ELSEVIER

Signal Processing 80 (2000) 1149–1166

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Innovative speech processing for mobile terminals: an annotated bibliography

André Gilloire*, Pascal Scalart, Claude Lamblin, Chafic Mokbel, Stéphane Proust

France Telecom CNET DIH/CMC, 2, Avenue Pierre Marzin, 22307 Lannion Cedex, France

Received 7 November 1997; received in revised form 12 January 1998

Abstract

This paper gives an overview of recent bibliographic references dealing with speech processing in mobile terminals. Its purpose is to point out state of the art issues in the area; thus a fairly large list of references taken from many conferences proceedings and journals is given and commented. General considerations about speech processing in mobile communications are firstly introduced; then we deal with audio processing for speech enhancement in mobile terminals and with low bit-rate speech coding. Speech recognition is addressed with some accent put on mobile applications. A short overview of implementation aspects of speech processing algorithms in mobile terminals is also given. Finally, open issues and problems are listed.

Zusammenfassung

Diese Arbeit gibt einen Überblick auf neue bibliografische Referenzen, die sich mit der Sprachverarbeitung in mobilen Einheiuten beschäftigen. Unser Ziel ist es, den Stand der technischen Veröffentlichungen auf diesem Gebiet zu ermitteln; somit ist eine recht lange Liste von Referenzen aus vielen Konferenzen und Zeitschiften und die zugehörigen Kommentare angegeben. Allgemeine Überlegungen zur Sprachverarbeitung in der Mobilkommunikation werden als erstes angeführt; anschließend behandeln wir Verfahren zur Spracheverbesserung in mobilen Systemen basierend auf niedrigen Bitraten. Bei der anschlieend behandelten Spracherkennung wird ein gewisser Schwerpunkt auf mobile Systeme gelegt. Ein kurzer Überblick zum Implementierungsaspekt der Sparchverarbeitungsverfahren in mobilen Endgeräten ist ebenfalls angegeben. Schlielich werden offene Themen und Probleme aufgezeigt.

Résumé

Cet article passe en revue la bibliographie récente concernant le traitement de la parole dans les terminaux mobiles. Son objectif est de mettre l'accent sur l'état de l'art dans le domaine; en conséquence une importante liste de références recueillies dans des actes de congrès et des revues est donnée et commentée. En premier lieu, des considérations générales sur le traitement de la parole dans les communications mobiles sont présentées; puis nous discutons de traitements de rehaussement de la parole dans les terminaux mobiles ainsi que du codage de la parole à bas débits. La reconnaissance vocale est traitée avec un accent particulier mis sur les applications mobiles. L'implantation des algorithmes de traitement

* Corresponding author.
E-mail address: andre.gilloire@cnet.francetelecom.fr (A. Gilloire).

0165-1684/00/$ - see front matter
PII: S0165-1684(00)00028-1

de la parole dans les terminaux mobiles est aussi brièvement traitée. Enfin, certaines questions et problèmes ouverts sont cités.

Keywords: Mobile telephony; Speech processing; Acoustic echo cancellation; Speech enhancement; Noise reduction; Speech coding; Speech recognition; Implementation

1. Introduction

The advent and wide dissemination of mobile radiotelephony has strongly increased the need of reliable wireless communication systems as well as the need of efficient speech processing techniques such as low bit rate speech coding, noise cancellation, acoustic echo control and robust speech recognition. During the last years, the literature has put growing emphasis on source coding and digital modulation issues for mobile telephony. Moreover, it has been soon recognized that speech communication provided by mobile services as well as access to these services by speech (e.g. voice dialing) needed specific adaptations and evolutions from earlier techniques as well as the development of new concepts. For example, mobile handsets are bound to be replaced by hands-free telephones when operated in cars, both for easier operation and for safety reasons; this fact led to focus research on new techniques like combined processing for acoustic echo cancellation and noise reduction and microphone arrays.

The main problems raised by speech processing in mobile telephony terminals are considered in some details in the paper. These problems can be briefly reviewed in the light of performance requirements for the mobile services users. The first requirement is that the quality of the transmitted speech which is picked up by the audio interface of the terminal must be high enough to provide the users with comfortable communication. Therefore, disturbances and distortions of speech such as background noise, echo, clipping and other signal degradations must be reduced to adequately low levels, possibly even cancelled (echo, noise). The reduction of disturbances is the task of front-end processings like acoustic echo and noise cancellation, associated with appropriate sound pick-up devices. In addition, speech must be encoded for transmission through the network at low bit rates since the radio links provide scarce bandwidth; the amount of speech degradations produced by the encoding schemes must be limited accordingly. Other requirements lie in the service aspects like voice dialing which can greatly improve easiness of use of mobile terminals; they typically rely on automatic speech recognition techniques that can be implemented either in the terminal itself or in the network. In both cases the challenge is to obtain acceptable recognition scores despite adverse and uncontrolled acoustic contexts and transmission channels.

Interoperation of mobile terminals through transmission networks needs standardization of critical parts of the speech processing. Speech coding (including auxiliary functions like discontinuous transmission, voice activity detection and comfort noise generation) is the most concerned one. Choice of front-end processing techniques is left free to the manufacturers, but standardization of performance requirements is needed to provide satisfactory voice quality. Choice of speech recognition techniques is presently left free; nevertheless possibilities which are being considered to share the required speech processing between the terminals and the network, would need some standardization.

The practical design of mobile terminals faces several challenges: small size, low-power consumption, easiness of use, and finally low cost. For that purpose, much effort has been done and is going on to integrate both baseband and radio functions in powerful – though cheap – VLSI devices. The availability of these devices is one major reason for the fast development of mobile communications which is currently observed. Speech processing functions which are parts of the baseband signal processing in mobile terminals (see Fig. 1), are generally implemented in customized digital signal processing chips (DSP).

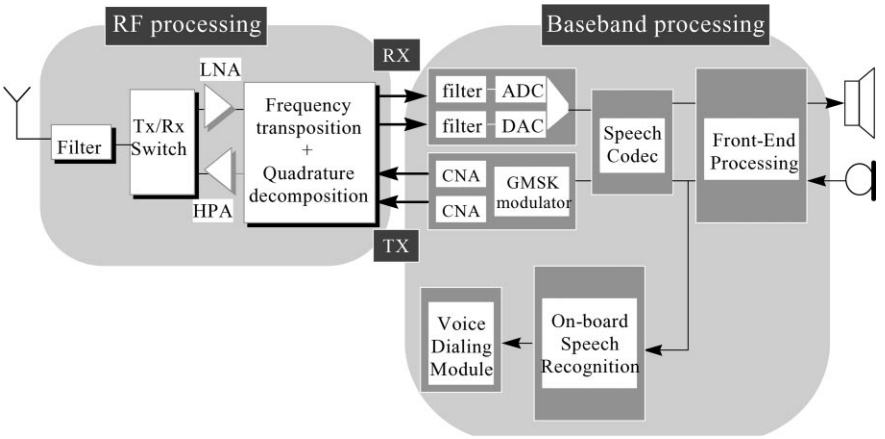

Fig. 1. Functional block-diagram of a mobile telephone terminal.

The paper is organized as follows. Section 2 addresses sound pick-up and speech processing techniques like acoustic echo cancellation and noise reduction bound to be implemented in mobile telephones, which purpose is to provide appropriate cleanliness of the speech signal even in adverse environments. In Section 3 an overview of low bit-rate speech coding techniques mainly developed for the GSM system is presented. Section 4 is dedicated to speech recognition. Implementation issues are briefly addressed in Section 5. Open issues are discussed in Section 6. A list of references organized according to the different topics addressed in the paper is provided at the end of the paper.

2. Sound pick-up and speech enhancement techniques

2.1. Acoustic problems in mobile terminals

The operation of hands-free as well as of hand-held mobile telephones may be impaired by the acoustic environment. Hands-free operation of vehicle-mounted mobile telephones during driving is now recommended or even made compulsory in many countries. However hands-free operation of mobile telephones involving microphones and loudspeakers located at some distance of the speaking/listening user is very sensitive to the car cabin's acoustical environment in terms of acoustic reflexions and background noise. Therefore, to obtain an acceptable speech quality, it is necessary to get rid both of the acoustic echo, which results from the acoustic coupling between the microphone and the loudspeaker, and of the background noise picked up by the microphone. Acoustic problems also occur in hand-held terminals. The acoustic transducers are separated by a short distance in many products, hence creating acoustic echo, and some amount of background noise is picked up by the microphone.

In this section we discuss various proposed acoustic echo cancellation algorithms and noise reduction techniques which can be used in hands-free mobile telephones.

2.2. Acoustic echo cancellation

In modern digital mobile radiocommunication systems such as the Pan-European GSM, elimination of the acoustic echo is of special concern, because inherent round-trip transmission delays of the order of 180 ms are quite common. With such delays, the echo becomes highly perceptible and high echo attenuation is needed in order to protect the interconnected telephone network from echo returning from the GSM network.

An effective way of reducing the echo annoyance is to use an acoustic echo canceller (AEC) which estimates the characteristics of the echo path by

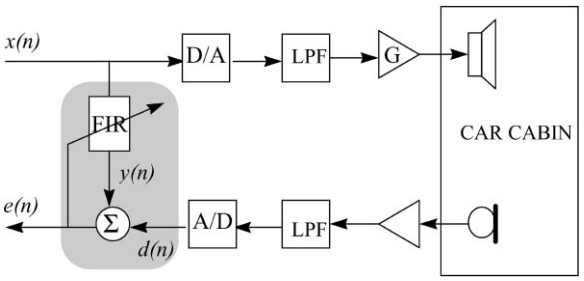

Fig. 2. Basic scheme of an acoustic echo canceller.

means of an adaptive filter (generally a FIR filter) and generates a replica of the echo signal, which is subtracted from the microphone signal to virtually eliminate the acoustic echo (Fig. 2).

Since the echo path is unknown and time-varying, adaptive filtering is a common way to obtain an estimate of the echo signal. Acoustic echo cancellers are a far better solution than echo control devices based on gain switching which are still used in many hands-free telephones, because they virtually provide full-duplex communication and avoid background noise switching effects.

Many algorithms have been proposed for reducing the acoustic echo in hands-free mobile telephones in cars. A substantial bibliography on these techniques can be found in [32,33]. Basically, considering acoustic echo cancellers, there are two families of algorithms for updating the tap weights of an adaptive filter, namely least mean squares (LMS) algorithms and recursive least squares (RLS) algorithms. Offsprings of these two families have been described e.g. in [22] in the context of acoustic echo cancellation.

Since the AEC algorithm should be implementable on a DSP chip it should exhibit reasonable complexity. Simple AECs based on time-domain LMS and Normalized LMS algorithms have been widely proposed for hands-free mobile telephones [19,36,39]. In [28] the NLMS and the second-order affine projection algorithm (APA2) are compared. Ref. [35] proposes an NLMS echo canceller associated with a sophisticated gain switching technique for general hands-free contexts. Furthermore, in order to cope with the presence of high levels of noise, adaptive algorithms for the computation of the adaptation step-size have been described [36,51].

Frequency domain implementations of the LMS like the multi-delay filter (MDF) and the Generalized MDF (GMDF) have also been proposed. These structures, based on the partition of the impulse response of the filter into sub-blocks, use small FFT size and thus exhibit small input/output delay [1,57]. A MDF algorithm with overlap has been evaluated in high background noise environments in [11].

Fast versions of the RLS (FRLS) with efficient numerical stabilization techniques and limited additional complexity have also been proposed: fast QR-RLS [10], fast transversal filter (FTF) [38]. Implementations in 16-bit and 24-bit fixed-point and 32-bit floating point arithmetics of the FTF algorithm have been adressed in [38]. However, for medium-sized filters (about 250 taps as discussed below) such as those encountered in mobile hands-free applications, FRLS algorithms still have high complexity. To overcome this problem, the use of a class of Newton-type algorithms known as fast Newton transversal filters (FNTF) has been proposed for mobile applications [55]. These algorithms have proved to be particularly well suited to speech since they assume low-order AR models for the input signals; moreover they exhibit better robustness to background noise than LMS-type algorithms.

Subband concepts have been introduced in AEC systems [10,20]. However, even if the analysis and synthesis filter banks satisfy the perfect reconstruction property, the aliasing components present after subband decomposition may be a serious problem for the identification of the acoustic echo path characteristics. Solutions to this problem can be provided by predicting adaptively neighbouring subbands signals [20] or by using highly selective all-pass based Power symetric QMF-IIR or aliasing cancellation QMF-FIR filter banks [63].

Comparative results in hands-free mobile environment between various AEC algorithms such as the NLMS, GMDF, etc, can be found in [53]. The echo reduction that can be achieved using an AEC is basically dependent upon the number of taps, the arithmetic precision used in the adaptive filter, and upon the characteristics of the acoustic impulse responses. Analysis of such impulse responses [19,38,60] in cars showed that a number of 256 taps

is typically needed for a sampling frequency of 8 kHz.

Satisfactory operation of AECs, especially in high background noise environments, needs reliable double talk detectors. [17] describes a detector which uses spectral and pitch cues, specifically designed for the context of mobile radiotelephones operating in cars.

It must be noted that opposite to speech coding discussed in the next section, acoustic echo control implementation is currently left free to the manufacturers of terminals. Only performance requirements are recommended, e.g. in [37].

2.3. Noise reduction

As stated earlier, one of the main problems arising in the operation of hands-free radiotelephones in cars is related to the high background noise that occurs when the car is riding [23]. Many techniques have been investigated in order to improve the overall quality and intelligibility of speech and to reduce the listener's tiredness in this context of high noise. In the sequel we classify the noise reduction techniques according to the number of input signals or channels that they use.

2.3.1. Multi-microphones techniques

Microphone arrays have attracted significant interest for speech pick-up in car environments [26,41]. They exploit spatial properties of the noise and speech sound fields which originate from various positions in the car. Several sensor array techniques can be used. In the usual structures of adaptive microphone arrays (also called adaptive beamformers), signals picked up by each microphone are filtered by optimal filters, whose outputs are combined to give an enhanced speech signal. In order to make the array work satisfactorily in the car interior context where the speaker is a source in the near-field subject to motions, preventive measures must be taken in order to avoid cancellation of the speaker's voice. A simple and straightforward method to control the optimisation of the filter is to constraint spatial filtering in the blocking matrix of the general sidelobe canceller, in order to select angular intervals over which the array is not allowed to cancel the target signal [54]. For fixed beamformers combined with post-processing treatments, an interesting choice of microphone positions can be found in [40]. Despite the fact that adaptive arrays are able to reduce the noise, they often introduce distorsions due to partial cancellation of the useful signal [27]. For that reason, and because a large number of microphones is still deemed impractical for consumer products, hands-free mobile manufacturers up to now focus on single or/and two microphone(s) solutions.

2.3.2. Two-microphones techniques

Several early investigations of two microphones speech enhancement systems in cars [24,25] have attempted to employ the adaptive noise cancellation (ANC) technique [58,69], which has proved to be efficient in a fighter cockpit environment [34]. In order to improve the performance of the basic system, concepts of sub-banding and multi-reference microphones [49,66,67] have been introduced. However, the ineffectiveness of ANC in noisy environments of car interiors has been observed and analyzed [12,25]; it is mainly due to the impossibility to prevent speech signals from entering both primary and secondary microphones.

Another way to reduce the influence of background noise is to exploit the differences between spatial coherences of speech and noise signals by using the coherence function between the two microphone signals. This method, first proposed for dereverberation purposes in [2], has been modified in order to compute an adaptive filter based on the magnitude squared coherence (MSC) between the two microphone signals. Extensions of this method have been proposed [42,44] for hands-free mobile applications in cars. Subjective tests have proved that this technique is able to remove the noise components in the high frequency band without introducing distorsions on the useful speech signal. However, low-frequency components of the noise with high levels still remain in the processed signal, because they exhibit significant coherence between the two microphones and thus are passed through the system.

2.3.3. Single microphone techniques

The far superiority in terms of speech quality of two-microphones techniques over single sensor

solutions has not yet been clearly demonstrated. Furthermore, for cost-effectiveness, algorithms should be implementable with reasonable cost on DSP chips. For these reasons, single microphone techniques still receive much attention from many research groups and from most manufacturers.

The "unimportance of phase in speech enhancement" was experimentally demonstrated in [68], where the authors showed that their enhancement system did not perform better if the "clean" (original) phase was used instead of the noisy one. However, it seems that this conclusion holds as long as local signal to noise ratio is at least about 6 dB [65]. Most single-microphone noise reduction techniques rely on the assumption of unimportance of phase. A common feature of these techniques is that the noise reduction operation can be related to the estimation of a short-time spectral suppression factor, which is evaluated for each frequency bin as a function of local signal to noise ratio estimators (see Fig. 3).

Many approaches, generally based on ad-hoc hypotheses, have been investigated in order to evaluate the short-time spectral suppression factor. They include well known algorithms like amplitude or power spectral subtraction [7,6,31,46,56], generalized Wiener filters [3,5,45], soft-decision estimation based on maximum likelihood (ML) estimation of the speech amplitude [8,52,70] or minimum mean square error (MMSE) estimation of each speech frequency component derived under a Gaussian assumption [9,13–16,30,59].

Because most single microphone speech enhancement techniques need to learn noise characteristics during speech pauses, they require a speech activity detection. Speech detection plays a key role in the performance of these noise reduction systems, hence in the output speech quality. Examples of such Speech/Pause detectors can be found in [43,61,62] where single- and two-microphones solutions have been proposed.

2.4. Combined acoustic echo cancellation and noise reduction

Combined systems performing simultaneously acoustic echo cancellation and noise reduction have been proposed [4,11,18,28,47,48,50,60]. Higher echo attenuation than the one provided by the echo canceller alone may be obtained owing to the Wiener-like operation of the noise reduction filter [50]. Since the annoyance of the residual echo signal is strongly dependent on the level of the background noise (this fact was recalled from experiments made in contexts of mobile communications issued from cars in [21]), the combined

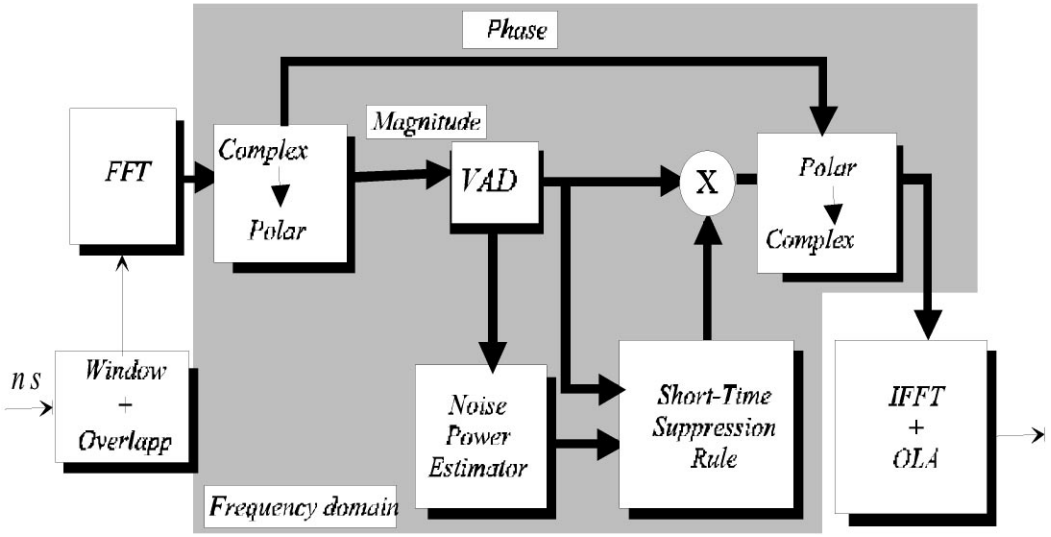

Fig. 3. Single-microphone noise reduction system based on spectral subtraction; ns: noisy speech, es: enhanced speech.

systems should be jointly optimized in order to take into account this dependence. Global approaches based on optimal filtering to reduce simultaneously echo and noise disturbances have been presented recently [4,29,64].

Generally speaking, noise reduction algorithms (and acoustic echo cancellation algorithms to a lesser extent) introduce additional delay which must be limited to avoid degradation of the conversational quality. The recent adoption within ETSI of an additional delay limit of 39 ms for hands-free processing has increased the GSM one-way delay budget by 40% and is likely to cause problems with end-to-end quality. Moreover, the Terminal Coupling Loss (overall echo attenuation between the receive input and the send output network interfaces) should be limited to the value of 40 dB for hands-free telephones. These constraints, added to the limitation of the computational complexity, make the design of acoustic echo cancellation and noise reduction algorithms still quite challenging.

3. Low bit-rate speech coding

Given the limited spectrum bandwidth allocated by the regulation authorities to the mobile services operators, digital mobile systems are one of the most important application areas for low bit rate speech coding. Some very specific constraints are imposed by the mobile context to the designers of speech coding algorithms, which make the design task highly challenging:

- Propagation conditions may change rapidly from good to very poor, with deep fading producing burst errors. If the current cell becomes saturated interferences between adjacent channels also produce transmission errors;
- mobile telephones are often used in high background noise environments;
- at the user end, speech coding is located in the mobile terminal where complexity is a key issue, because it impacts on cost and power consumption.

Many techniques can be used for speech encoding at low bit-rates. In this paper we put the accent on speech encoding techniques developed specifically for the pan-European GSM system, since the development of the GSM system has led to – and still demands – the development of a whole family of speech encoding techniques for different bit rates and quality levels. All speech codecs[1] for mobile applications (and for GSM in particular) operate in the time domain on blocks of the input signal called frames. The frame size has an influence on the overall transmission delay and is consequently limited to 20 ms in the GSM system. These codecs are based on the source-filter model with an analysis by synthesis structure [72]: an excitation signal is filtered through a long-term prediction filter to produce the harmonic structure and through a short-term prediction filter to produce the formantic structure (outline of the spectrum). The coded and transmitted parameters are: the excitation signal (positions and amplitudes of pulses for the first generation GSM RPE-LTP (regular pulse excitation with long-term predictor) codec or index of a codeword drawn from an excitation dictionary for more recent CELP (Code Excited Linear Prediction) codecs, parameters of the long term prediction filter (delay and gain), and coefficients of the short term prediction filter (generally 10 coefficients). The analysis by synthesis structure means that the excitation parameters are searched by synthesizing for each set of these parameters the corresponding synthesized signal and comparing it to the original one using a perceptual criterion (see Fig. 4).

The selected parameters are those that produce the synthesized signal which is closest to the original one with respect to the perceptual distance. Other techniques can be used as well for low bit-rate speech coding: harmonic or sinusoidal coders (IMBE (Improved Multiband Excitation), transform coding), etc.; [71] give an overview of these techniques.

The first-generation GSM codec (RPE-LTP) [73–75,79,80][2] needs a bit rate of 13 kb/s. Some additional bit rate (9.8 kb/s) is used for channel protection: a redundancy is added in the bitstream

[1] The term "codec" will be used in the sequel as an acronym for the overall speech coding–decoding algorithm.

[2] ETSI denotes the European Telecommunications Standards Institute.

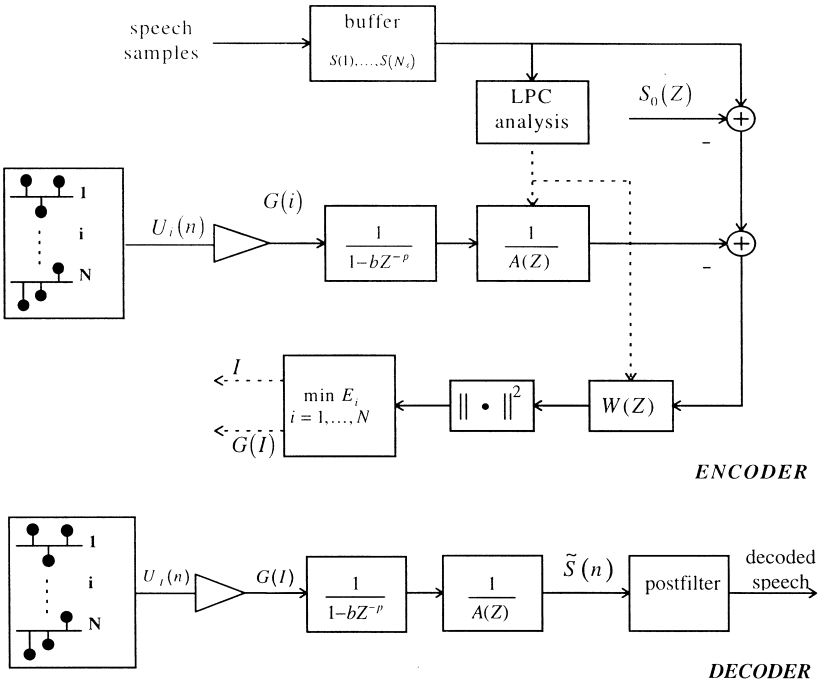

Fig. 4. Basic CELP codec scheme. $1/A(z)$: short-term synthesis filter, $1/(1-bz^{-P})$: long-term synthesis filter $W(z)$: perceptual filter, $S_0(z)$: synthesized signal from previous excitation and zero input on the current frame.

by a convolutional code and a CRC (Cyclic Redundancy Check). Interleaving of frames is also used to decorrelate as much as possible the burst of errors at the expense of 40 ms delay. The bits of the codec are classified in a few classes from the most to the less sensitive bits to transmission errors. The less sensitive bits are not protected.

With the strong increase of the number of GSM subscribers the first problem encountered was the need to increase the capacity of the system. This led to the standardization in 1994 of a "half-rate" system [81–84,88]. The selected speech coding algorithm was a VSELP (vector sum excited linear prediction) 5.6 kb/s codec based on the CELP technique (excitation signals are codewords of a dictionary which are all tested by an analysis by synthesis scheme; the particular VSELP technique uses a linear combination of basis vectors). This codec having 4 times the complexity of the RPE-LTP, yields the same level of quality as the "full rate" GSM codec for clean speech, but it is more affected by the background noise and by multiple transcodings that occur during mobile to mobile calls. Given this problem of quality, some uncertainty remains about the effective use of this codec in the GSM networks.

Both full-rate and half-rate codecs yield a speech quality that is very significantly below the "wire-line" usual telephone quality ranging from PCM (Pulse Code Modulation) G711 at 64 kb/s to ADPCM (Adaptive Differential Pulse Code Modulation) G726 at 32 kb/s. Since GSM subscribers use their mobile telephones more and more as their usual phones in car, office or even at home, the level of speech quality yielded by the GSM system is now found too low. In 1996 an enhanced GSM full-rate speech codec (called EFR) has been standardized [89–92,98]. The bit-rate needed by this codec is 12.2 kb/s. The channel protection is the same as the one used in the first generation with an additional 8 bits CRC. This enhanced codec provides wire-line quality for speech in good or medium channel conditions (which are the most frequent conditions in well-designed networks) but it does not perform

better than the first generation full-rate system in bad channel conditions. This codec has been also adopted for the American PCS1900 system. The EFR speech codec is a CELP codec of the ACELP (algebraic code excited linear prediction) family (algebraic codebooks requiring no storage of the codewords with focused searches) and it is related to the ITU-T[3] G729 standard at 8 kb/s [99,100]. Note that a derived version of the ITU standard G729 has been selected as the enhanced full rate codec for the mobile TDMA (Time Division Multiple Access) American standard IS136 (it replaces the first generation codec IS54 which was a VSELP codec) [97].

The reason for the limited robustness of the EFR codec is that, like all the other GSM codecs, it uses a fixed amount of channel protection whereas it operates in a wide range of channel conditions. In very good conditions, 10.6 kb/s for channel protection is too much, but it is insufficient in very poor conditions. Multi-rate coding schemes would help to optimize the use of the overall available channel capacity since they allow flexible distribution of the bit rate between the useful information (i.e. speech parameters) and the channel protection. In an early work [101] a multi-rate CELP codec was described, which operated at bit rates between 4.8 ad 8 kb/s. A new standardization work has been started at ETSI to design an adaptive multi-rate coding scheme, which should be completed (for the speech coding scheme itself) at the end of 1998, although some uncertainty remains on the feasibility of this scheme. The principle is that within one channel (half-rate or full-rate), the codec will be able to adapt its bit-rate and the corresponding channel protection according to the measured quality of the channel. Indeed, this particular operation of the codec will need to measure and to forward downlink information to the terminal on the current channel quality. The signalling and the channel quality indicator will be transmitted in-band. The speech coding rates could be for example 8, 6 and 4 kb/s for half rate channels and 12, 8 and 4 kb/s for full-rate channels. In order to achieve a capacity gain, the possibility to switch between full-rate channels and half-rate channels during a call is under study. The half-rate codec will be used only in good channel conditions since not enough protection is available at high error rates to guarantee a close to wire-line quality.

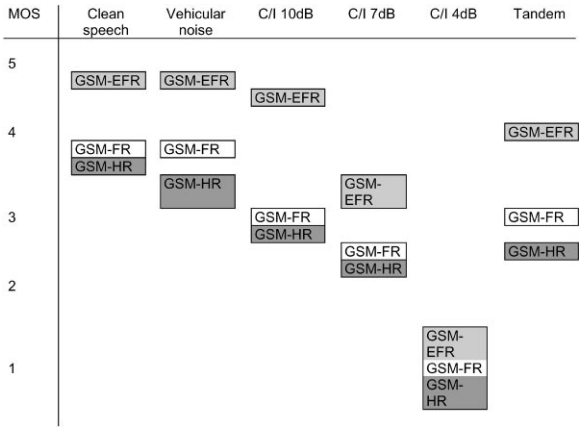

Fig. 5. Performance of GSM codecs in typical situations: FR: full-rate, HR: half-rate, EFR: enhanced full-rate.

Discontinuous transmission (DTX) has been designed and standardized for all GSM speech coding schemes. The purpose of DTX is to allow transmission of coded speech only during active voice periods in the communication, which results both in limitation of power consumption in the terminal and in increase of the system capacity by reducing co-channel interference. DTX needs efficient and reliable voice activity detectors. The coded signal is not transmitted during non-speech periods; only some bits (some LPC spectrum and energy parameters) are sent to generate a comfort noise at the decoder [76–78,85–87,93–96].

To conclude this section, the performance of the three GSM codecs in typical situations is shown in Fig. 5. It can be shown that robustness against transmission errors is still a critical problem (see Section 6).

4. Speech recognition in terminals

4.1. Automatic speech recognition basics

Basically, automatic speech recognition (ASR) aims to determine the pronounced words from an

[3] ITU-T denotes the Telecommunications Standardization Body of the International Telecommuications Union.

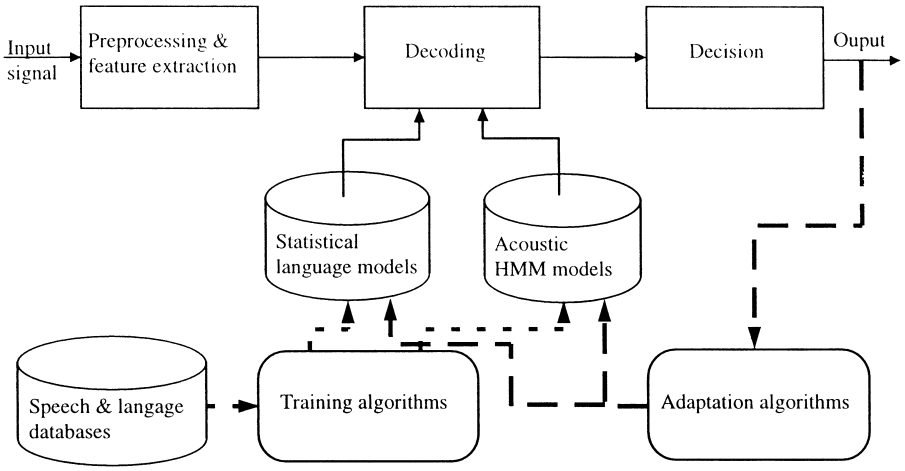

Fig. 6. Main blocks of an ASR system.

observed acoustic signal. ASR allows oral interaction between humans and machines, which is very useful to build applications with a user friendly interface. In telecommunications services, typical applications are voice activated dialing [123], directory assistance automation [108] and voice-activated interactive voice response servers [114]. These applications need that the system be able to recognize voice commands.

The Fig. 6 shows the main processing blocks generally found in an ASR system.

An acoustic signal is observed at the input of the preprocessing module. This latter one may be preceded by an echo canceller in order to allow the anticipation of the user's responses, which may be allowed while the system produces synthesized speech utterances (e.g. information and hints to the user) to improve the interaction with the recognition system. Techniques described in Section 2.2 can be used in order to perform echo cancellation. The signal observed at the output of the echo canceller is preprocessed and analyzed to provide a sequence of feature vectors. Preprocessing is generally used in order to reduce the perturbing components, i.e. additive noise and audio (telephone) channel effects. Noise reduction is performed using classical techniques such as the ones described in Section 2.3. However, these techniques are often tuned differently for speech recognition applications, and there is no need to compute the enhanced signal since only feature vectors are used for speech recognition. Considering channel effects on the speech spectrum, reducing this effect is equivalent to blind equalization of the channel transfer function. This equalization is generally based on long-term statistics of speech signals. For example, an adaptive filter is used in [111] where the filter parameters are adjusted in such a way that the equalized signal has a long-term spectrum equal to a predefined one. A speech endpoints detection is performed in order to decide which parts of the signal contain speech to be recognized. This detector is generally based on a local measure of the energy.

During the recognition process the feature vectors are decoded on the basis of statistical models compatible with the application (typically hidden Markov models (HMM) for acoustical modeling and N-grams models for language modeling), and the pronounced utterance is identified in the decision stage under the most likely hypothesis. Training algorithms are used to determine the parameters of the acoustical and language models [106,113] from appropriate databases. EM-type algorithms are generally used to train acoustical models. The parameters of the language models are generally estimated in order to minimize the perplexity (or the entropy) [105,113]. For large

vocabulary continuous speech recognition, the lexical description of the vocabulary words is a crucial task. Actually, for such applications, subword models (context dependent phoneme models) are generally used, and the lexical description must consider pronunciation variations (several possible transcriptions for a given word) and coarticulation effects. Indeed, such complex applications are up to now far above the capabilities of mobile terminals.

Adaptation algorithms may be used in order to increase the robustness of the ASR to variations in the recognition contexts; they adapt the ASR models parameters to the specific conditions of the application (speaker, ambient noise, telephone channel, etc.). Bayesian adaptation [110] and spectral transformation adaptation [115] may be used to perform this task. Adaptation techniques make use of the existing model of the application and of few examples of speech data collected in the target conditions to compute new estimate of the speech model parameters. The rejection of out of vocabulary (OOV) words or noises has also a great impact on performance and acceptability of ASR systems. Several techniques are used to perform rejection [109,122]. In general, "garbage" models are used to model the non-vocabulary words or utterances.

Looking at the Fig. 6, it can be easily noted that an ASR system has a high complexity and may require large memory and CPU resources. It is to be noted also that the whole processing must be done in real time or nearly real time. The computational and memory costs increase when going from applications involving isolated words and small vocabulary in a speaker-dependent mode to applications running with continuous speech and large vocabulary in a speaker independent mode. Specific normalization of the different variables should be done since the feature vectors have generally high rank and the likelihood computed may undergo quickly the machine precision. Such normalizations are generally limited when the Viterbi algorithm is used for training and recognition. Moreover, fast training algorithms must be used if the possibility to define on-line customized vocabulary (for personal applications like voice dialing) is given to the user. The corresponding extra cost makes the implementation of speech recognition in the terminals only reasonable for small vocabulary, isolated or connected words, and in speaker-dependent mode.

4.2. Speech recognition in mobile terminals

Already in the 1980s, several telephone terminals provided voice activated dialing. The ASR engine was based on the well-known "dynamic time warping" (DTW) algorithm. However, with the second generation of ASR systems based on HMM algorithms better performances could be achieved. Recently several telephone terminals integrating such ASR systems were introduced on the market. ASR finds a large application domain with the rapid development of the wireless networks and the size reduction of the wireless terminals. Several GSM terminals integrate an ASR module for voice activated dialing (such products were presented at the CEBIT'97 by well-known manufacturers). ASR may also be used in such terminals to voice activate some telecommunication services. As noted before, since ASR requires important memory and CPU resources, only small vocabulary isolated or connected words recognition may be integrated in a handset. On the opposite, large vocabulary ASR may be used through the telephone network, like in the automated directory assistance application. A list of recent available or announced mobile terminals and semiconductor chips that include speech recognition is given in Table 1.

Robustness to variations in the application environment is an important issue for the success of ASR in real life applications. Much research is going on to improve the robustness of ASR systems; as seen above, two classes of techniques are generally considered: preprocessing and adaptation. Preprocessing techniques aim to reduce the disturbing components in the observed signals while adaptation techniques aim to update the ASR parameters in order to better match the specific application environment. The main techniques in both classes are described in [107,112].

Preprocessing techniques like spectral subtraction and audio channel equalization are more suitable for implementation in the terminal. Currently available adaptation techniques have higher computational and memory costs. Moreover, preprocessing techniques (spectral subtraction for

Table 1
Mobile terminal products and chips including speech recognition (see [116,116–121] in the reference list)

Reference	Usage	Characteristics
Parrot [116]	Organizer/Pager Voice Activated directory Wireless Speech dialer	Speaker dependent 700 names in directory $300
General Motors & Hughes network systems [117]	Cellular phones (M6200) with voice dialing (names + digits) in some GM cars	Speaker dependent and independant
Philips Spark [18]	GSM mobile phone with voice dialing by name	Speaker dependent 10 names in directory
Nortel PCS1930 (announced on the web site of Nortel)	Mobile phone with voice dialing by name	Speaker dependent 20 names in directory
VPTI [119]	Professional voice organizer, Organizer/Pager voice dialing by name recognition of time for scheduling speech recognition to retrieve messages	Speaker dependent
Union Electric Corporation [120]	Telephone voicer autodialer	Speaker dependent 50 names in directory RSC-164 (Sensory Inc) microcontroller from speech recognition
Philips PR31100 UCB1100 chips [121]	Personal Digital Assistant (PDA) Personal Intelligent Communicators (PIC)	Speaker dependent Discrete models 100 names 38$

example) may also be used for speech enhancement. Thus, several DSP and terminals manufacturers are highly interested in developing such techniques [102,104,126]. Another aspect that highly interests these manufacturers is the reduction of the computational cost of the ASR algorithm [124,125].

To conclude this section, one can say that ASR can be fully implemented in a terminal if operating on small vocabulary and in isolated or connected words modes. The main challenge is to increase the robustness of the system to variations in the acquisition conditions of the speech signal (noise, channel effects, reverberation, echo, etc.). More complex ASR applications can be implemented within servers that can be accessed through the telephone network. For these latter applications, robustness remains also a crucial issue. Another solution may consist in distributing the tasks between the terminals and the servers. For example, feature extraction can be performed in the terminal and the decoding can be implemented in the server; however, such distributed processing needs standardization of information exchanged between the terminal and the server as well as some communication facilities. Achieving such a solution is the objective of the AURORA project proposed by several European manufacturers of GSM terminals [103].

5. Implementation aspects

The implementation of speech processing algorithms in mobile telephone terminals is a challenge since the required processing capability and memory tend to increase while power consumption must be held at a low level. Many papers have been published on that subject; the few references given in this section point out some currently active topics. New technologies and methodological aspects for low-power electronics are discussed in several papers: [129,130] propose reconfigurable

processing, i.e. dynamic reconfiguration of hardware modules, which is said to yield an order of magnitude of power reduction. The proposed architecture is centered around a reconfigurable communication network; it permits to match architectural granularity, to preserve the locality inherent in the algorithm and to use energy on demand. In [128] DSP cores for mobile communications are considered under the light of a "concept for application tailored signal processors". Major VLSI manufacturers now provide VLSI "kits" customized according to the requirements of manufacturers of mobile telephones. One can cite for example the DSP ST18950/D950 manufactured by SGS-Thomson [127], which can be customized to support speech enhancement algorithms as well as speech codecs.

Optimization can also be performed at the processing level. The different speech processing functions discussed in the previous sections may share some processing modules: for example, endpoints detection necessary for speech enhancement, echo cancellation and speech recognition may be implemented by the same module. The same speech enhancement algorithm could be used for improving both speech recognition performance and speech quality for transmission as well. Besides, combined echo cancellation and speech enhancement as mentioned in Section 2.4 may be more efficient than performing separately these two processings.

6. Open issues

The techniques currently available to perform speech processing in mobile telephone terminals suffer from well identified limitations. Improving these techniques and developing better ones yields many open issues.

Considering speech enhancement functions, new prospects are opened by combining acoustic echo cancellation and noise reduction in the same algorithm. Speech quality should also benefit from the adaptation of multi-sensor sound pick-up techniques to mobile environments.

Speech coding techniques still need much investigation effort. The main research directions concerning speech codecs for GSM applications can be listed as follows:

- design of a flexible coding scheme with a better adaptation of the channel protection according to the quality of this channel;
- improvement of the coding of speech mixed with background noise at strong levels (low and medium bit-rate CELP coders show difficulties to achieve satisfactory performance in such contexts);
- design of low bit-rate (4–5 kb/s) coding algorithms to obtain a good quality half-rate codec (a standardization work for a speech codec at 4 kb/s has started at the ITU-T but the constraints for GSM are stronger, especially as far as robustness to transmission errors is concerned);
- improvement of robustness to high error rates mainly based on three appraches: the joint optimization of channel and source coding, source decoding using channel decoding information, and improvement of the frame erasure recovery procedure.

Considering speech recognition for mobile applications, robustness in presence of noisy speech remains an open issue. Speech enhancement techniques and model parameters adaptation should be further studied and developed. Processing sharing between the network and the terminals is also an open issue.

7. Conclusion

The different sections of the paper have presented typical aspects of speech processing in mobile terminals. The focus was put on digital mobile telephones; no specific attention was paid to other kinds of mobile terminals like DECTs since the use and the network constraints (delay, bit rate) are thought to be less stringent. It is hoped that the annotated list of references which is provided in the paper will help the reader to get an overview of what is going on in the field of speech processing for mobile terminals.

Note: The ETSI documents listed in the speech coding section can be got from the ETSI Secretariat, F-06921 Sophia Antipolis CEDEX, France - secretariat@etsi.fr.

References

Acoustic echo cancellation and Noise reduction

[1] O. Ait Amrane, E. Moulines, M. Charbit, Y. Grenier, Low-delay frequency domain LMS algorithm, Proceedings of the International Conference on Acoustic Speech and Signal Processing, San Francisco, USA, 1992, pp. IV.9–IV.12.

[2] J.B. Allen, A. Berkley, J. Blauert, Multimicrophone signal processing technique to remove room reverberation from speech signals, J. Acoust. Soc. Am. 62 (4) (October 1977).

[3] L. Arslan, A. McCree, V. Viswanathan, New methods for adaptive noise suppression, Proceedings International Conference on Acoustic Speech and Signal Processing, Detroit, USA, 1995, pp. 812–815.

[4] B. Ayad, R. Le Bouquin-Jeannès, G. Faucon, Acoustic echo and noise reduction: a novel approach, Proceedings of the International Workshop on Acoustic Echo and Noise Control, London, September 1997, pp. 168–171.

[5] A. Akbari Azirani, R. Le Bouquin-Jeannès, G. Faucon, Speech enhancement using a Wiener Filtering under signal presence uncertainty, Signal Process. VIII: Theories Appl. Trieste (1996) 971–974.

[6] M. Berouti, R. Schwartz, J. Makhoul, Enhancement of speech corrupted by acoustic noise, Proceedings of International Conference on Acoustics Speech and Signal Processing (1979) 208–211.

[7] S.F. Boll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust. Speech Signal Process. ASSP-27 (2) (April 1979) 113–120.

[8] A. Brancaccio, C. Pelaez, Experiments on noise reduction techniques with robust voice detector in car environment, Proceedings of the EUROSPEECH'93, 1993, pp. 1259–1262.

[9] O. Cappé, Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor, IEEE Trans. Speech Audio Process. 2 (2) (April 1994) 345–349.

[10] F. Capman, J. Boudy, P. Lookwood, Acoustic echo cancellation using a fast QR-RLS algorithm and multirate scheme, Proceedings of the International Conference on Acoustics Speech and Signal Processing, Detroit, USA, 1995, pp. 969–972.

[11] F. Capman, J. Boudy, P. Lockwood, Acoustic echo cancellation and noise reduction in the frequency-domain: a global optimisation, Signal Process. VIII: Theories Appl. Trieste (1996) 29–32.

[12] N. Dal Degan, C. Prati, Acoustic noise analysis and speech enhancement techniques for mobile radio applications, Signal Processing 15 (1988) 43–56.

[13] G. Doblinger, Computationally efficient speech enhancement by spectral minima tracking in subbands, Proceedings of the fourth European Conference on Speech Communication and Technology, September 1995, pp. 1513–1516.

[14] Y. Ephraim, D. Malah, Speech enhancement using optimal non-linear spectral amplitude estimator, Proceedings of the International Conference on Acoustic Speech and Signal Processing, Boston, 1983, pp. 1118–1121.

[15] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. ASSP-32 (6) (December 1984) 1109–1121.

[16] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. ASSP-33 (2) (April 1985) 443–445.

[17] H. Ezzadi, J. Rouat, I. Bourmeyster, A new algorithm for double talk detection and separation in the context of digital mobile radio telephone, Proceedings of the International Conference on Acoustic Speech and Signal Processing, Munich, 1997, pp. 1897–1900.

[18] G. Faucon, R. Le Bouquin-Jeannès, Joint system for echo cancellation and noise reduction, Proceedings of fourth European Conference on Speech Communication and Technology, Madrid, 1995, pp. 1525–1528.

[19] M. Festa, V. Carmagnola, V. Lazzari, R. Montagna, An acoustic echo canceller for mobile hands-free telephone, Proceedings of the third International Workshop on Acoustic Echo Control, Plestin les Grèves, France, 7–8 September 1993, pp. 197–201.

[20] A. Gilloire, M. Vetterli, Adaptive filtering in subbands with critical sampling: analysis, experiments and application to acoustic echo cancellation, IEEE Trans. Signal Process. 40 (8) (August 1992) 1862–187.

[21] A. Gilloire, Performance evaluation of acoustic echo control: required values and measurement procedures, Ann. Télécommun. 49 (7–8) (1994) 368–372.

[22] A. Gilloire, E. Moulines, D. Slock, P. Duhamel, State of the Art in Acoustic Echo Cancellation, in: A.R. Figueiras-Vidal (Ed.), Digital Signal Processing in Telecommunications, Springer, Berlin, 1996, pp. 45–91.

[23] I. Goetz, Acoustic Noise environment for the GSM AMR codec, SMG-11, Heathrow, UK, 9–13 December 1996, SMG11 Tdoc 24/96.

[24] R.A. Goubran, H.M. Hafez, Background acoustic noise reduction in mobile telephony, Proceedings of the 36th IEEE International Conference on Veh. Technol. 1986, pp. 72–76.

[25] M.M. Goulding, J.S.Bird, Speech enhancement for mobile telephony, IEEE Trans. Veh. Technol. 39 (4) (November 1990) 316–326.

[26] Y. Grenier, M. Xu, An adaptive microphone array for speech input in cars, Proceedings of ISATA, Florence, 1990, pp. 485–492.

[27] Y. Grenier, A microphone array for car environments, Speech Commun. 12 (1993) 25–39.

[28] Y. Guelou, A. Benamar, P. Scalart, Analysis of two structures for combined acoustic echo and noise reduction, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Atlanta, USA, 1996, pp. 637–640.

[29] S. Gustafsson, R. Martin, Combined acoustic echo control and noise reduction based on residual echo estimation, Proceedings of the International Workshop on Acoustic Echo and Noise Control, London, September 1997, pp. 160–163.
[30] J. Häkkinen, M. Väänänen, Background noise suppressor for a car hands-free microphone, Proceedings of the fourth International Conference on Signal Processing Applications and Technology, Santa Clara, USA, 1993, pp. 300–307.
[31] P. Händel, Low-distortion spectral subtraction for speech enhancement, Proceedings of EUROSPEECH'95, Madrid, September 1995, pp. 1549–1552.
[32] E. Hänsler, The Hands-free telephone problem: an annotated bibliography, Signal Processing 27 (1992) 259–271.
[33] E. Hänsler, The hands-free telephone problem: an annotated bibliography update, Annales des Télécommun. 49 (7-8) (1994) 360–367.
[34] W.A. Harrison, J.S. Lim, E. Singer, Adaptive noise cancellation in a fighter cockpit environment, Proceedings of the International Conference on Acoustic Speech and Signal Processing, 1984, pp. 18.A.4.1–18.A.4.4.
[35] P. Heitkämper, M. Walker, Adaptive gain control and echo cancellation for hands-free telephone systems, Proceedings of European Conference on Speech Communication and Technology, Berlin, September 1993, pp. 1077–1080.
[36] A. Hirano, A. Sugiyama, A Noise-robust stochastic gradient algorithm with an adaptive step-size suitable for mobile hands-free telephones, Proceedings of the International Conference on Acoustic Speech and Signal Processing, Detroit, USA, 1995, pp. 1392–1395.
[37] Recommendation G.167: Acoustic echo control devices, ITU-T, Geneva, 1992.
[38] S.H. Jensen, Acoustic echo canceller for hands-free mobile radiotelephony, Signal Process. VI: Theories Appl. Brussels (1992) 1629–1632.
[39] M. Koya, M. Tsukamoto, Y. Iwata, K. Shimazu, Y. Fujiwara, A hands-free mobile telephone using echo canceller technique, Proceedings of the International Conference on Acoustic Speech and Signal Processing, 1986, pp. 32.4.1–32.4.5.
[40] K. Kroschel, K. Lange, Subband array processing for speech enhancement, Proceedings of the EUROSPEECH'93, 1993, pp. 621–624.
[41] K. Kroschel, A. Czyzewksi, M. Ihle, M. Kuropatwinski, Adaptive noise cancellation of speech signals in a noise reduction system based on a microphone array, 102nd AES Convention, Preprint 4450, Munich, Germany, 22–25 March 1997.
[42] R. Le Bouquin-Jeannès, G. Faucon, On using the coherence function for noise reduction, Proceedings V European Conference of the Signal Processing, Barcelone, 18–21 September 1990, pp. 1103–1106.
[43] R. Le Bouquin-Jeannès, G. Faucon, Voice activity detector based on the averaged magnitude squred coherence, Proceedings of the International Conference on Signal Processing Applications and Technology, October 1995, pp. 1964–1968.
[44] R. Le Bouquin, Enhancement of noisy speech signals: applications to mobile radio communications, Speech Commun. 18 (1996) 3–19.
[45] J.S. Lim, A.V. Oppenheim, Enhancement and bandwidth compression of noisy speech, Proceedings of the IEEE, 37 (12) (December 1979) 1586–1604.
[46] R. Martin, Spectral subtraction based on minimum statistics, Signal Process. VII: Theories Appl. Edinburgh, UK (1994) 1182–1185.
[47] R. Martin, P. Vary, Combined acoustic echo cancellation, dereverberation, and noise reduction: a two microphone approach, Ann. Télécommun. 49 (7–8) (1994) 429–438.
[48] R. Martin, J. Altenhoner, Coupled adaptive filters for acoustic echo control and noise reduction, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Detroit, USA, 1995, pp. 3043–3046.
[49] R. Martin, Design and optimization of a two microphone speech enhancement system, Proceedings of the Fourth European Conference on Speech Communication and Technology, Madrid, September 1995, pp. 2009–2012.
[50] R. Martin, P. Vary, Combined acoustic echo control and noise reduction for hands-free telephony – state of the art and perspectives, Signal Process. VIII: Theories Appl. Trieste (1996) 1107–1110.
[51] K. Mayyas, T. Aboulnasr, A robust variable step-size LMS-type algorithm: analysis and simulations, Proceedings of the International Conference on Acoustic Speech and Signal Processing, Detroit, USA, 1995, pp. 1408–1411.
[52] R.J. McAulay, M.L. Malpass, Speech enhancement using a soft-decision noise suppression filter, IEEE Trans. Acoust. Speech Signal Process. ASSP-28 (2) (April 1980) 137–145.
[53] P. Naylor, J. Alcazar, J. Boudy, Y. Grenier, Enhancement of hands-free telecommunications, Ann. Télécommun. 49 (7–8) (1994) 373–379.
[54] S. Nordholm, I. Claesson, B. Bengtsson, Adaptive array noise suppression of handsfree speaker input in cars, IEEE Trans. Veh. Technol. 42 (4) (November 1993) 514–518.
[55] T. Petillon, A. Gilloire, S. Theodoridis, The Fast Newton transversal filter: an efficient scheme for acoustic echo cancellation in mobile radio, IEEE Trans. Signal Process. 42 (3) (March 1994) pp. 509–518.
[56] P. Pollak, P. Sovka, J. Uhlir, Noise suppression system for a car, Proceedings EUROSPEECH, Berlin, Germany, 21–23 September 1993, pp. 1073–1076.
[57] J. Prado, E. Moulines, Frequency-domain adaptive filtering with application to acoustic echo control, Ann. Télécommun. 49 (7–8) (1994) 414–428.
[58] M.R. Sambur, Adaptive noise cancelling for speech signals, IEEE Trans. Acoust. Speech Signal Process. ASSP-26 (5) (October 1978), pp. 419–423.

[59] P. Scalart, J. Vieira Filho, Speech enhancement based on a priori signal to noise estimation, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Atlanta, USA, 1996, pp. 629–632.
[60] P. Scalart, A. Benamar, A system for speech enhancement in the context of hands-free radiotelephony with combined noise reduction and acoustic echo cancellation, Speech Commun. 20 (1996) 203–214.
[61] P. Sovka, P. Polläk, The study of speech/pause detectors for speech enhancement methods, Procedings of the fourth European Conference on Speech Communication and Technology, Madrid, 1995, pp. 1575–1578.
[62] P. Sovka, V. Davidek, P. Polläk, J. Uhlir, Speech/pause detection for real-time implementation of spectral subtraction algorithm, Proceedings of the International Conference on Signal Processing Applications and Technology, October 1995, pp. 1955–1959.
[63] O. Tanrikulu, B. Baykal, A.G. Constantinides, J.A. Chambers, Residual signal in subband acoustic echo cancellers, Signal Process. VIII: Theories Appl. Trieste (1996) pp. 21–24.
[64] V. Turbin, A. Gilloire, P. Scalart, C. Beaugeant, Using psychoacoustic criteria in acoustic echo cancellation algorithms, Proceedings of the International Workshop on Acoustic Echo and Noise Control, London, September 1997, pp. 53–56.
[65] P. Vary, Noise Suppression by spectral magnitude estimation – mechanism and theoretical limits, Signal Processing VIII (1985) 387–400.
[66] R.B. Wallace, R.A. Goubran, Parallel adaptive filter structures for acoustic noise suppression, Proceedings of the IEEE International Symposium on Circuits and Systems, San Diego, USA, 10–13 May 1992, pp. 525–528.
[67] R.B. Wallace, R.A. Goubran, Improved tracking adaptive noise canceler for nonstationary environments, IEEE Trans. on Signal Process. 40 (3) (March 1992) 700–703.
[68] D.L. Wang, J.S. Lim, The unimportance of phase in speech enhancement, IEEE Trans. Acoust. Speech Signal Process. ASSP-30 (4) (August 1982) 679–681.
[69] B. Widrow et al., Adptive noise cancelling: principles and applications, Proc. of the IEEE 63 (12) (December 1975) 1692–1716.
[70] J. Yang, Frequency domain noise suppression approaches in mobile telephone systems, Proceedings of the International Conference on Acoustic Speech and Signal Processing, Minneapolis, USA, April 1993, pp. II.363–II.366.

General speech coding techniques (Analysis-by-Synthesis, harmonic coding)

[71] C. Garcia-Mateo, D. Docampo-Amodeo, Modeling techniques for speech coding: a selected survey, in: A.R. Figueiras-Vidal (Ed.), Digital Signal Processing in Telecommunications, Springer, Berlin, 1996, pp. 1–43.
[72] P. Kroon, B.S. Atal, Predictive Coding of Speech Using Analysis-by-Synthesis Techniques, in: S. Furui, M. Sondhi (Eds.), Advances in Speech and Signal Processing, Marcel Dekker, New York, 1991, pp. 141–164.

GSM Full-rate codec

[73] GSM 06.01 Digital cellular telecommunications system; Full rate speech; Processing functions.
[74] GSM 06.10 Digital cellular telecommunications system; Full rate speech; Transcoding.
[75] GSM 06.11 Digital cellular telecommunications system; Full rate speech; Substitution and muting of lost frames for full rate speech traffic channels.
[76] GSM 06.12 Digital cellular telecommunications system; Full rate speech; Comfort noise aspects for full rate speech traffic channels.
[77] GSM 06.31 Digital cellular telecommunications system; Full rate speech; Discontinuous Transmission (DTX) for full rate speech traffic channels.
[78] GSM 06.32 Digital cellular telecommunications system; Voice Activity Detection (VAD).
[79] K. Hellwig, P. Vary, D. Massaloux, J.P. Petit, C. Galand, M. Rosso, Speech codec for the European Mobile Radio system, Proceedings of GLOBECOM 89, pp. 1065–1069.
[80] P. Kroon, E.F. Deprettere, R.J. Sluyter, Regular-pulse excitation: a novel approach to effective and efficient multipulse coding of speech, IEEE Trans. Acoust. Speech Signal Process. ASSP 34 (1986) 1054–1063.

GSM Half-rate codec

[81] GSM 06.02 Digital cellular telecommunications system; Half rate speech; Half rate speech processing functions.
[82] GSM 06.06 Digital cellular telecommunications system Half rate speech; ANSI-C code for the GSM half rate speech codec.
[83] GSM 06.20 Digital cellular telecommunications system; Half rate speech transcoding.
[84] GSM 06.21 Digital cellular telecommunications system; Substitution and muting of lost frames for half rate speech traffic channels.
[85] GSM 06.22 Digital cellular telecommunications system; Comfort noise aspects for half rate speech traffic channels.
[86] GSM 06.41 Digital cellular telecommunications system; Discontinuous Transmission (DTX) for half rate speech traffic channels.
[87] GSM 06.42 Digital cellular telecommunications system; Voice Activity Detection (VAD) for half rate speech traffic channels.
[88] I. Gerson, M. Jasiuk, A 5600 bps VSELP speech coder candidate for half-rate GSM, Proceedings of the EUROSPEECH, September 1993, pp. 253–257.

GSM Enhanced Full-Rate codec (EFR)

[89] GSM 06.51 Digital cellular telecommunications system; Enhanced Full Rate (EFR) speech processing functions; General description.

[90] GSM 06.53 Digital cellular telecommunications system; ANSI-C code for the GSM Enhanced Full Rate (EFR) speech codec.
[91] GSM 06.60 Digital cellular telecommunications system; Enhanced Full Rate (EFR) speech transcoding.
[92] GSM 06.61 Digital cellular telecommunications system; Substitution and muting of lost frames for Enhanced Full Rate (EFR) speech traffic channels.
[93] GSM 06.62 Digital cellular telecommunications system; Comfort noise aspects for Enhanced Full Rate (EFR) speech traffic channels.
[94] GSM 06.81 Digital cellular telecommunications system; Discontinuous Transmission (DTX) for Enhanced Full Rate (EFR) speech traffic channels.
[95] GSM 06.82 Digital cellular telecommunications system; Voice Activity Detector (VAD) for Enhanced Full Rate (EFR) speech traffic channels.
[96] D.K. Freeman, G. Cosier, C.B. Southcott, I. Boyd, The voice activity detector for the pan-European digital cellular phone mobile telephone service, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Glasgow, UK, May 1989, pp. 369–372.
[97] T. Honkanen, J. Vainio, K. Jarvinen, P. Haavisto, R. Salami, C. Laflamme, J.-P. Adoul, Enhanced full rate speech codec for IS136 digital cellular terminal, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, April 1997, pp. 731–734.
[98] K. Jarvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, J.-P. Adoul, GSM enhanced full rate speech codec, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, April 1997, pp. 771–774.
[99] R. Salami, C. Laflamme, J.-P. Adoul,, A. Kataoka, S. Hayashi, C. Lamblin, D. Massaloux, S. Proust, Description of the proposed ITU-T 8 kb/s speech coding standard, IEEE Workshop on Speech Coding for Telecommunications, Annapolis, USA, September 1995.
[100] R. Salami, C. Laflamme, B. Bessette J.-P. Adoul, Description of the ITU-T recommendation G.729 annex A: reduced complexity 8 kbit/s CS-ACELP codec, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, April 1997, pp. 775–778.

Multirate speech coding

[101] P. Kroon, K. Swaminathan, A high quality mutirate real-time CELP coder, IEEE J. Selected Areas Commun. 10 (5) (June 1992) 850–857.

Automatic Speech Recognition

[102] L. Arslan, A. McGree, V. Viswanathan, New methods for adaptive noise suppression, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Detroit, USA, 1995, pp. 812–815.
[103] AURORA Project for Distributed Speech Recognition, Briefing Paper for ETSI, 1997 (this document may be provided by Mr. Chris Ellis, project coordinator, ChrisEllis@BCS.org.uk).
[104] S. Dufour, C. Glorion, P. Lockwood, Evaluation of the Root-Normalised Front-End (RN-LFCC) for Speech Recognition in Wireless GSM Network Environments, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, 1996, pp. 77–80.
[105] F. Jelinek, UP FROM TRIGRAMS! The struggle for improved language models, Proceedings of the EUROSPEECH, 1991, pp. 1037–1040.
[106] D. Jouvet, Reconnaissance de mots connectés indépendamment du locuteur par des méthodes statistiques, Ph.D. Thesis, ENST, June 1988 (in French).
[107] C.-H. Lee, On feature and model compensation approach to robust speech recognition, Proceedings of the ESCA-NATO Workshop on Robust Speech Recognition Over Unknown Communication Channels, Pont-à-Mousson, France, 17–18 April 1997, pp. 45–54.
[108] M. Lening, G. Biebly, J. Massicotte, Directory assistance automation in Bell Canada: trial results, Speech Commun. J 17 (3-4) (November 1995) 227–234.
[109] L. Mauuary, Improving the performance of interactive voice response services, Ph.D. Thesis, Université de Rennes, January 1994 (in French).
[110] C.G. Miglietta, C. Mokbel, D. Jouvet, J. Monné, Bayesian adaptation of speech recognizers to field speech data, Proceedings ICSLP, 1996, pp. 917–920.
[111] C. Mokbel, D. Jouvet, J. Monné, Deconvolution of telephone line effects for speech recognition, Speech Commun. J. 19 (3) (September 1996) 185–196.
[112] C. Mokbel, L. Mauuary, D. Jouvet, J. Monné, C. Sorin, J. Simonin, K. Bartkova, Towards improving ASR robustness for PSN and GSM telephone applications, Proceedings of IVTTA, New Jersey, September 30–October 1, 1996, pp. 73–76.
[113] L. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, Prentice Hall, Englewood Cliffs, NJ, 1993.
[114] C. Sorin, D. Jouvet, C. Gagnoulet, D. Dubois, D. Sadek, M. Toularhoat, Operational and experimental French telecommunication services using CNET speech recognition and text-to-speech synthesis, Speech Commun. J. 17 (3-4) (November 1995) 273–286.
[115] T. Soulas, C. Mokbel, D. Jouvet, J. Monné, Adapting PSN recognition models to the GSM environment by using spectral transformation, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, 1997, pp. 1003–1006.
[116] "Voice Dialing Organizer Available in U.S.," Speech Recognition Update, n. 33, March 1996, p. 17.
[117] "GM Includes Voice-Activated Cellular Phone with some Models," Speech Recognition Update, n. 34, April 1996, p. 9.
[118] "Philips Selling Voice-Dialing Cellular Telephone," Speech Recognition Update, n. 47, May 1997, p. 8.
[119] "VPTI Ships New Consumer Products Using Speech Recognition," Speech Recognition Update, n. 41, November 1996, p. 18.

[120] "Sensory Recognition Chip to be Used in Telephone Dialer," Speech Recognition Update, n. 43, January 1997, p. 6.
[121] "Cost-Effective Chips for Portable Devices," Speech Recognition Update, n. 35, May 1996, p. 6.
[122] R.A. Sukkar, A.R. Setlur, M.G. Rahim, C.-H. Lee, Utterance verification of keyword strings using Word-Based Minimum Verification Error (WB-MVE) Training, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Atlanta, USA, 1996, pp. 518–521.
[123] G.J. Vystosky, VoiceDialing^{SM} – The first speech recognition service delivered to customer's home from the telephone network, Speech Commun. J. 17 (3-4) (November 1995) 235–248.
[124] T. Watanabe, K. Shinoda, K. Takagi, K. Iso, High speed speech recognition using tree-structured probability density function, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Detroit, USA, 1995, pp. 556–559.
[125] M. Yamada, H. Yamamoto, T. Kosaka, Y. Komori, Y. Ohora, Fast output probability computation using scalar quantization and independent dimension multi-mixture, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Atlanta, USA, 1996, pp. 893–896.
[126] R. Yang, P. Haavisto, An improved noise compensation algorithm for speech recognition in noise, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Atlanta, USA, 1996, pp. 49–52.

Implementation issues

[127] P. Blouet, The D950 core and its coprocessor concept, Proceedings of ICSPAT, 1995, pp. 800–804.
[128] G. Fettweis, DSP cores for mobile communications: where are we going? Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, 1997, pp. 279–282.
[129] Special issue on low power electronics, Proceedings of the IEEE 83 (4) (April 1995).
[130] J. Rabaey, Reconfigurable processing: the solution to low-power programmable DSP, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Munich, 1997, pp. 275–278.