Abstract

The development in digital technologies has facilitated speech signal to be reduplicated and edited at high fidelity. Although many applications benefit from these developments, new social issues related to malicious attacks and unauthorized tampering to speech have arisen. For example, by using advanced speech analysis/synthesis tools, ordinary people are capable to produce high naturalness of tampered speech without leaving perceptual clues. Since these tools enable the speech to be tampered in a much easier and credible way, it is becoming difficult to confirm the originality of speech signal. As an important information carrier, the originality of speech signals should be strictly confirmed. To avoid the unauthorized tampering as well as the negative influence that they may cause, it is necessary to conduct relevant research about speech protection and tampering detection to protect speech signal.

Information hiding technique which can hide or embed digital data such as copyright notice or serial number in the original speech signal has been considered as an effective solution for the above issues. The embedded digital data is generally referred as watermarks, and this kind of information hiding methods is specified as watermarking methods. To be effective, watermarking methods should satisfy several requirements: (1) inaudibility to human auditory system, (2) blindness for watermark extraction, (3) robustness against allowable speech processing and common attacks, and (4) fragility against tampering. The first three requirements are required for general watermarking methods, and the last one is an additional requirement when watermarking methods are used for tampering detection. However, it is proven to be difficult for watermarking methods to satisfy all these requirements simultaneously. Our research aim is to solve the problem of unauthorized tampering with information hiding and watermarking methods. The first target is to realize a general watermarking method that can satisfy all the first three requirements. After that, this watermarking method will be applied to other applications, such as tampering detection by exploring the fragility, and hybrid watermarking.

Since human auditory system is usually not sensitive to tiny changes of speech parameters, watermarks are possible to be inaudibly embedded by subtly modifying speech parameters. According to the source-filter model, the linear prediction (LP) coefficients can provide accurate estimation of formants. The line spectral frequencies (LSFs), as substitute parameters of LP coefficients, can not only represent the formants but also have several excellent properties: (i) they are less sensitive to noise and (ii) the influences caused by the deviation of LSFs can be limited to the local spectral, thus the distortion introduced by LSFs deviation in both spectral and sound quality can be minimized. In addition, since LSFs are universal features in different speech codecs, if watermarks are embedded into LSFs, they are possible to survive from the encoding/decoding process. Therefore, embedding watermarks into LSFs also enables the watermarking method to be robust against difficult speech codecs.

Since LSFs can directly represent the formants, the modifications to LSFs made by watermark embedding can be physically considered as make tuning to the formants of speech signal. Therefore, our main concept for watermarking is formant tuning. Based on this concept, we propose two watermarking schemes. One is watermarking based on quantizing LSFs with quantization index modulation (QIM) (LSFs-QIM based watermarking). In this method, different watermarks are embedded into the LSFs of speech signal with different quantization steps. In the watermarking extraction process, watermarks are blindly extracted by re-quantizing the LSFs obtained from the watermarked signal with the same quantization steps. However, it is found that, since the QIM based modifications to LSFs are quite unintentional, the original formant structure of speech signal is easily disrupted, which will degrade the sound quality of speech signal. Moreover, the performance of this method is characterized by the quantization step, i.e., small quantization step is benefit for good sound quality of watermarked signal but strong robustness cannot be obtained, and vice versa. Therefore, it is difficult for this method to get a trade-off between inaudibility and robustness.

As to overcome these drawbacks, the original formant structure of speech signal should be considered for better performance. As we have found, in the field of speech synthesis, formant which is a crucial acoustic feature for speech perception, can be enhanced to improve the quality and intelligibility of speech when the speech is impaired by environmental noise or other reasons. Since formant can be enhanced to improve the speech quality, and such modifications do not cause perceptual distortion to the original speech, watermarking based on formant enhancement is possible to be inaudible to human auditory system. Based on this concept, we propose another watermarking scheme, i.e., watermarking based on formant enhancement (formant-enhancement based watermarking). In this method, different watermarks are embedded by enhancing different formants: ``0" is embedded by enhancing the sharpest formant and ``1" is embedded by enhancing the second sharpest formants, after which different bandwidth relationships between the sharpest and the second sharpest formants are established. These different bandwidth relationships can be used to blindly extract watermarks in the extraction process.

We evaluate the proposed two watermarking methods with respect to inaudibility and robustness (both methods are blind). For the LSFs-QIM based watermarking, the performance of inaudibility and robustness are evaluated with different quantization steps. The results from inaudibility evaluation reveal that the proposed method can satisfy inaudibility when quantization steps are small. The results from robustness evaluation suggest that the proposed method has good bit detection rate for normal extraction and some of general speech processing. However, the weak robustness of this method against speech codecs, down-sampling, and low-bit quantization has greatly restricted its effectiveness. For the formant enhancement based watermarking, evaluations are carried out for both this method and other watermarking methods to make a comparison study. The LP order and the modification level for the formant enhancement based method are well examined for achieving good performance in inaudibility and robustness. Based on the evaluation results, watermark embedding through formant enhancement does not cause severe degradation to the original speech quality, and the watermark extraction by identifying bandwidth relationship is able to tolerate slight distortions of frequency components caused by other processing. Therefore, the formant enhancement based method can satisfy the requirements of inaudibility, blindness, and robustness, especially the robustness against speech codecs.

Since the formant enhancement based method can satisfy the three basic requirements for watermarking, we apply it to tampering detection scheme of speech signal. Ideally, if the watermarking method can satisfy fragility, tampering can be detected with the mismatched bits between embedded watermarks and the extracted watermarks. The tampering detection ability of the proposed scheme is evaluated against several kinds of tampering. The embedding bit rate of watermarks is 4 bps, each embedded bit is able to account for 0.25 s speech segment when locating the tampering. The evaluation results show that when tampering has been made to the watermarked speech, watermarks in the tampered segment will be destroyed. Therefore, the proposed scheme is fragile against tampering, and it has the ability to detect tampering as well as checking the originality of speech signals. The formant enhancement based watermarking is also applied to hybrid watermarking method, where the formant enhancement based watermarking and cochlear delay based watermarking are combined together. The evaluation results suggest that the robustness of hybrid method can be improved compared with each single method, since the disadvantage of one watermarking method can be concealed by the other watermarking method.

Based on these results, we conclude that the formant enhancement based method can satisfy the first three requirements for general watermarking. It can also satisfy fragility when used for tampering detection. Therefore, it has the ability to solve the problems of speech tampering.

Keywords: Information hiding, speech watermarking, formant enhancement, tampering detection, hybrid watermarking