Chapter 10

Transmission and Storage

10.1 Overview

Isabel Trancoso

Instituto de Engenharia de Sistemas e Computadores, Lisbon, Portugal
and Instituto Superior Tecnico, Lisbon, Portugal

This chapter is devoted to two closely linked areas of speech processing: coding and enhancement. For many years, these have been active areas of research, motivated by the increasing need for speech compression for bandlimited transmission and storage, and, on the other hand, for the need to improve the intelligibility of speech contaminated by noise.

In an age where the word gigabit became common when talking about channel or disk capacity, the aim of compression is not clear to everyone and one needs to justify it by describing the myriad of new applications demanding less and less bits per second and the rapidly expanding corpora.

Until the late seventies, research in speech compression followed two different directions: *vocoders* (abbreviation of voice coders) and *waveform coders*. The two approaches substantially differ in their underlying principles and performance. Whereas the first explore our knowledge of speech production, attempting to represent the signal spectral envelope in terms of a small number of slowly varying parameters, the latter aim at a faithful reproduction of the signal either in the time or frequency domains. They also represent two opposite choices in terms of the interleaving of the four main dimensions of the performance of speech coding: bit rate, speech quality, algorithm complexity and communication delay. Vocoders achieve considerable bit rate savings at the cost of quality degradation, being aimed at bit rates below 2 to 4 kbps (Tremain, 1982). For waveform coders, on the other hand, the preservation of the quality of the synthesized speech is the prime goal, which demands bit rates well above 16 kbps (Jayant & Noll, 1984). For an excellent overview of the main speech coding activities at the end of that decade, see Flanagan et al. (1979).

The next decade saw an explosion of work on speech coding, although most of the new coders could hardly be classified according to the waveform-coder/vocoder distinction. This new generation of coders overcame the limitations of the dual-source excitation model typically adopted by vocoders. Complex prediction techniques were adopted, the masking properties of the human ear were exploited, and it became technologically feasible to quantize parameters in blocks (VQ—vector quantization), instead of individually, and use computationally complex analysis-by-synthesis procedures. CELP (Schroeder & Atal, 1985) multi-pulse (Atal & Remde, 1982) and regular-pulse (Kroon, Deprettere, et al., 1986) excitation methods are some of the most well-known *new generation* coders in the time domain, whereas in the frequency domain one should mention sinusoidal/harmonic (Almeida & Silva, 1984; McAulay & Quatieri, 1986) and multi-band excited coders (Griffin & Lim, 1988). Variants of these coders have been standardized for transmission at bit rates ranging from 13 down to 4.8 kbps, and special standards have also been derived for low-delay applications (LD-CELP) (Chen, 1991). (See also Atal, Cuperman, et al., 1991 and Furui & Sondhi, 1991 for collections of extended papers on some of the most prominent coding methods of this decade.)

Nowadays, the standardization effort in the cellular radio domain that motivated this peak of coding activity is not so visible, and the research community is seeking new avenues. The type of quality that can be achieved with the so-called telephone bandwidth (3.2 kHz) is no longer enough for a wide range of new applications demanding wide-band speech or audio coding. At these bandwidths (5 to 20 kHz), waveform coding techniques of the sub-band and transform coding type have been traditionally adopted for high bit rate transmission. The need for 8-to-64 kbps coding is pushing the use of techniques such as linear prediction for these higher bandwidths, despite the fact that they are typical of telephone speech. The demand for lower bit rates for telephone bandwidth is, however, far from exhausted. New directions are being pursued to cope with the needs of the rapidly evolving digital telecommunication networks. Promising results have been obtained with approaches based, for instance, on articulatory representations, segmental time-frequency models, sophisticated auditory processing, models of the uncertainty in the estimation of speech parameters, etc. The current efforts to integrate source and channel coding are also worthy of mention.

Although the main use of speech coding so far has been transmission, speech encoding procedures based on Huffman coding of prediction residuals have lately become quite popular for the storage of large speech corpora.

The last part of this chapter covers an area closely related to coding and recognition, denoted as speech enhancement. The goal of speech enhancement is quality and/or intelligibility increase for a broad spectrum of applications, by (partly) removing the noise which overlaps with the speech signal in both time and frequency. The first noise-suppression techniques using only one microphone adopted single-filter approaches, either of the spectral-subtraction type or based on MAP or MMSE estimators. In the last few years, several pattern matching techniques have been proposed, neural networks have become quite popular as well and a number of robust parameterization methods and better metrics have emerged to improve the recognition of noisy speech. Multiple-microphone approaches can also be adopted in several applications. For an extended overview of enhancement methods, see Lim and Oppenheim (1979) and Boll (1991).

10.2 Speech Coding

Bishnu S. Atal & Nikil S. Jayant

AT&T Bell Laboratories, Murray Hill, New Jersey, USA

Coding algorithms seek to minimize the bit rate in the digital representation of a signal without an objectionable loss of signal quality in the process. High quality is attained at low bit rates by exploiting signal redundancy as well as the knowledge that certain types of coding distortion are imperceptible because they are masked by the signal. Our models of signal redundancy and distortion masking are becoming increasingly more sophisticated, leading to continuing improvements in the quality of low bit rate signals. This section summarizes current capabilities in speech coding, and describes how the field has evolved to reach these capabilities. It also mentions new classes of applications that demand quantum improvements in speech compression, and comments on how we hope to achieve such results.

Vocoders and Waveform Coders

Speech coding techniques can be broadly divided into two classes: waveform coding that aims at reproducing the speech waveform as faithfully as possible and vocoders that preserve only the spectral properties of speech in the encoded signal. The waveform coders are able to produce high-quality speech at high enough bit rates; vocoders produce intelligible speech at much lower bit rates, but the level of speech quality—in terms of its naturalness and uniformity for different speakers—is also much lower. The applications of vocoders so far have been limited to low-bit-rate digital communication channels. The combination of the once-disparate principles of waveform coding and vocoding has led to significant new capabilities in recent compression technology. The main focus of this section is on speech coders that support application over digital channels with bit rates ranging from 4 to 64 kbps.

10.2.1 The Continuing Need for Speech Compression

The capability of speech compression has been central to the technologies of robust long-distance communication, high-quality speech storage, and message encryption. Compression continues to be a key technology in communications in spite of the promise of optical transmission media of relatively unlimited bandwidth. This is because of our continued and, in fact, increasing need to use band-limited media such as radio and satellite links, and bit-rate-limited storage media such as CD-ROMs and silicon memories. Storage and archival of large volumes of spoken information makes speech compression essential even in the context of significant increases in the capacity of optical and solid-state memories.

Low bit-rate speech technology is a key factor in meeting the increasing demand for new digital wireless communication services. Impressive progress has been made during recent years in coding speech with high quality at low bit rates and at low cost. Only ten years ago, high quality speech could not be produced at bit rates below 24 kbps. Today, we can offer high quality at 8 kbps, making this the standard rate for the new digital cellular service in North America. Using new techniques for channel coding and equalization, it is possible to transmit the 8 kbps speech in a robust fashion over the mobile radio channel, in spite of channel noise, signal fading and intersymbol interference. The present research is focussed on meeting the critical need for high quality speech transmission over digital cellular channels at 4 kbps. Research on properly coordinated source and channel coding is needed to realize a good solution to this problem.

Wireless communication channels suffer from multipath interference producing error rates in excess of 10%. The challenge for speech research is to produce digital speech that can be transmitted with high quality over communication networks in the presence of up to 10% channel errors. A speech coder operating at 2 kbps will provide enough bits for correcting such channel errors, assuming a total transmission rate on the order of 4 to 8 kbps.

The bit rate of 2 kbps has an attractive implication for voice storage as well. At this bit rate, more than 2 hours of continuous speech can be stored on a single 16 Mbit memory chip, allowing sophisticated voice messaging services on personal communication terminals, and extending significantly the capabilities of digital answering machines. Fundamental advances in our understanding of speech production and perception are needed to achieve high quality speech at 2 kbps.

Applications of wideband speech coding include high quality audioconferencing with 7 kHz-bandwidth speech at bit rates on the order of 16 to 32 kbps, and high-quality stereoconferencing and dual-language programming over a basic ISDN link. Finally, the compression of a 20 kHz-bandwidth to rates on the order of 64 kbps will create new opportunities in audio transmission and networking, electronic publishing, travel and guidance, teleteaching, multilocation games, multimedia memos, and database storage.

10.2.2 The Dimensions of Performance in Speech Compression

Speech coders attempt to minimize the bit rate for transmission or storage of the signal while maintaining required levels of speech quality, communication delay, and complexity of implementation (power consumption). We will now provide brief descriptions of the above parameters of performance, with particular reference to speech.

Speech Quality: Speech quality is usually evaluated on a five-point scale, known as the mean-opinion score (MOS) scale, in speech quality testing—an average over a large number of speech data, speakers, and listeners. The five points of quality are: *bad, poor, fair, good,* and *excellent.* Quality scores of 3.5 or higher generally imply high levels of intelligibility, speaker recognition and naturalness.

Bit Rate: The coding efficiency is expressed in bits per second (bps).

Communication Delay: Speech coders often process speech in blocks and such processing introduces communication delay. Depending on the application, the permissible total delay could be as low as 1 msec, as in network telephony, or as high as 500 msec, as in video telephony. Communication delay is irrelevant for one-way communication, such as in voice mail.

Complexity: The complexity of a coding algorithm is the processing effort required to implement the algorithm, and it is typically measured in terms of arithmetic capability and memory requirement, or equivalently in terms of cost. A large complexity can result in high power consumption in the hardware.

10.2.3 Current Capabilities in Speech Coding

Figure 10.1 shows the speech quality that is currently achievable at various bit rates from 2.4 to 64 kbps for narrowband telephone (300–3400 Hz) speech. The intelligibility of coded speech is sufficiently high at these bit rates and is not an important issue. The speech quality is expressed on the five-point MOS scale along the ordinate in Figure 10.1.

PCM (pulse-code modulation) is the simplest coding system, a memoryless quantizer, and provides essentially transparent coding of telephone speech at 64 kbps. With a simple adaptive predictor, adaptive differential PCM (ADPCM) provides high-quality speech at 32 kbps. The speech quality is slightly inferior to that of 64 kbps PCM,

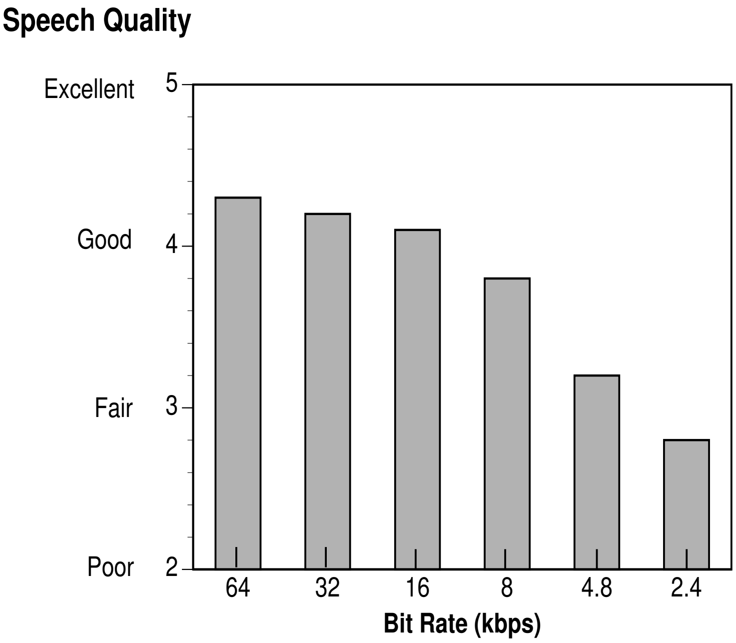

Figure 10.1: The speech quality mean opinion score for various bit rates.

although the telephone handset receiver tends to minimize the difference. ADPCM at 32 kbps is widely used for expanding the number of speech channels by a factor of two, particularly in private networks and international circuits. It is also the basis of low-complexity speech coding in several proposals for personal communication networks, including CT2 (Europe), UDPCS (USA) and Personal Handyphone (Japan)

For rates of 16 kbps and lower, high speech quality is achieved by using more complex adaptive prediction, such as linear predictive coding (LPC) and pitch prediction, and by exploiting auditory masking and the underlying perceptual limitations of the ear. Important examples of such coders are multi-pulse excitation, regular-pulse excitation, and code-excited linear prediction (CELP) coders. The CELP algorithm combines the high quality potential of waveform coding with the compression efficiency of model-based vocoders. At present, the CELP technique is the technology of choice for coding speech at bit rates of 16 kbps and lower. At 16 kbps, a low-delay CELP (LD-CELP) algorithm provides both high quality, close to PCM, and low communication delay and has been accepted as an international standard for transmission of speech over telephone networks.

At 8 kbps, which is the bit rate chosen for first-generation digital cellular telephony in North America, speech quality is good, although significantly lower than that of the 64 kbps PCM speech. Both North American and Japanese first generation digital standards are based on the CELP technique. The first European digital cellular standard is based on regular-pulse excitation algorithm at 13.2 kbps.

The rate of 4.8 kbps is an important data rate because it can be transmitted over most local telephone lines in the United States. A version of CELP operating at 4.8 kbps has been chosen as a United States standard for secure voice communication. The other such standard uses an LPC vocoder operating at 2.4 kbps. The LPC vocoder produces intelligible speech but the speech quality is not natural.

The present research is focussed on meeting the critical need for high quality speech transmission over digital cellular channels at 4 and 8 kbps. Low bit rate speech coders are fairly complex, but the advances in VLSI and the availability of digital signal processors have made possible the implementation of both encoder and decoder on a single chip.

10.2.4 Technology Targets

Given that there is no rigorous mathematical formula for speech entropy, a natural target in speech coding is the achievement of high quality at bit rates that are at least a factor of two lower than the numbers that currently provide high quality: 4 kbps for telephone speech, 8 kbps for wideband speech and 24 kbps for CD-quality speech. These numbers represent a bit rate of about 0.5 bit per sample in each case.

Another challenge is the realization of robust algorithms in the context of real-life imperfections such as input noise, transmission errors and packet losses.

Finally, an overarching set of challenges has to do with realizing the above objectives with usefully low levels of implementation complexity.

In all of these pursuits, we are limited by our knowledge in several individual disciplines, and in the way these disciplines interact. Advances are needed in our understanding of *coding, communication and networking, speech production and hearing*, and *digital signal processing.*

In discussing directions of research, it is impossible to be exhaustive, and in predicting what the successful directions may be, we do not necessarily expect to be accurate. Nevertheless, it may be useful to set down some broad research directions, with a range that covers the obvious as well as the speculative. The last part of this section is addressed to this task.

10.2.5 Future Directions

Coding, Communication, and Networking: In recent years, there has been significant progress in the fundamental building blocks of source coding: flexible methods of time-frequency analysis, adaptive vector quantization, and noiseless coding. Compelling applications of these techniques to speech coding are relatively less mature. Complementary advances in channel coding and networking include coded modulation for wireless channels and embedded transmission protocols for networking. Joint designs of source coding, channel coding, and networking will be especially critical in wireless communication of speech, especially in the context of multimedia applications.

Speech Production and Perception: Simple models of periodicity, and simple source models of the vocal tract need to be supplemented (or replaced) by models of articulation and excitation that provide a more direct and compact representation of the speech-generating process. Likewise, stylized models of distortion masking need to be replaced by models that maximize masking in the spectral and temporal domains. These models need to be based on better overall models of hearing, and also on experiments with real speech signals (rather than simplified stimuli such as tones and noise).

Digital Signal Processing: In current technology, a single general-purpose signal processor is capable of nearly 100 million arithmetic operations per second, and one square centimeter of silicon memory can store about 25 megabits of information. The memory and processing power available on a single chip are both expected to continue to increase significantly over the next several years. Processor efficiency as measured by mips-per-milliwatt of power consumption is also expected to improve by at least one order of magnitude. However, to accommodate coding algorithms of much higher complexity on these devices, we will need continued advances in the way we match processor architectures to complex algorithms, especially in configurations that permit graceful control of speech quality as a function of processor cost and power dissipation. The issues of power consumption and battery life are particularly critical for personal communication services and portable information terminals.

For further reading, we recommend Jayant and Noll (1984), Jayant, Johnston, et al. (1993), Lipoff (1994), Jayant (1992), Atal and Schroeder (1979), Atal (1982), Schroeder and Atal (1985), and Chen (1991).

10.3 Speech Enhancement

Dirk Van Compernolle

K.U. Leuven—ESAT, Heverlee, Belgium

Speech enhancement in the past decades has focused on the suppression of additive background noise. From a signal processing point of view additive noise is easier to deal with than convolutive noise or nonlinear disturbances. Moreover, due to the bursty nature of speech, it is possible to observe the noise by itself during speech pauses, which can be of great value.

Speech enhancement is a very special case of signal estimation as speech is nonstationary, and the human ear—the final judge—does not believe in a simple mathematical error criterion. Therefore subjective measurements of intelligibility and quality are required.

Thus the goal of speech enhancement is to find an *optimal estimate* (i.e., preferred by a human listener) $\hat{s}(t)$, given a noisy measurement $y(t) = s(t) + n(t)$. A number of overview papers can be found in Ephraim (1992) and Van Compernolle (1992).

10.3.1 Speech Enhancement by Spectral Magnitude Estimation

The relative unimportance of phase for speech quality has given rise to a family of speech enhancement algorithms based on spectral magnitude estimation. These are frequency-domain estimators in which an estimate of the clean-speech spectral magnitude is recombined with the noisy phase before resynthesis with a standard overlap-add procedure (Figure 10.2). The name *spectral subtraction* is loosely used for many of the algorithms falling in this class (Boll, 1979; Berouti, Schwartz, et al., 1979).

Power Spectral Subtraction: This is the simplest of all variants. It makes use of the fact that power spectra of additive independent signals are also additive and that this property is approximately true for short-time estimates as well. Hence, in the case of stationary noise, it suffices to subtract the mean noise power to obtain a least squares estimate of the power spectrum.

$$|\hat{S}(f)|^2 = |Y(f)|^2 - E[|N(f)|^2] \approx |Y(f)|^2 - |\overline{N(f)}|^2 \tag{10.1}$$

$$\hat{S}(f) = |\hat{S}(f)| \angle Y(f) \tag{10.2}$$

The greatest asset of spectral subtraction lies in its simplicity and the fact that all that is required, is an estimate of the mean noise power and that the algorithm doesn't need

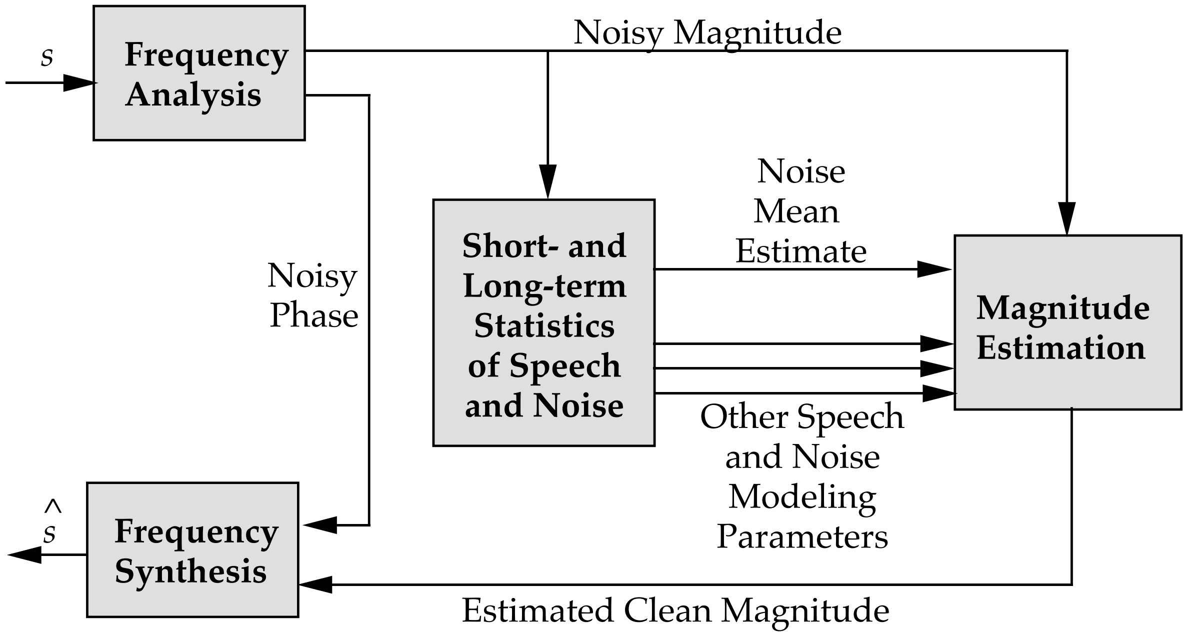

Figure 10.2: Speech Enhancement by Spectral Magnitude Estimation.

any signal assumptions. At the same time the latter is its great weakness. Within the framework occasional negative estimates of the power spectrum can occur. To make the estimates consistent some artificial flooring is required, which yields a very characteristic *musical noise*, caused by the remaining isolated patches of energy in the time-frequency representation.

Much effort has been put into reducing this annoying musical noise. One effective way is smoothing over time of the short-time spectra. This has the contrary effect, however, of introducing echoes. While reducing the average level of the background noise substantially, plain spectral subtraction has been rather ineffective in improving intelligibility and quality for broadband background noise.

Minimum Mean Square Error Estimators: Power spectral subtraction is a minimum mean square estimator with little or no assumptions about the prior distributions for power spectral values of speech and noise. This is the underlying reason why ad hoc operations like clipping are necessary. Within the framework of spectral magnitude estimation two major improvements are: (i) modeling of realistic a priori statistical distributions of speech and noise spectral magnitude coefficients (Ephraim & Malah, 1984), (ii) minimizing the estimation error in a domain which is perceptually more relevant than the power spectral domain (e.g., log magnitude domain) (Porter & Boll, 1984; Ephraim & Malah, 1985; Van Compernolle, 1989).

Minimum mean square error estimators (MMSEEs) have been developed under various assumptions such as Gaussian sample distributions, lognormal distribution of spectral magnitudes, etc. While improving on quality, these estimators tend to be complex and computationally demanding.

Time Varying Speech Models and State-based Methods: In a first generation the MMSEEs used a single distribution modeling *all speech* and one modeling *all noise*. Significant improvement is still possible if one takes into account the nonstationarity of the speech signal (and the noise). The use of *local* speech models implies much smaller variances in the models and tighter estimates. There are two possible approaches: (i) the incoming speech is aligned with an ergodic (fully-connected) HMM in which a separate MMSEE is associated with each state (Ephraim, Malah, et al., 1989), (ii) the parameters in a simple parametric speech model can be continuously adapted on the basis of the observations (Xie & Van Compernolle, 1993). In the first approach a set of possible *states* has to be created during a training phase and this should be a complete set. In the second approach no explicit training is required, but a simpler model may be needed to make the continuous parameter updates feasible.

It is obvious that neither the state association nor the parameter updates will be trivial operations and that this adds another level of complexity to the spectral estimation problem. A side effect of these methods is that they require dynamic time alignment which is inherently noncausal. While at most a *few frames* extra delay is inserted, this may be a concern in some applications.

10.3.2 Wiener Filtering

The Wiener filter obtains a least squares estimate of $s(t)$ under stationarity assumptions of speech and noise. The construction of the Wiener filter requires an estimate of the power spectrum of the clean speech and the noise:

$$W(f) = \frac{\Phi_{ss}(f)}{\Phi_{ss}(f) + \Phi_{nn}(f)}$$

The previous discussion on global and local speech and noise models equally applies to Wiener filtering. Wiener filtering has the disadvantage, however, that the estimation criterion is fixed.

10.3.3 Microphone Arrays

Microphone arrays exploit the fact that a speech source is quite stationary and therefore, by using beamforming techniques, can suppress nonstationary interferences more effectively than any single sensor system. The simplest of all approaches is the delay and sum beamformer that phase aligns incoming wavefronts of the desired source before adding them together (Flanagan, 1985). This type of processing is robust and needs only limited computational hardware, but requires a large number of microphones to be effective. An easy way to achieve uniform improvement over the wide speech bandwidth is to use a subband approach together with a logarithmically spaced array. Different sets of microphones are selected to cover the different frequency ranges (Silverman, 1987). A much more complex alternative is the use of adaptive beamformers, in which case each incoming signal is adaptively filtered before being added together. These arrays are most powerful if the noise source itself is directional. While intrinsically much more powerful, the adaptive beamformer is prone to signal distortion in strong reverberation. A third class of beamformers is a mix of the previous schemes. A number of digital filters are predesigned for optimal wideband performance for a set of look directions. The adaptation now exists in selecting the optimal filter at any given moment using a proper tracking mechanism. Under *typical* reverberant conditions, this last approach may prove the best overall solution. It combines the robustness of a simple method with the power of digital filtering.

While potentially very powerful, microphone arrays bring about a significant hardware cost due to the number of microphones and/or required adaptive filters. As a final remark it should be said that apart from noise suppression alone, microphone arrays help to dereverberate the signals as well.

10.3.4 Future Directions

The most substantial progress in the past decade has come from the incorporation of a model of the nonstationary speech signal into the spectral subtraction and Wiener filtering frameworks. The models under consideration have mostly been quite simple. It may be expected that the use of more complex models, borrowed from speech recognition work, will take us even further (cf. section 1.4). This line of work is promising from a quality point of view but implies much greater computational complexity as well. At the same time these models may have problems in dealing with events that did not occur during the *training phase.* Therefore the truly successful approaches will be those who strike the optimal balance between sufficiently detailed modeling of the speech signal to have a high quality estimator and a sufficiently weak model to allow for plenty of uncertainties.

Microphone arrays are promising but are expensive and have to develop further. The combination of single-sensor and multiple-sensor noise suppression techniques remains a virtually unexplored field.

10.4 Chapter References

Almeida, L. and Silva, F. (1984). Variable-frequency synthesis: an improved harmonic coding scheme. In *Proceedings of the 1984 International Conference on Acoustics, Speech, and Signal Processing*, page 27.5. Institute of Electrical and Electronic Engineers.

Atal, B. S. (1982). Predictive coding of speech at low bit rates. *IEEE Transactions on Communication*, COM-30(4):600–614.

Atal, B. S., Cuperman, V., and Gersho, A., editors (1991). *Advances in Speech Coding*. Kluwer Academic, Boston.

Atal, B. S. and Remde, J. R. (1982). A new model of LPC excitation for producing natural-sounding speech at low bit rates. In *Proceedings of the 1982 International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 614–617. Institute of Electrical and Electronic Engineers.

Atal, B. S. and Schroeder, M. R. (1979). Predictive coding of speech signals and subjective error criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, pages 247–254.

Berouti, M., Schwartz, R., and Makhoul, J. (1979). Enhancement of speech corrupted by additive noise. In *Proceedings of the 1979 International Conference on Acoustics, Speech, and Signal Processing*, pages 208–211. Institute of Electrical and Electronic Engineers.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-27(2):113–120.

Boll, S. (1991). Speech enhancement in the 1980s. In Furui, S. and Sondhi, M. M., editors, *Advances in Speech Signal Processing*. Marcel Dekker.

Chen, J.-H. (1991). A robust low delay CELP speech coder at 16 kb/s. In Atal, B. S., Cuperman, V., and Gersho, A., editors, *Advances in Speech Coding*, pages 25–35. Kluwer Academic, Boston.

Ephraim, Y. (1992). Statistical-model-based speech enhancement systems. *Proceedings of the IEEE*, 80(10):1526–1555.

Ephraim, Y. and Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-32(6):1109–1121.

Ephraim, Y. and Malah, D. (1985). Speech enhancement using a minimum mean-square log-spectral amplitude estimator. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-33(2):443–445.

Ephraim, Y., Malah, D., and Juang, B. H. (1989). Speech enhancement based upon hidden markov modeling. In *Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing*, pages 353–356, Glasgow, Scotland. Institute of Electrical and Electronic Engineers.

Flanagan, J. et al. (1979). Speech coding. *IEEE Transactions on Communications*, COM-27(4):710–737.

Flanagan, J. L. (1985). Use of acoustic filtering to control the beamwidth of steered microphone arrays. *Journal of the Acoustical Society of America*, 78(2):423–428.

Furui, S. and Sondhi, M. M., editors (1991). *Advances in Speech Signal Processing*. Marcel Dekker.

Griffin, D. and Lim, J. (1988). Multiband excitation vocoder. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(8).

ICASSP (1984). *Proceedings of the 1984 International Conference on Acoustics, Speech, and Signal Processing*. Institute of Electrical and Electronic Engineers.

ICASSP (1989). *Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, Scotland. Institute of Electrical and Electronic Engineers.

Jayant, N., Johnston, J., and Safranek, R. (1993). Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10).

Jayant, N. S. (1992). Signal compression: Technology targets and research directions. *IEEE Jour. Select. Areas Comm.*, 10(5):796–818.

Jayant, N. S. and Noll, P. (1984). *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice-Hall, Englewood Cliffs, New Jersey.

Kroon, P., Deprettere, E., and Sluyter, R. (1986). Regular-pulse excitation: a novel approach to effective and efficient multi-pulse coding of speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-34:1054–1063.

Lim, J. and Oppenheim, A. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67:1586–1604.

Lipoff, S. J. (1994). Personal communication networks: Bridging the gap between cellular and cordless phones. *Proceedings of the IEEE*, pages 564–571.

McAulay, R. and Quatieri, T. (1986). Speech analysis-synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-34:744–754.

Porter, J. E. and Boll, S. F. (1984). Optimal estimators for spectral restoration of noisy speech. In *Proceedings of the 1984 International Conference on Acoustics, Speech, and Signal Processing*, pages 18.A.2.1–4. Institute of Electrical and Electronic Engineers.

Schroeder, M. R. and Atal, B. S. (1985). Code-excited linear prediction CELP: High quality speech at very low bit rates. In *Proceedings of the 1985 International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 937–940, Tampa, Florida. Institute of Electrical and Electronic Engineers.

Silverman, H. (1987). Some analysis of microphone array for speech data acquisition. *IEEE Transactions on Acoustics Speech and Signal Processing*, 35(12):1699–1712.

Tremain, T. (1982). The government standard linear predictive coding algorithm: LPC-10. *Speech Technology Magazine*, pages 40–49.

Van Compernolle, D. (1989). Spectral estimation using a log-distance error criterion applied to speech recognition. In *Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing*, pages 258–261, Glasgow, Scotland. Institute of Electrical and Electronic Engineers.

Van Compernolle, D. (1992). DSP techniques for speech enhancement. In *Proceedings of the ESCA Workshop on Speech Processing in Adverse Conditions*, pages 21–30.

Xie, F. and Van Compernolle, D. (1993). Speech enhancement by nonlinear spectral estimation—a unifying approach. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 1, pages 617–620, Berlin. European Speech Communication Association.