berlin, pp. 189–196.