Effects of the Voice over Internet Protocol on Perturbation Analysis of Normal and Pathological Phonation

Yanmei Zhu Rachel E. Witt Julia K. MacCallum Jack J. Jiang

Department of Surgery, Division of Otolaryngology, Head and Neck Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisc., USA

Key Words

Jitter • Shimmer • Audio compression • Voice over Internet Protocol

Abstract

Objective: In this study, a Voice over Internet Protocol (VoIP) communication based on G.729 protocol was simulated to determine the effects of this system on acoustic perturbation parameters of normal and pathological voice signals.

Patients and Methods: Fifty recordings of normal voice and 48 recordings of pathological voice affected by laryngeal paralysis were transmitted through a VoIP communication system. The acoustic analysis programs of CSpeech and MDVP were used to determine the percent jitter and percent shimmer from the voice samples before and after VoIP transmission. The effects of three frequently used audio compression protocols (MP3, WMA, and FLAC) on the perturbation measures were also studied. **Results:** It was found that VoIP transmission disrupts the waveform and increases the percent jitter and percent shimmer of voice samples. However, after VoIP transmission, significant discrimination between normal and pathological voices affected by laryngeal paralysis was still possible. It was found that the lossless compression method FLAC does not exert any influence on the perturbation measures. The lossy compression methods MP3 and

WMA increase percent jitter and percent shimmer values.

Conclusion: This study validates the feasibility of these transmission and compression protocols in developing remote voice signal data collection and assessment systems.

Copyright © 2010 S. Karger AG, Basel

Introduction

Acoustic perturbation measures, such as jitter and shimmer, describe the frequency and amplitude perturbation characteristics of voices. These parameters have been applied to noninvasively and objectively assess laryngeal function and voice quality since the 1960s [1, 2]. Numerous studies have shown that the acoustic perturbation measures can reflect the internal functioning of the voice production system [1–15], and it has been reported that acoustic measures have accuracies of over 90% for detecting vocal pathologies [16]. Therefore, these measures provide an important method for noninvasive discrimination between normal voice and pathological voice. Moreover, because the voice signal can be noninvasively recorded and easily transferred across extended geographic distances, acoustic perturbation analysis could potentially be applied for the assessment of voice disorders in remote regions.

The high quality of the voice signal is an important precondition to ensure the accuracy of voice perturbation analysis [17], and consequently, the assessment of that voice. The quality of the voice signal and the accuracy of subsequent perturbation analysis can be affected by various factors when the signal is recorded, stored and transferred. It has been found that jitter and shimmer are sensitive to variations in microphone type and placement [18, 19], sampling rate or signal length [20], noise [21], extraction algorithms [22], and analysis systems [23, 24]. The errors introduced by these factors affect the perturbation values, which directly affects the accuracy of assessment regarding laryngeal pathologies. For example, an increase of 1% in percent jitter is likely to lead to the conclusion by a speech-language pathologist that the perturbation value does not fall within the normal range [6].

Today, Internet development and the proliferation of Internet Protocol (IP) telephones make the transmission of voice signal data increasingly more convenient and accessible, which creates the potential to meet the needs of medically underserved populations in remote regions. However, for long-distance voice communication systems, audio compression is often used to achieve a low bit rate in the digital representation of the original audio signal with minimum perceivable loss of quality. The compressed audio files are obviously smaller than the WAV file, so these formats are not only popularly used for network real-time transfer, but also for the audio data store and share. Most of the digital recorders also store the audio file in compression formats, e.g. Free Lossless Audio Codec (FLAC), MPEG-1 Audio Layer 3 (MP3, ISO 11172-3) and Windows Media Audio (WMA, Microsoft). Audio compression, referring to an audio encoding technique, is usually employed for efficient transmission and storage of voice signal data [25]. These techniques typically achieve far greater file compression by discarding data considered to be less critical, such as high-frequency sounds, or sounds that occur simultaneously with other louder sounds [26]. After a voice signal is compressed, the information in the original data that was discarded during compression cannot be completely recovered. Therefore, this kind of compression is referred to as lossy compression. We anticipate that in a remote acoustic perturbation analysis system, the lossy compression protocol could be a new factor influencing voice quality and voice perturbation measurements. Therefore, before a network-based acoustic analysis system or compressed audio data can be used for practical clinical applications, an investigation of the effects of the Voice over Internet Pro-

tol (VoIP) and compression on voice perturbation analysis is necessary.

In this study, we examined the potential influence of compression protocols over VoIP systems on voice perturbation parameters. A specific protocol for human voice, G.729 [27], is an International Telecommunication Union (ITU) standard speech data compression algorithm that compresses voice audio at 8 kb/s. This protocol encodes speech in a high-quality form using a relatively low bit rate. It is one of the most frequently used protocols for speech transmission over IP; therefore, the voice signal transmission through VoIP used in this study was based on G.729. In our computer simulation, a sender and a receiver of an IP communication system were simulated by an encoder and a decoder based on G.729. Percent jitter and percent shimmer were measured in the voice signals before and after transmission over IP by using CSpeech [28] and MDVP [29]. SigmaPlot 3.0 and SigmaStat 2.0 (Jandel Scientific, San Jose, Calif., USA) were used to statistically compare perturbation measurement results. In addition, MP3, WMA, and FLAC are three of the most popular forms of music compression and storage for Internet transmission. The potential effects of these codecs on the perturbation analyses of voice signals were also examined in this study.

Patients and Methods

Subject Voice Data

The voice data analyzed in this study were obtained from the Disordered Voice Database, model 4337, version 1.03 (Kay Elemetrics Corp., Lincoln Park, N.J., USA), developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. The sample included 50 normal voice recordings from people with healthy laryngeal systems and 48 pathological voice recordings from patients with laryngeal paralysis.

The voice data contained in the Disorder Voice Database are the sustained vowel /a/. Voice recordings were made in a sound-proof booth on a DAT recorder in mono channel at a sampling rate of $f_s = 50,000$ Hz (Kay Elemetrics Corp.). Recordings were made in PCM format with a bit rate of 16. All of the voice recordings used in this study were resampled at $f_s = 44,100$ Hz using GoldWave software (version 5.23) (Goldwave Inc., St. John's, Nfld., Canada) to mimic the most widely used sampling rate present on the Internet. The original voice samples and the transmitted voice samples are subsequently referred to as the resampled recordings in this study.

Simulation of VoIP Telecommunication Based on G.729

The G.729 protocol is an 8-kb/s Conjugate-Structure Algebraic-Code-Excited Linear Prediction speech compression algorithm approved by ITU Telecommunication Standardization Sector (ITU-T) [27]. It is the most regularly used compression

Fig. 1. Diagram of VoIP communication system based on the G.729 protocol.

method in VoIP applications because of its low bandwidth requirement.

An encoder and a decoder based on the G.729 protocol were used to simulate the sender and receiver of a VoIP system. Voice samples with 16 bit sample size, 44,100 Hz sampling rate, and PCM format were entered into the G.729 encoder and converted to an 8-kb/s compressed data stream. Then, the compressed binary data stream was sent to the receiver and decoded to a 16-bit, 44,100-Hz sampled Wave (WAV) file. Figure 1 presents a diagram of the simulated VoIP communication system. In this study, the encoder and decoder were programmed in the Microsoft C++ Language from G.729 (ITU-T G.729).

Voice Compression with MP3, WMA, and FLAC

The MP3 protocol is an International Organization for Standardization and International Electrotechnical Commission (ISO/IEC) standard released in 1991 as a means of lossy audio compression. This format is currently the most common way to store a single audio segment because the file can be easily organized and transferred between computers or other digital devices, such as MP3 players. This audio coding, commonly referred to as Perceptual Coding, compresses the audio data by reducing the accuracy of certain parts of sound deemed to be beyond the auditory resolution ability of most people [30]. In this study, voice data with 16 bit sample size and 44,100 kHz sampling ratio were compressed by the MP3 encoder (LAME, version 3.97) with a bit ratio of 128 kb/s and saved as MP3 files in the computer. Usually, these MP3 files are about 1/10th the size of the uncompressed WAV file from the original audio source.

WMA is another type of popular audio encoder. This audio encoder is also based on the study of psychoacoustics, which may discard imperceptible audio signals during the compression process. Therefore, WMA is a lossy audio encoder which could result in the irreversible loss of audio quality and may introduce undesired compression artifacts not present in the source material. In this study, the WMA encoder (GoldWave V.5.23) was used to compress voice data with 16 bit sample size and 44,100 kHz sampling ratio to WMA files with a bit ratio of 128 kb/s.

FLAC has been recognized as the fastest and most widely supported lossless audio codec. FLAC is a lossless audio codec, which means that audio is compressed by FLAC without removing any information from the audio stream. FLAC can typically reduce the size of the audio stream to between 40 and 50% of the original signal. In this study, a 16-bit mono channel FLAC encoder (GoldWave V.5.23) was used to encode the original voice signals into FLAC files and to decode FLAC signals back to PCM signed 16-bit mono channel WAV signals.

Perturbation Measurements

Two perturbation measures, percent jitter and percent shimmer, were analyzed in this study.

(1) Percent jitter is a cycle-to-cycle frequency perturbation measure, defined as

$$\text{Jitter}(\%) = \frac{\left[\frac{1}{(K-1)} \right] \sum_{i=1}^{K-1} \left| T_0^{(i)} - T_0^{(i+1)} \right|}{(1/K) \sum_{i=1}^K T_0^{(i)}} \quad (1)$$

(2) Percent shimmer is a cycle-to-cycle amplitude perturbation measure, defined as

$$\text{Shimmer}(\%) = \frac{\left[\frac{1}{(K-1)} \right] \sum_{i=1}^{K-1} \left| A^{(i)} - A^{(i+1)} \right|}{(1/K) \sum_{i=1}^K A^{(i)}} \quad (2)$$

In equations 1 and 2, K is the number of pitch periods, $T_0^{(i)}$ and $A^{(i)}$ are the extracted pitch period and the peak-to-peak amplitude of a voice signal, respectively.

MDVP and CSpeech are two popular programs for speech analysis in clinics. Both of them have the ability to estimate jitter and shimmer from normal/pathological voice signals. However, the two programs use different algorithms to estimate these parameters. The difference in algorithm may cause slight differences between their respective calculated results. Therefore, in this study, MDVP and CSpeech software were both used to estimate the perturbation measurements of the voice signal. The percent jitter and the percent shimmer calculated by the two types of acoustic analysis software are reported in the 'Results' section.

Procedure and Statistical Analysis

Fifty normal voice recordings and 48 pathological voice recordings were transferred through the simulated VoIP telecommunication system. The measurements of percent jitter and percent shimmer were extracted from both the original voice signal before transmission and from the voice signals recovered at the receiver side using both MDVP and CSpeech software. The Mann-Whitney rank sum test was employed, using percent jitter and shimmer as dependent variables and the subject groups (normal voice before transmission, pathological voice before transmission, normal voice after transmission, pathological voice after transmission) as independent variables. SigmaStat 2.0 (Jandel Scientific) software was used for statistical analysis.

Similar procedures were also used to study the effects of MP3, WMA, and FLAC compression on the perturbation measurements. The percent jitter and percent shimmer of 50 original normal voice recordings and 48 original pathologic voice recordings were estimated two separate times by using MDVP and CSpeech.

Then, the original recordings were compressed using the MP3, WMA, and FLAC encoders, respectively. However, because MDVP and CSpeech cannot directly analyze a voice signal in MP3, WMA, or FLAC formats, we converted these files to a WAV format before analysis.

Results and Discussion

Figure 2 compares an original waveform (fig. 2a) with the same waveforms after VoIP transmission and audio compression (fig. 2b–e). The waveform of the voice after FLAC compression (fig. 2e) is identical to the original one. This is because FLAC is a lossless compression method, and the compressed FLAC file contains all of the information that is contained in the original audio signal. The difference between the original and WMA compressed signals (fig. 2d) is very small; we calculated difference series by subtracting WMA compressed signal from the original signal. To quantitatively measure the disruption of the WMA compression on the voice signal, the standard deviation of this difference series was calculated. It is only 3.54×10^{-3} , which is much smaller than the amplitude of the original signal. There also exists a difference between the original and MP3 compressed signals (fig. 2c). Figures 2c, d coincide in that MP3 and WMA are lossy compression methods. In addition, the waveform of the voice signal after transmission through a VoIP communication system is presented in figure 2b. The VoIP transmission based on the G.729 protocol significantly disturbs the waveform of the voice signal. Moreover, in comparison with the original signal, the waveform shows a time delay after VoIP transmission. This is because the G.729 protocol is based on the Code-Excited Linear Prediction model, in which each 80-bit frame produced contains linear prediction coefficients, excitation code book indices, and gain parameters that are used by the decoder in order to reproduce speech. The protocol requires 10-ms input frames and generates frames of 80 bits in length. With the G.729 processing signals in 10-ms frames and 5-ms look-ahead, the total algorithmic delay is 15 ms.

The above result shows that the VoIP transmission or lossy compression could disturb the waveform of the voice signal. This suggests that perturbation measures obtained from these disturbed voice signals may be different from the original signals. To study the effects of the voice transmission and compression on perturbation measures, CSpeech and MDVP software were used to analyze the voice signals.

Fig. 2. The waveform of a normal voice signal. **a** Waveform of the original normal voice signal. **b** Waveform after G.729 VoIP transmission. **c** Waveform after MP3 compression encoding. **d** Waveform after WMA compression encoding. **e** Waveform after FLAC encoding.

Figures 3 and 4 present the perturbation measures estimated by CSpeech and MDVP, respectively, where the first column and the second column correspond to the results of the normal voice and the pathological voice, respectively. Tables 1 and 2 illustrate the mean values of percent jitter and percent shimmer. The transmission by VoIP and compression MP3 and WMA change the perturbation values of voice signals compared to the original signals.

Fig. 3. The distributions of **a** percent jitter for normal voice signals, **b** percent shimmer for normal voice signals, **c** percent jitter for pathological voice signals, and **d** percent shimmer for pathological voice signals as estimated by CSpeech. Distributions are shown for original signals, signals after G.729 VoIP transmission,

signals after MP3 compression encoding, signals after WMA compression encoding, and signals after FLAC encoding. The line inside the box marks the median, whiskers show 10th and 90th percentiles, and the dots are the outlying points.

Original Voice

For normal voice data, both percent jitter (0.236% for CSpeech; 0.590% for MDVP) and percent shimmer (1.571% for CSpeech; 2.199% for MDVP) values of the original signals are low. However, for the pathological voice data, the percent jitter (1.197% for CSpeech; 2.703% for MDVP) and the percent shimmer (7.445% for CSpeech; 8.898% for MDVP) values are significantly higher than the corresponding values of the normal voice ($p \leq 0.001$).

In addition, we may also notice that the perturbation parameters estimated by MDVP and CSpeech are different. This is because different algorithms are used in anal-

ysis by different programs. Therefore, the influence of VoIP on the perturbation parameters estimated by the two programs was investigated in this study.

VoIP Transmission Based on G.729 Protocol

After transmission through a VoIP system, the percent jitter values of the normal voice and the pathological voice are both increased (tables 1, 2). Percent jitter values of the normal voice after transmission are significantly greater than those of the normal voice before transmission ($p \leq 0.001$ for both CSpeech and MDVP). Moreover, this same pattern of results is observed in the percent jit-

Fig. 4. The distributions of **a** percent jitter for normal voice signals, **b** percent shimmer for normal voice signals, **c** percent jitter for pathological voice signals, and **d** percent shimmer for pathological voice signals as estimated by MDVP. Distributions are shown for original signals, signals after G.729 VoIP transmission,

signals after MP3 compression encoding, signals after WMA compression encoding, and signals after FLAC encoding. The line inside the box marks the median, whiskers show 10th and 90th percentiles, and the dots are the outlying points.

ter values of the pathological voice after the signal is transmitted by VoIP ($p \leq 0.001$ for both CSpeech and MDVP).

Similarly, the percent shimmer values of the normal and pathological voices are increased after the VoIP transmission based on G.729 protocol. Percent shimmer values of both normal and pathological voices are significantly greater after transmission than those values before transmission ($p \leq 0.001$). Moreover, the effect of the VoIP on the percent shimmer values is relatively greater than that on the percent jitter values. After transmission, the percent jitter is increased 2.32 times accord-

ing to CSpeech and 1.88 times according to MDVP, respectively. However, the percent shimmer is increased 3.31 times according to CSpeech and 2.59 times according to MDVP, respectively.

In addition, it should be noted that after VoIP transmission, the perturbation measures of the pathological voice are still significantly higher than those of the normal voice ($p \leq 0.001$). Though these results suggest that VoIP transmission could influence the values of jitter and shimmer, voice signals after VoIP transmission may be used for distinguishing the normal voice from the voice of patients with laryngeal paralysis through the compar-

Table 1. Percent jitter and percent shimmer values estimated by CSpeech (mean $\pm$ SD)

	Percent jitter, %		Percent shimmer, %	
	normal	pathological	normal	pathological
Original	0.236 $\pm$ 0.0914	1.197 $\pm$ 1.215	1.571 $\pm$ 0.699	7.445 $\pm$ 6.754
VoIP	0.418 $\pm$ 0.127	2.926 $\pm$ 4.059	7.690 $\pm$ 2.030	22.129 $\pm$ 14.471
MP3	8.263 $\pm$ 9.697	6.072 $\pm$ 8.830	30.609 $\pm$ 25.045	26.237 $\pm$ 8.830
WMA	0.377 $\pm$ 0.163	1.377 $\pm$ 0.910	2.241 $\pm$ 0.696	8.617 $\pm$ 5.810
FLAC	0.236 $\pm$ 0.091	1.197 $\pm$ 1.215	1.571 $\pm$ 0.699	7.445 $\pm$ 6.754

Table 2. Percent jitter and percent shimmer values estimated by MDVP (mean $\pm$ SD)

	Percent jitter, %		Percent shimmer, %	
	normal	pathological	normal	pathological
Original	0.590 $\pm$ 0.440	2.703 $\pm$ 1.995	2.199 $\pm$ 0.946	8.898 $\pm$ 4.632
VoIP	1.171 $\pm$ 1.065	5.028 $\pm$ 4.063	10.815 $\pm$ 2.745	17.684 $\pm$ 6.424
MP3	0.644 $\pm$ 0.483	2.913 $\pm$ 2.167	3.342 $\pm$ 1.016	9.673 $\pm$ 4.467
WMA	0.598 $\pm$ 0.468	2.771 $\pm$ 2.061	2.452 $\pm$ 0.934	9.040 $\pm$ 4.575
FLAC	0.588 $\pm$ 0.441	2.703 $\pm$ 1.995	2.208 $\pm$ 0.939	8.898 $\pm$ 4.632

ison of relative, rather than absolute, values for the acoustic parameters of jitter and shimmer.

MP3 Compression

In applying MDVP to the voice signals compressed by the MP3 encoder, it is found that the changes in the percent jitter and shimmer values are very small. In fact, the Mann-Whitney rank sum test confirms that for percent jitter and the percent shimmer, there is not a statistically significant difference between the original pathological voice signals and the MP3-compressed voice signals ($p = 0.626$ and $p = 0.321$, respectively). Moreover, after MP3 compression, the perturbation measures of the pathological voice are significantly higher than those of the normal voice ($p \leq 0.001$).

However, after applying CSpeech on the same voice signals after MP3 compression, the obtained results are different from those estimated by MDVP. The jitter and shimmer values after compression are significantly greater than those before compression ($p \leq 0.001$). This significant change could be related to the internal algorithm of MP3 and CSpeech. MP3 compression is based on the study of psychoacoustics. This audio coding compresses the audio data by reducing the accuracy of certain parts of sound deemed to be beyond the auditory resolution

ability of most people. Therefore, imperceptible audio signals could be discarded during the compression process, which could result in the irreversible loss of audio quality and may introduce significant increases in jitter and shimmer. Furthermore, after MP3 compression, no significant difference can be found between the normal and the pathological groups for jitter and shimmer values ($p = 0.842$ and $p = 0.779$, respectively).

These analysis results indicate that the MP3-compressed voice signal may be used for distinguishing the normal voice from the voice of patients with laryngeal paralysis through the comparison of acoustic parameters which are estimated by MDVP; however, the perturbation measures estimated from the MP3 voice signal by CSpeech may not be appropriate for making this distinction.

WMA Compression

The increases of jitter and shimmer values induced by WMA compression are lower than those induced by MP3 compression. In contrast to MP3 compression, the percent jitter and shimmer values estimated by CSpeech are similar to those estimated by MDVP (tables 1, 2). Percent jitter and percent shimmer values after WMA compression are not significantly different from those before

compression. However, after WMA compression, the perturbation measures of the pathological voice are still significantly higher than those of the normal voice ($p \leq 0.001$).

These analysis results suggest that the WMA-compressed voice signal may be used for distinguishing the normal voice from the voice of patients with laryngeal paralysis through the comparison of percent jitter and shimmer, which are estimated by MDVP or CSpeech.

FLAC Compression

Perturbation analysis was also applied to the voice signal after FLAC compression. FLAC uses a four-stage method to compress audio: (1) The blocking stage divides the audio signal into blocks of a specified size. (2) The interchannel decorrelation stage is performed to remove redundancy in the stereo signals' left and right channels. (3) The prediction stage calculates the way in which each block could be most closely modeled by using linear prediction and run-length encoding; the prediction requires less memory to store than does the original signal. (4) Finally, the difference between the predicted and the original signals is compressed using Rice coding. Because the residual signal is stored, FLAC preserves the exact original signal and reduces the file size. Because FLAC is a lossless compression algorithm, perturbation measures are not influenced by this compression algorithm. The values for percent jitter and percent shimmer of the voice after compression are identical to those values prior to compression.

Conclusion

In summary, the lossless audio compression method FLAC does not exert any influence on the perturbation measures. However, after the application of the lossy compression protocol MP3, perturbation measures of the output voice signals estimated by CSpeech are significantly different from those measures before compression. Additionally, discrimination between the normal and pathological voice recordings from patients with laryngeal paralysis is impossible. The VoIP transmission and the lossy compression protocol WMA also disrupt the waveform of the voice samples and increase the percent jitter and shimmer of voice samples. However, significant discrimination between the normal and pathological voice recordings from patients with laryngeal paralysis is still possible. Therefore, the VoIP communication system based on the G.729 protocol has the potential to be used

as an appropriate telecommunications system for clinical application in the remote assessment of laryngeal disorders due to paralysis.

In the current study, we chose two groups of typical voice data to examine the influence of VoIP transmission and different audio compression protocols on the acoustic parameters of jitter and shimmer in voice signals. One group included normal voice recordings; another group included pathological voice recordings from patients with laryngeal paralysis. However, as a primary study, we chose to exclude other possible voice types and laryngeal diseases, such as hoarse voice and breathy voice. This is a limitation of the current study. We found that VoIP transmission and lossy audio compression will change the values of jitter and shimmer compared to those measured in the original voice signal. Therefore, the critical values of these acoustic parameters necessary for the discrimination between normal and pathological voices should be modified for the voice signal after transmission or compression. Further studies are necessary to examine the effects of digital transmission and compression on perturbation analysis for more general pathological situations.

Acknowledgment

This study was supported by NIH grant No. 1-RO1DC006019 and No. 1-RO1DC05522 from the National Institute of Deafness and Other Communication Disorders.

References

- 1 Lieberman P: Perturbations in vocal pitch. *J Acoust Soc Am* 1961;33:597–603.
- 2 Lieberman P: Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *J Acoust Soc Am* 1963; 35:344–353.
- 3 Hecker MHL, Kreul EJ: Description of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency. *J Acoust Soc Am* 1970;49:1275–1282.
- 4 Deal RE, Emanuel FW: Some waveform and spectral features of vowel roughness. *J Speech Hear Res* 1978;21:250–264.
- 5 Horii Y: Fundamental frequency perturbation observed in sustained phonation. *J Speech Hear Res* 1979;22:5–19.
- 6 Horii Y: Jitter and shimmer differences among sustained vowel phonation. *J Speech Hear Res* 1982;25:12–14.
- 7 Sorensen D, Horii Y: Frequency and amplitude perturbation in the voice of female speakers. *J Commun Disord* 1983;16:57–61.

- 8 Sorensen D, Horii Y: Directional perturbation factors for jitter and for shimmer. *J Commun Disord* 1984;17:143–151.
- 9 Klingholz F, Martin F: Quantitative spectral evaluation of shimmer and jitter. *J Speech Hear Res* 1985;28:169–174.
- 10 Hall KD: Variations across time in acoustic and electroglottographic measures of phonatory function in women with and without vocal nodules. *J Speech Hear Res* 1995;38:783–793.
- 11 Koike Y: Vowel amplitude modulations in patients with laryngeal diseases. *J Acoust Soc Am* 1969;45:839–844.
- 12 Iwata S: Periodicities of pitch perturbations in normal and pathologic larynges. *Laryngoscope* 1972;82:87–96.
- 13 Pinto NB, Titze IR: Unification of perturbation measures in speech signals. *J Acoust Soc Am* 1990;87:1278–1289.
- 14 Wendahl RW: Laryngeal analog synthesis of harsh voice quality. *Folia Phoniatri* 1963;15:241–250.
- 15 Wendahl RW: Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. *Folia Phoniatri* 1966;15:241–250.
- 16 Rosen KM, Kent RD, Delaney AL: Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. *J Speech Hear Res* 2006;49:395–411.
- 17 Titze I, Horii Y, Scherer R: Some technical considerations in voice perturbation measurements. *J Speech Hear Res* 1987;30:252–260.
- 18 Gelfer MP, Fendel DM: Comparisons of jitter, shimmer, and signal-to-noise ratio from directly digitized versus taped voice samples. *J Voice* 1995;9:378–382.
- 19 Jiang J, Lin E, Hanson DG: Effect on tape recording on perturbation measures. *J Speech Hear Res* 1998;41:1031–1041.
- 20 Karnell MP: Laryngeal perturbation analysis: minimum length of analysis window. *J Speech Hear Res* 1991;34:544–548.
- 21 Carson CP, Ingrisano DS, Eggleston KD: The effect of noise on computer-aided measures of voice: a comparison of CSpeechSP and the multi-dimensional voice program software using the CSL 4300B module and multi-speech for windows. *J Voice* 2003;17:12–20.
- 22 Titze I, Liang H: Comparison of F0 extraction method for high-precision voice perturbation measurements. *J Speech Hear Res* 1993;36:1120–1133.
- 23 Bielałowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS: Comparison of voice analysis systems for perturbation measurements. *J Speech Hear Res* 1996;39:126–134.
- 24 Karnell MP, Hall KD, Landahl KL: Comparison of fundamental frequency and perturbation measurements among three analysis systems. *J Voice* 1995;9:383–393.
- 25 Jayant NS, Johnston JD, Safranek RJ: Signal compression based on models of human perception. *Proc IEEE* 1993;10:1385–1422.
- 26 Kohlrausch A: The perceptual basis for audio compression. *Physics Today* 2007;6:80.
- 27 ITU-T Rec: G.729 Annex A. Reduced Complexity 8 kbit/s CS-ACELP Speech Codec. 1996.
- 28 Milenkovic P, Read C: CSpeech Version 4 User's Manual. 1992.
- 29 Kay Elemetrics Corp: Multi-Dimensional Voice Program: Software Instruction Manual. 1993.
- 30 Andres T, Fancisco M: Perceptual cryptography on MPEG layer III bit-streams. *IEEE Trans Comp Elec* 2002;48:1046.