Enhancement and Bandwidth Compression of Noisy Speech

JAE S. LIM, MEMBER, IEEE, AND ALAN V. OPPENHEIM, FELLOW, IEEE

Invited Paper

Abstract—Over the past several years there has been considerable attention focused on the problem of enhancement and bandwidth compression of speech degraded by additive background noise. This interest is motivated by several factors including a broad set of important applications, the apparent lack of robustness in current speech-compression systems and the development of several potentially promising and practical solutions. One objective of this paper is to provide an overview of the variety of techniques that have been proposed for enhancement and bandwidth compression of speech degraded by additive background noise. A second objective is to suggest a unifying framework in terms of which the relationships between these systems is more visible and which hopefully provides a structure which will suggest fruitful directions for further research.

I. INTRODUCTION

THERE ARE a wide variety of contexts in which it is desired to enhance speech. The objective of enhancement may perhaps be to improve the overall quality, to increase intelligibility, to reduce listener fatigue, etc. Depending on the specific application, the enhancement system may be directed at only one of these objectives or several. For example, a speech communication system may introduce a low-amplitude long-time delay echo or a narrow-band additive disturbance. While these degradations may not by themselves reduce intelligibility for the purposes for which the channel is used, they are generally objectionable and an improvement in quality perhaps even at the expense of some intelligibility may be desirable. Another example is the communication between a pilot and an air traffic control tower. In this environment, the speech is typically degraded by background noise. Of central importance is the intelligibility of the speech and it would generally be acceptable to sacrifice quality if the intelligibility could be improved. Even with normal undegraded speech, it is sometimes useful or desirable to provide enhancement. As a simple example high-pass filtering of normal speech is often used to introduce a "crispness" which is generally perceived as an improvement in quality.

The speech-enhancement problem covers a broad spectrum of constraints, applications and issues. Environments in which an additive background signal has been introduced are common. The background may be noise-like such as in aircraft, street noise, etc. or may be speech-like such as an environment with competing speakers. Other examples in which the need for speech enhancement arises include correcting for reverberation, correcting for the distortion of the speech of underwater divers breathing a helium–oxygen mixture, and correcting the distortion of speech due to pathological difficulties of the speaker or introduced due to an attempt to speak too rapidly. Even for these examples, the problem and techniques vary, depending on the availability of other signals or information. For example, for enhancement of speech in an aircraft a separate microphone can be used to monitor the background noise so that the characteristics of the noise can be used to adjust or adapt the enhancement system. At the air-traffic control tower, however, the only signal available for enhancement is the degraded speech.

Another very important application for speech enhancement is in conjunction with speech bandwidth compression systems. Because of the increasing role of digital communication channels coupled with the need for encrypting of speech and increased emphasis on integrated voice-data networks, speech-bandwidth-compression systems are destined to play an increasingly important role in speech-communication systems. The conceptual basis for narrow-band speech-compression systems stems from a model for the speech signal based on what is known about the physics and physiology of speech production. Because of this reliance on a model for the signal it is not unreasonable to expect that as the signal deviates from the model due to distortion such as additive noise, the performance of the speech compression system with regard to factors such as quality, intelligibility, etc., will degrade. It is generally agreed that the performance of current speech-compression systems degrades rapidly in the presence of additive noise and other distortions and there is currently considerable interest and attention being directed at the development of more robust speech compression systems. There are two basic approaches which are typically considered either of which may be preferable in a given situation. One approach is to base the bandwidth compression on the assumption of undistorted speech and develop a preprocessor to enhance the degraded speech in preparation for further processing by the bandwidth compression system. It is important to recognize that in enhancing speech in preparation for bandwidth compression the effectiveness of the preprocessor is judged on the basis of the output of the bandwidth-compression system in comparison with the output if no preprocessor is used. Thus, for example, it is possible that the output of the preprocessor would be judged by a listener to be inferior (by some measure) to the input but that the output of the bandwidth-compression system with the preprocessor is preferred to the output without it. In this case, the preprocessor would clearly be considered to be effective

Manuscript received June 22, 1979; revised August 28, 1979. This work was supported in part by the Defense Advance Research Projects Agency monitored by the Office of Naval Research under Contract N00014-75-C-0951-NR049-328 at M.I.T. Research Laboratory of Electronics and in part by the Department of the Air Force under Contract F19628-78-C-0002 at M.I.T. Lincoln Laboratory.

The authors are with M.I.T. Research Laboratory of Electronics and M.I.T. Lincoln Laboratory, Cambridge, MA 02139.

0018-9219/79/1200-1586$00.75

in enhancing the speech in preparation for bandwidth compression. Another approach to bandwidth compression of degraded speech is to incorporate into the model for the signal information about the degradation. A number of systems based on such an approach have recently been proposed and will be discussed in detail in this paper.

As is evident from the above discussion, the general problem of enhancing speech is broad and the constraints, information, and objectives are heavily dependent on the specific context and applications. In this paper, we consider only a small subset of possible topics, specifically the enhancement and bandwidth compression of speech degraded by additive noise. Furthermore, we assume that the only signal available is the degraded speech and that the noise does not depend on the original speech. Many practical problems, some of which have already been discussed, fall into this framework and some problems that do not can be transformed so that they do. For example, multiplicative noise or convolutional noise degradation can be converted to an additive noise degradation by a homomorphic transformation [1], [2]. As another example, signal-dependent quantization noise in pulse-code modulation (PCM) signal coding can be converted to a signal independent additive noise by a pseudo-noise technique [3]-[5].

Even within the limited framework outlined above, there is a diversity of approaches and systems. One objective of this paper is to provide an overview of the variety of techniques that have been proposed for enhancement of speech degraded by additive background noise both for direct listening and as a preprocessor for subsequent bandwidth compression. Many of these systems were developed independently of each other and on the surface often appear to be unrelated. Thus another objective of the paper is to provide a unifying framework in terms of which the relationship between these systems is more visible, and which hopefully will provide a structure which will suggest further fruitful directions for research.

In Section II, we present an overview of the general topic. In this overview we classify the various enhancement systems based on the information assumed about the speech and the noise. Some systems based on time-invariant Wiener filtering, for example, rely only on an assumed noise power spectrum and on long-time average characteristics of speech, such as the fact that the average speech spectrum decays with frequency at approximately 6 dB/octave. Other systems rely on aspects of speech perception or speech production in general or on a detailed model of speech.

Sections III–V present a more detailed discussion of several of these categories of speech-enhancement systems. In particular, Section III is concerned with the general principle of speech enhancement based on estimation of the short-time spectral amplitude of the speech. This basic principle encompasses a variety of techniques and systems including the specific methods of spectral subtraction, parametric Wiener filtering, etc. In Section IV, speech enhancement techniques which rely principally on the concept of the short-time periodicity of voiced speech are reviewed, including comb-filtering and related systems. Section V discusses a variety of systems that rely on more specific modeling of the speech waveform. As we will discuss in detail, in some cases, parameters of the model are obtained from an analysis of the degraded speech and used to synthesize the enhanced speech. In other cases, the results of an analysis based on a model for speech are used to control an enhancement filter, perhaps with the procedure being iterative so that the output of an enhancement filter is then subjected to further analysis, etc. Many of these systems also incorporate a number of the techniques introduced in Section III, including Wiener filtering and spectral subtraction.

In Sections III–V, the focus is entirely on systems for enhancement with the evaluation of the systems being based on listening without further processing. In Section VI, we consider the related but separate problem of bandwidth compression of speech degraded by additive noise.

In Section VII, we discuss in some detail the evaluation of the performance of the various systems presented in the earlier sections. In general, the performance evaluation of a speech-enhancement system is extremely difficult, in large measure because the appropriate criteria for evaluation are heavily dependent on the specific application of the system. Relative importance of such factors as quality, intelligibility, listener fatigue, etc., may vary considerably with the application. In Section VII, we summarize the performance evaluations that have been reported for the various systems presented in this paper. Since the evaluation of different systems has generally been based on different procedures, environments, etc., no attempt is made in the section to *compare* individual systems. In general, however, we will see that while many of the enhancement systems reduce the apparent background noise and thus perhaps increase quality, many of them to varying degrees, reduce intelligibility. In the context of bandwidth compression, however, various systems provide an increase in intelligibility over that obtained without the incorporation of speech enhancement.

II. Overview of Systems for Enhancement and Bandwidth Compression of Noisy Speech

As indicated in the previous section, our focus in this paper is on degradation due to the presence of additive noise. Even within this limited context there are a wide variety of approaches which have been proposed and explored. Conceptually any approach should attempt to capitalize on available information about the signal, i.e., the speech, and the background noise. Speech is a special subclass of audio signals and there are reasonable models in terms of which the speech waveform can be described and categorized. The more specifically we attempt to model the speech signal, the more potential for separating it from the background noise. On the other hand, the more we assume about the speech the more sensitive the enhancement system will be to inaccuracies or deviations from these assumptions. Thus incorporating assumptions and information about the speech signal represents tradeoffs which are reflected in the various systems. In a similar manner systems can attempt to incorporate detailed information about the background noise. For example, the type of processing suggested if the background noise is a competing speaker is different than if it is wide-band random noise. Thus enhancement systems also tend to differ in terms of the assumptions made regarding the background noise. As with assumptions related to the signal, the more an enhancement system attempts to capitalize on assumed characteristics of the noise the more susceptible it is likely to be to deviations from these assumptions.

Another important consideration in speech enhancement stems from the fact that the criteria for enhancement ultimately relate to an evaluation by a human listener. In different contexts the criteria for evaluation may differ depending on whether quality, intelligibility, or some other attribute is the

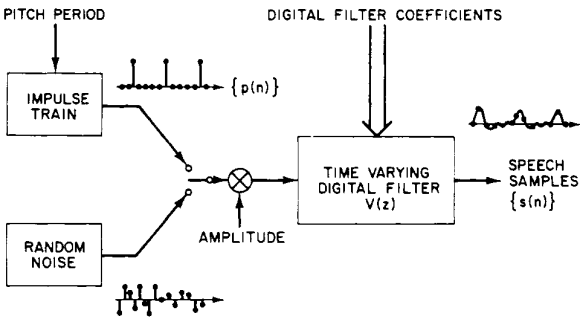

Fig. 1. A speech production model.

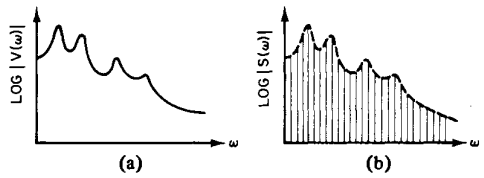

Fig. 2. An example of resonant frequencies of an acoustic cavity. (a) Vocal-tract transfer function. (b) Magnitude spectrum of a speech sound with the resonant frequencies shown in (a).

most important. Thus speech enhancement must inevitably take into account aspects of human perception. As we will indicate shortly, some systems are heavily motivated by perceptual considerations, others rely more on mathematical criteria. In such cases, of course, the mathematical criteria must in some way be consistent with human perception, and, while an optimum mathematical criterion is not known, some mathematical error criteria are understood to be a better match than others to aspects of human perception.

In the following discussion we briefly describe some aspects of speech production and speech perception that in varying degrees play a role in speech-enhancement systems. Following that we present a brief overview of a representative collection of speech-enchancement systems, with the intent of categorizing these systems in terms of the various aspects of speech production and perception on which they attempt to capitalize.

Speech is generated by exciting an acoustic cavity, the vocal tract, by pulses of air released through the vocal cords for voiced sounds, or by turbulence for unvoiced sounds. Thus a simple but useful model for speech production consists of a linear system, representing the vocal tract, driven by an excitation function which is a periodic pulse train for voiced sounds and wide-band noise for unvoiced sounds, as illustrated in Fig. 1. Furthermore, since the linear system represents an acoustic cavity, its response is of a resonant nature, so that its transfer function is characterized by a set of resonant frequencies, referred to as formants, as illustrated in Fig. 2(a). Thus, if the excitation and vocal-tract parameters are fixed, then as indicated in Fig. 2(b), the speech spectrum has an envelope representing the vocal-tract transfer function of Fig. 2(a) and a fine structure representing the excitation.

Many of the techniques for speech enhancement, particularly those in Sections III and V are conceptually based on the representation of the speech signal as a stochastic process. This characterization of speech is clearly more appropriate in the case of unvoiced sounds for which the vocal tract is driven by wide-band noise. The vocal tract of course changes shape as different sounds are generated and this is reflected in a time varying transfer function for the linear system in Fig. 1. However, because of the mechanical and physiological constraints on the motion of the vocal tract and articulators such as the tongue and lips, it is reasonable to represent the linear system in Fig. 1 as a slowly varying linear system so that on a short-time basis it is approximated as stationary. Thus some specific attributes of the speech signal, which can be capitalized on in an enhancement system are that it is the response of a slowly varying linear system, that on a short-time basis its spectral envelope is characterized by a set of resonances, and that for voiced sounds, on a short-time basis it has a harmonic structure. This simplified model for speech production has generally been very successful in a variety of engineering contexts including speech enchancement, synthesis, and bandwidth compression. A more detailed discussion of models for speech production can be found in [6]-[8].

The perceptual aspects of speech are considerably more complicated and less well understood. However, there are a number of commonly accepted aspects of speech perception which play an important role in speech-enchancement systems. For example, consonants are known to be important in the intelligibility of speech even though they represent a relatively small fraction of the signal energy. Furthermore, it is generally understood that the short-time spectrum is of central importance in the perception of speech and that, specifically, the formants in the short-time spectrum are more important than other details of the spectral envelope. It appears also, that the first formant, typically in the range of 250 to 800 Hz, is less important perceptually, than the second formant [9], [10]. Thus it is possible to apply a certain degree of high pass filtering [11], [12] to speech which may perhaps affect the first formant without introducing serious degradation in intelligibility. Similarly low-pass filtering with a cutoff frequency above 4 kHz, while perhaps affecting crispness and quality will in general not seriously affect intelligibility. A good representation of the magnitude of the short-time spectrum is also generally considered to be important whereas the phase is relatively unimportant. Another perceptual aspect of the auditory system that plays a role in speech enhancement is the ability to mask one signal with another. Thus, for example, narrow-band noise and many forms of artificial noise or degradation such as might be produced by a vocoder are more unpleasant to listen to than broad-band noise and a speech-enhancement system might include the introduction of broad-band noise to mask the narrow-band or artificial noise.

All speech-enhancement systems rely to varying degrees on the aspects of speech production and perception outlined above. One of the simplest approaches to enhancement is the use of low-pass or bandpass filtering to attenuate the noise outside the band of perceptual importance for speech. More generally, when the power spectrum of the noise is known, one can consider the use of Wiener filtering, based on the long-time power spectrum of speech. While in some cases such as the presence of narrow-band background noise, this is reasonably successful, Wiener filtering based on the long-time power spectrum of the speech and noise is limited because speech is not stationary. Even if speech were truly stationary, mean-square error which is the error criterion on which Wiener filtering is based is not strongly correlated with perception and thus is not a particularly effective error criterion to apply to speech processing systems. This is evidenced, for example, in the use of masking for enhancement. By adding broad-band

noise to mask other degradation, we are, in effect, increasing the mean-square error. Another example that suggests that mean-square error is not well matched to the perceptually important attributes in speech is the fact that distortion of the speech waveform by processing with an all-pass filter results in essentially no audible difference if the impulse response of the all-pass filter is reasonably short but can result in a substantial mean-square error between the original and filtered speech. In other words, mean-square error is sensitive to phase of the spectrum whereas perception tends not to be.

Masking and bandpass filtering represent two simple ways in which perceptual aspects of the auditory system can be exploited in speech enhancement. Another system whose motivation depends heavily on aspects of speech perception was proposed by Thomas and Niederjohn [12] as a preprocessor prior to the introduction of noise in those applications where noise-free speech is available for processing. In essence, their system applies high-pass filtering to reduce or remove the first formant followed by infinite clipping. The motivation for the system lies in the observation that at a given signal-to-noise ratio infinite clipping will increase, relative to the vowels, the amplitude of the perceptually important low-amplitude events such as consonants thus making them less susceptible to masking by noise. In addition, for vowels the filtering will increase the amplitude of higher formants relative to the first formant, thus making the perceptually more important higher formants less susceptible to degradation. In the speech enhancement problem considered in this paper, noise-free speech is not available for processing as required in the above system. Thomas and Ravindran [13], however, applied high-pass filtering followed by infinite clipping to noisy speech as an experiment. While quality may be degraded by the process of filtering and clipping, they claim a noticeable improvement in intelligibility when applied to enhance speech degraded by wide-band random noise. One possible explanation may be that the high-pass filtering operation reduces the masking of perceptually important higher formants by the relatively unimportant low-frequency components.

Another system which relies heavily on human perception of speech was proposed by Drucker [14]. Based on some perceptual tests, Drucker concluded that one primary cause for the intelligibility loss in speech degraded by wide-band random noise is the confusion among the fricative and plosive sounds which is partly due to the loss of short pauses immediately before the plosive sounds. By high-pass filtering one of the fricative sounds, the /s/ sound, and inserting short pauses before the plosive sounds (assuming that their locations can be accurately determined), Drucker claims a significant improvement in intelligibility.

In discussing perceptual attributes we indicated that the short-time spectral magnitude is generally considered to be important whereas the phase is relatively unimportant. This forms the basis for a class of speech enhancement systems which attempt in various ways to estimate the short-time spectral magnitude of the speech without particular regard to the phase and to use this to recover or reconstruct the speech. This class of systems includes spectral subtraction techniques originally due to Weiss *et al.* [15], [16], and which have recently received a great deal of attention [17]–[22] and optimum filtering techniques such as Wiener filtering and power spectrum filtering. These systems will be discussed in considerable detail in Section III. As we will see, many of these systems which appear on the surface to be different are in fact identical or very closely related.

In addition to directly or indirectly utilizing perceptual attributes most enhancement systems rely to varying degrees on aspects of speech production. For example, in Section IV, we describe in detail a variety of systems that attempt, in some way, to capitalize on short-time periodicity of speech during voiced sounds. As a consequence of this periodicity, during voiced intervals the speech spectrum has a harmonic structure which suggests the possibility of applying comb filtering or as proposed by Parsons [23] attempting to extract in other ways, the components of the speech spectrum only at the harmonic frequencies. In essence, knowledge of the harmonic structure of voiced sounds allows us in principle to remove the noise in the spectral bands between the harmonics.

As discussed in Section IV, speech enhancement by comb filtering can also be viewed in terms of averaging successive periods of the noisy speech to partially cancel the noise. Another system, which attempts to take advantage of the quasi-periodic nature of the speech was proposed by Sambur [24]. As developed in more detail in Section IV, his system is based on the principles of adaptive noise cancelling. Unlike the classical procedure Sambur's method is designed to cancel out the clean speech signal, taking advantage of the quasi-periodic nature of the speech to form an estimate of the speech at each time instant from the value of the signal one period earlier.

In the model of speech production, we represented the speech signal as generated by exciting a quasi-stationary linear system with a pulse train for voiced speech and noise for unvoiced speech. Based on this model, an approach to speech enhancement is to attempt to estimate parameters of the model rather than the speech itself and to then use this to synthesize the speech, i.e., to enhance speech through the use of an analysis-synthesis system. A particularly novel application of this concept was used by Miller [25] to remove the orchestral accompaniment from early recordings of Enrico Caruso. In this system homomorphic deconvolution was used to estimate the impulse response of the model in Fig. 1. A similar approach to noise reduction was proposed by Suzuki [26], [27] whereby the short-time correlation function of the degraded speech is used as an estimate of the impulse response of the linear system. This system is referred to as splicing of auto correlation function (SPAC). A modification of SPAC is referred to as splicing of cross-correlation function (SPOC). A number of systems also attempt to model the vocal-tract impulse response in more detail. As we discussed previously the vocal-tract transfer function is characterized by a set of resonances or formants that are perceptually important. This suggests the possibility of representing the vocal-tract impulse response in terms of a pole-zero model with the analysis procedure directed at estimating the associated parameters. The poles in particular would provide a reasonable representation of the formants.

All-pole modeling of speech has had notable success in analysis–synthesis systems for clean speech. A number of recent efforts have been directed toward estimating the parameters in an all-pole model from noisy observations of the speech such as the systems by Magill and Un [28], Lim and Oppenheim [29], Lim [18], and Done and Rushforth [30]. Extensions to pole-zero modeling have also been proposed

by Musicus and Lim [31] and Musicus [32]. These various approaches are described and compared in detail in Section V.

The above discussion was intended as a brief overview of the general approaches to speech enhancement. In the next three sections we explore in more detail many of the systems mentioned above. In particular, in Section III, we focus on speech-enhancement techniques based on short-time spectral amplitude estimation. In Section IV our focus is on speech enhancement based on periodicity of voiced speech and in Section V on speech-enhancement techniques using an analysis-synthesis procedure.

III. Speech Enhancement Techniques Based on Short-Time Spectral Amplitude Estimation

In general, in enhancement of a signal degraded by additive noise, it is significantly easier to estimate the spectral amplitude associated with the original signal than it is to estimate both amplitude and phase. As we discussed in Section II, it is principally the short-time spectral amplitude rather than phase that is important for speech intelligibility and quality. As we discuss in this section, there are a variety of speech-enhancement techniques that capitalize on this aspect of speech perception by focusing on enhancing only the short-time spectral amplitude. The techniques to be discussed can be broadly classified into two groups. In the first, presented in Section III-A, the short-time spectral amplitude is estimated in the frequency domain, using the spectrum of the degraded speech. Each short-time segment of the enhanced speech waveform in the time domain is then obtained by inverse transforming this spectral amplitude estimate combined with the phase of the degraded speech. In the second class, discussed in Section III-B the degraded speech is first used to obtain a filter which is then applied to the degraded speech. Since these procedures lead to zero-phase filters, it is again only the spectral amplitude that is enhanced, with the phase of the filtered speech being identical to that of the degraded speech.

In both classes of systems discussed below no conceptual distinction is made between voiced and unvoiced speech and in particular in contrast to the techniques to be discussed in Section IV the periodicity of voiced speech is not exploited. Both classes of systems in this section are most easily interpreted in terms of a stochastic characterization of the speech signal. While this characterization is more justifiable for unvoiced speech it has been shown empirically to also lead to successful procedures for voiced speech.

A. Speech Enhancement Based on Direct Estimation of Short-Time Spectral Amplitude

When a stationary random signal $s(n)$ has been degraded by uncorrelated additive noise $d(n)$ with a known power density spectrum, the power density spectrum or spectral amplitude of the signal is easily estimated through a process of spectral subtraction. Specifically, if

$$y(n) = s(n) + d(n) \tag{1}$$

and $P_y(\omega)$, $P_s(\omega)$, and $P_d(\omega)$ represent the power density spectra of $y(n)$, $s(n)$, and $d(n)$, respectively, then

$$P_y(\omega) = P_s(\omega) + P_d(\omega). \tag{2}$$

Consequently, a reasonable estimate for $P_s(\omega)$ is obtained by subtracting the known spectrum $P_d(\omega)$ from an estimate of $P_y(\omega)$ developed from the observations of $y(n)$.

Speech, of course, is not a stationary signal. However, with $s(n)$ in (1) now representing a speech signal and with the processing to be carried out on a short-time basis we consider $s(n)$, $d(n)$, and $y(n)$ multiplied by a time-limited window $w(n)$. With $y_w(n)$, $d_w(n)$, and $s_w(n)$ denoting the windowed signals $y(n)$, $d(n)$, and $s(n)$ and $Y_w(\omega)$, $D_w(\omega)$, and $S_w(\omega)$ as their respective Fourier transforms we have

$$y_w(n) = s_w(n) + d_w(n) \tag{3}$$

and

$$|Y_w(\omega)|^2 = |S_w(\omega)|^2 + |D_w(\omega)|^2 + S_w(\omega) \cdot D_w^*(\omega) + S_w^*(\omega) \cdot D_w(\omega) \tag{4}$$

where $D_w^*(\omega)$ and $S_w^*(\omega)$ represent complex conjugates of $D_w(\omega)$ and $S_w(\omega)$. The function $|S_w(\omega)|^2$ will be referred to as the short-time energy spectrum of speech. For speech enhancement based on the short-time spectral amplitude, the objective is to obtain an estimate $|\hat{S}_w(\omega)|$ of $|S_w(\omega)|$ and from this, an estimate $\hat{s}_w(n)$ of $s_w(n)$.

From the estimate $\hat{s}_w(n)$, speech can be generated in a variety of different ways. One approach is to use an analysis window function $w(n)$ that generates $s(n)$ when all the frames of $s_w(n)$ are overlapped and added with the appropriate time registration. Such a window function satisfies the equation

$$\sum_i w_i(n) = 1, \quad \text{for all } n \text{ of interest} \tag{5}$$

where $w_i(n)$ represents the ith window frame. Two such examples are overlapped triangular and hamming windows. Using such a window function, speech is then generated by adding up the estimates of the windowed segments.

Various speech-enhancement techniques discussed in this section differ primarily in how $|S_w(\omega)|$ is specifically estimated from the noisy speech. In one spectral subtraction technique referred to as *power spectrum subtraction*,[1] $|S_w(\omega)|$ is estimated based on (4). From the observed data $y_w(n)$, $|Y_w(\omega)|^2$ can be obtained directly. The terms $|D_w(\omega)|^2$, $S_w(\omega) \cdot D_w^*(\omega)$ and $S_w^*(\omega) \cdot D_w(\omega)$ cannot be obtained exactly and in the power spectrum subtraction technique they are approximated by $E[|D_w(\omega)|^2]$, $E[S_w(\omega) \cdot D_w^*(\omega)]$ and $E[S_w^*(\omega) \cdot D_w(\omega)]$ where $E[\cdot]$ denotes the ensemble average. For $d(n)$ zero mean[2] and uncorrelated with $s(n)$, $E[S_w(\omega) \cdot D_w^*(\omega)]$ and $E[S_w^*(\omega) \cdot D_w(\omega)]$ are zero and an estimate $|\hat{S}_w(\omega)|^2$ of $|S_w(\omega)|^2$, is suggested from (4) as

$$|\hat{S}_w(\omega)|^2 = |Y_w(\omega)|^2 - E[|D_w(\omega)|^2], \tag{6}$$

where $E[|D_w(\omega)|^2]$ is obtained either from the assumed known properties of $d(n)$ or by an actual measurement from the background noise in the intervals where speech is not present. The estimate $|\hat{S}_w(\omega)|^2$ based on (6) is not guaranteed to be nonnegative since the right-hand side can become negative, and a number of somewhat arbitrary choices have been made. In some studies, the negative values are made positive by changing the sign. In some other studies $|\hat{S}_w(\omega)|^2$ is set to zero if

[1] The name "power spectrum subtraction" comes from the close similarity between (2) and (6).

[2] The zero mean assumption for the additive random noise is made only for notational convenience.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.