A SPECTRAL CUES PRESERVING COMPRESSION ALGORITHM FOR DIGITAL HEARING AID

Xianbo Xiao, Guangshu Hu, and Chunhong Liu

Department of Biomedical Engineering, Tsinghua University and
Department of Otolaryngology, XuanWu Hospital, Beijing, China
email:{xianbo.xiao@gmail.com}

ABSTRACT

Spectral cues are important for speech intelligibility, while they are distorted, in most popular multi-band compression algorithms for hearing aid, by splitted into multiply bands and processed by different amplifiers. This paper presented a morphology-base method, which could preserve primary spectral cues and take low computation cost. Speech testing results demonstrated its efficiency.

1. INTRODUCTION

Compression algorithm is the fundamental part of hearing aid, which maps all the sounds in normal people's hearing dynamic range to impaired people's, and keeps sounds comfortable and intelligible [1]. Spectral cues, including spectral envelop, formants locations and magnitude contrast, are significant features to guarantee high speech intelligibility. But traditional bandwidth fixed multi-band compression methods didn't take those into consideration, and distorted spectral structures by discontinuous amplifiers between neighbor bands [2].

Recently, Researchers set to look for new methods [3, 4, 5], which can apply nonlinear amplification to different frequency sounds, and keep spectral cues farthest. The smearing effect seems inevitable in bandwidth fixed filters, so here comes two kinds of solutions. One is to design filters instantly for every frame. Filters can be adaptive, but should be adjusted in time to keep up with the variance of phonemes. The other solution is to go back to single channel path [3]. In early days, single channel methods appeared to meet the constraint of hardware, and eliminated by multi-band methods subsequently. But now they came back to reparation multi-band methods' limitation. For example, [4] used sinusoidal modeling to find primary peaks and preserved them during compression, which also implied peaks indicated by LPC could be kind of targets.

This paper followed the second way. Here presented a morphology-based single channel compression method, which met both the needs of compression and cues-keeping, using only logical judgments and addition operations. Speech testing results compared to multi-band compression methods showed progress. Finally, a technology of enlarging the peak-to-valley magnitude contrast to improve frequency resolution ability was tested here. But no noticeable progress was observed.

2. LPC-BASE METHOD

Linear Predictive Coding, as well known, using an AR model to estimate spectral of speech, can locate formants and illustrate energy distribution reasonably. So spectral cues picked up from LPC spectral have more meanings in speech and hearing. A LPC-based spectral cues preserving compression method can be illustrated by following:

256 points, sampled by 16 kHz and overlapped by 128 points, are filtered by a 16 msec hamming window, and then estimated by a LPC spectral. Cues points could be found on FFT spectral corresponding to the peaks and valleys of LPC spectral. Find gains for cues points in a lookup table, and all 256 gains will be achieved by interpolation. A RIFFT operation finally gives the waveform in time domain [6, 7].

Fig.1 shows a frame in vowel /   /. Spectral envelope and cue peaks can be found well preserved after compression. These undistorted spectral details are important for better speech intelligibility. But LPC method consumes too much to be suitable for realizing in hearing aid system. So a new method should be found to replace the LPC procedure, which should have similar locating function, and lower computation consumption.

3. MORPHOLOGY-BASED METHOD

Morphology provided a solution here. In given opinion, morphology comes from geometry and studies image mechanism both in macrostructure and in microstructure [8]. Its special ability in macrostructure can find cue peaks and valleys from FFT spectral for us. Following sections introduced morphology operators used in this paper and describe the morphology-based method in detail.

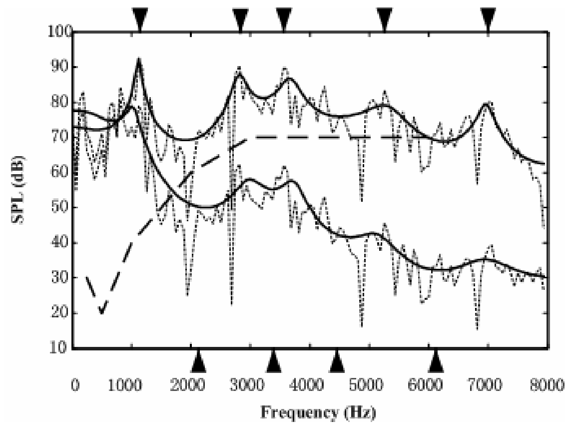

Figure 1: FFT spectral (dotted line) and LPC spectral (solid line), before amplified (top group) and after amplified (bottom group), the dashed line is an individual audiogram which algorithms were prescribed to. Arrows on top point to cue peaks and arrows on bottom point to cue valleys

5.1. Morphology operations

1. Erosion and dilation

Erosion and dilation are two basic operators in the area of mathematical morphology. In binary image, the erosion operator takes a set of coordinate points known as a structuring element, to erode away the boundaries of regions of foreground pixels. Dilation work reversely.

2. Opening and closing

Opening and closing are two very important operators from mathematical morphology. An opening is defined as an erosion followed by a dilation using the same structuring element for both operations, and closing is dilation after erosion. They are normally applied to gray-value images. Closing is the dual of opening.

Here, opening procedure served to smooth the contours of the object, break narrow isthmuses, and eliminate thin protrusions and small objects. And the closing was performed to fill in gaps and eliminate small holes.

3. Top-Hat routine and Bottom-Hat routine

Top-hat routine is also based on a sliding structuring element. An opening, removes all objects with a size below the structuring element leaving the background unaffected. The top-hat is then obtained by subtracting this opening from the original image to represent the objects thinner than the top-hat diameter. Bottom-hat routine does the same based on closing.

Top-hat and Bottom-Hat here are applied on FFT spectral instead of grey-scale image, to find out peaks or valleys narrower than the structuring element, and the routine results will be criterions for selecting reasonable one from several overlapped cue points.

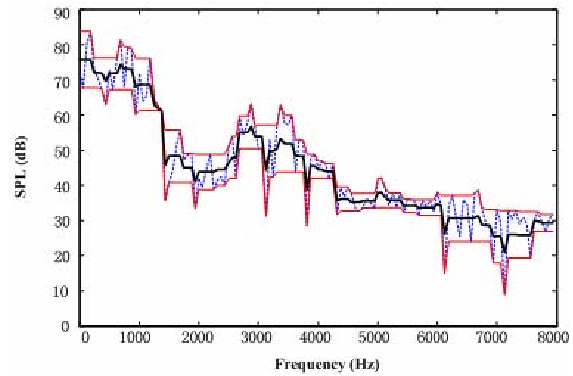

Figure2: Envelopes of FFT spectral (thin solid) and averaging curve (bold solid)

5.2. Morphology-based method

The morphology-based method is called MORM below, and LPC-based method LPCM.

MORM takes 5 steps to accomplish compression:

1. An opening operation and a closing operation are applied to the FFT spectral to get the envelopes, using a line (s1 points in length). Averaging the top envelope and bottom envelope to get a curve which will reveal spectral energy distribution (as in Fig.2).
2. Take closing operation and Top-Hat routine on previous averaged curve separately to eliminate peaks no wider than s2 points (smaller peaks will be considered as more details than trends). Remaining peaks will be candidates for cues. Similarly, valley candidates are picked out by opening operation and Bottom-Hat routine.
3. Choose cue points on FFT spectral. They can be central points of the cue peaks and valleys, or can be the highest or deepest points at the central parts of the candidates.
4. Previous incomplete detection might bring missorts. Remediation procedure will get the missed back, and pruning procedure will cut some out (as in Fig3):
   a) If the first cue point is for a peak, then it shouldn't be positioned less than 200Hz. Because most phoneme's first formants are not lower than 200Hz, and low frequency noise always made fake peaks there.
   b) If the last cue point is for a valley, then it shouldn't be positioned more than 7200Hz;
   c) Space between two neighbor cue points shouldn't be less than 300Hz, otherwise, keep the deeper or higher one, and cut the other out.
   d) If two neighbor cue points are both for peaks, add one for valley in the middle place of them.
   e) If tow neighbor cue points are both for valleys, add one for peak in the middle place of them.

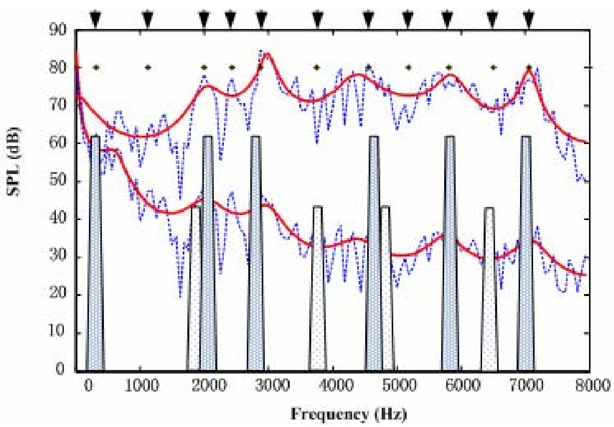

Figure3: Effect of Remediation and pruning procedure. The dark stems and light stems stand for peaks and valleys candidates. Arrows on the top point to the cue points including for peaks and for valleys after remediation and pruning, and they seem reasonable.

It's important that above rules should be carried out in turn; otherwise the result will be unpredictable.

5. Compute the gains for cue points by prescription formula, interpolate them to amplify all 128 points, and then return waveform by RIFFT.

5.2. Simulation results

Fig4 shows an example of a girl's voice 'he who whatever'. MORM is found similar to PLCM by a great extent both on waveform and spectrogram. All Mandarin vowels and consonants went through simulation experiments, and statistic data were integrated. Mismatches existed mainly in three aspects: (a) Inaccuracy of corresponding cue points. The error is ± 2 points generally, which won't make perceptible error in spectral compression results. (b) Missing cues. When peaks or valleys are too narrow or too flat, they might be missed. Most of them can be got back in remediation procedure. (c) Mistaking some peaks for cues. This case seldom happens in fact.

Cue points location accuracy decreased slightly in MORM than in LPCM, while computation cost dropped dramatically. All the structuring elements in opening operation, closing operation, Top-Hat and Bottom-Hat here are chosen to be series of points equal to 1, so MORM method can pick out cue points, by only logical judgments, addition and subtraction operations.

In this case, the computation consumption and storage space needed are reduced in a great extend, and complexity of computation and requirement to system are minimized. This will be an impressive superiority when implemented on real-time system, for example, on low-cost fixed-point DSP.

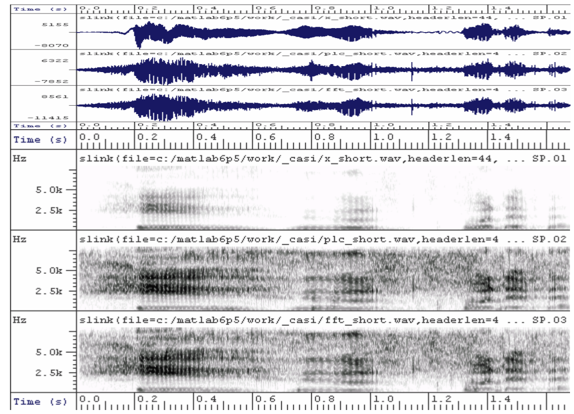

Fig4: Comparison of PLCM and MORM process results of voice 'he who whatever' in waveform (top group) and spectrogram (bottom group). original signal is on top, LPCM result in middle, and MORM result on bottom (s1=8, s2=22)

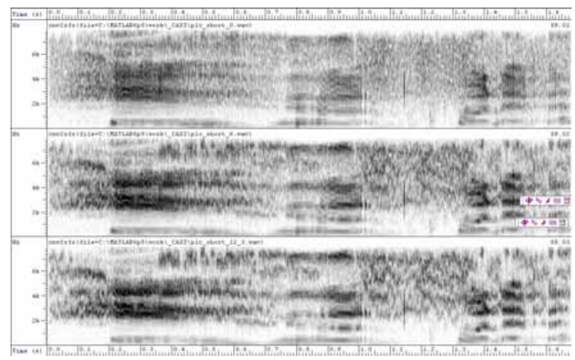

Fig5: Effect of CONT+MORM procedure. The top one is for only MORM procedure, middle one for a mild CONT (valleys were decreased by 6dB, peaks were increased by 3dB), and bottom one for an intense CONT (valleys were decreased by 12dB, peaks were increased by 3dB)

4. CONTRAST TECHNOLOGY

In addition to loss of sensitivity, the impaired ear exhibits reduced frequency resolution, which also contributes to deterioration of speech intelligibility. Because this loss is due to damaged outer hair-cells and the flattened tuning curves, no signal processing method seems to have capability to fix it [9, 10]. However, some reports showed, enlarging fall of valleys and peaks could be helpful. As MORM could separate primary peaks and valleys, this technology (referred as CONT technology) was embedded in MORM and evaluated here.

The degree valleys are decreased will affect the result. And to compensation sounds energy, peaks should be driven up. Fig.5 showed effect of CONT technology. In CONT result, formants stripes are more conspicuous. But the fluctuation shouldn't be too much, or some frequencies would become inaudible.

5. EXPERIMENT AND RESULT

To evaluate MORM and CONT, 5 moderate impaired persons with flat hearing loss participated in speech intelligibility testing. Testing materials are 30 monosyllabic Mandarin words, processed by MORM, MORM+CONT, and an 8-channel compression algorithm (MBCM) separately. Listeners were asked to choose vowels or consonants from 3 selective branches immediately after he/she had heard a word. Testing signal amplitude was set to MCL mode (chosen by listeners themselves at preview) and soft mode (10 dB below MCL). Testing with noise was conducted too.

Testing results are listed in Table 1. MORM shows better intelligibility than MBCM, for both vowels and consonants. When there is background noise and SNR is still high, MORM is better than MBCM. But when SNR falls, MORM results show no advantage than MBCM. Thus it can be seen, that MORM is more sensitive to noise, and may deteriorate when SNR falls further.

Intelligibilities scores of MORM and CONT+MORM were compared, and no distinctive progress was founded. But this doesn't mean CONT technology is noneffective. More detailed testing should be employed in this.

		Soft	Soft-noise	MCL	MCL-noise
MORM	consonant	*58 (10)*	*24 (15)*	*62 (5)*	*40 (19)*
	vowel	*64 (4)*	*34 (11)*	*70 (2)*	*64 (6)*
MORM CONT	consonant	*59 (18)*	*21 (7)*	*62 (11)*	*30 (8)*
	vowel	*64 (10)*	*37 (10)*	*67 (7)*	*65 (8)*
MBC	consonant	*47 (16)*	*13 (10)*	*52 (15)*	*25 (23)*
	vowel	*61 (5)*	*31 (8)*	*68 (8)*	*46 (24)*

Table1: Averaged speech Intelligibility (%) results. Numbers in bracket are stand deviations.

6. CONCLUSIONS

Modern hearing aids require better output sound quality and higher speech intelligibility. But traditional multi-band compression methods distorted spectral cues and deteriorated speech intelligibility. A new morphology-base method was provided in this paper with much lower computation cost, and confirmed to be superior in most situations. Although behaving not very satisfying in noisy situation, MORM shows superiority compared to traditional multi-band compression method. Noise cancellation algorithms will be expected to deal with background noise after all. In addition, the capability that MORM can pick up primary formants of speech signal could also be significative for other algorithms or applications, such as speech enhancement and recognition.

Except for ability of accurate location the spectral cue points, the morphology-based method reduced computation consumption greatly and its simple execution requirement made it practical in most real-time hearing rehabilitation devices, including digital hearing aids. Now MORM is immigrated to a DSP-based digital hearing aid development platform – pDHA [11]. Long-term wearing tests will be carried out in future.

6. REFERENCES

[1] Xianbo X, Guangshu H. The progress of algorithms applied in digital hearing aid, Journal of Biomedical Engineering, Vol.21, No.4, pp.694-698, 2004

[2] Anna O B. More channels are better, right? http://www.audiologyonline.com

[3] A revolution: the first ChannelFree^{TM}: digital hearing system in the world, http://www.bernafon.com

[3] Young-Cheol P, Hae-su L, Yong-chun Kim, et al. High performance digital hearing aid processor with psychoacoustic loudness correction, International Conference on Consumer Electronics, June 11-13, 1997

[4] Juan C, Tejero C, Janet C R. Preserving spectral contrast in amplitude compression for hearing aids. Proceedings of the 23rd Annual EMBS International Conference, October 25-28, 2001

[5] Li M, McAllister HG, Black ND. Wavelet-based nonlinear AGC method for hearing aid loudness compensation. IEE Proceedings Vision Image Signal Processing. Vol. 147, No 6, pp.502-507. December 2000

[6] J. C. Tejero, S. Bernal, J. A. Hidaldo, et al. A digital hearing aid that compensates loudness for sensorineural hearing impairments. Proc. of ICASSP, vol. 1, pp. 2991-2994, 1995

[7] F. Asano,Y. Suzuki, T. Sone, et al. A digital hearing aid that compensates loudness for sensorineural impaired listeners. Proc. ICASSP91, pp. 3625-3628, 1991

[8] Yi C. Image processing and analysis: morphology method and application, Science Publishing House, 2002

[9] Dillon H. Hearing aids. Boomerang Press, Sydney, Australia, 2001

[10] A Maynard Engebretson. Benefits of digital hearing aids. IEEE Engineering in Medicine and Biology Magazine, 13(2), 1994, 238–248.

[11] Hui Zh, Zi G, Xianbo X, et al. An dsp-based digital hearing aid development platform, CBME2003, Wuxi, China, 2003