On the Origin of Multiplication of the Dominant Cepstral Peak

Yasuo Koike, Ghen Ohyama, Naoya Takeda and Kazuyo Sogawa

Department of Otolaryngology, The University of Tokushima School of Medicine
3 Kuramoto, Tokushima, JAPAN

Introduction

For some time it has been known that the dominant or highest cepstral peak, derived from the acoustic speech, is sensitive to the existence of laryngeal lesions[1]. Since this fact seemed clinically meaningful, a series of cepstral analyses were made on the acoustic speech of patients with various different laryngeal diseases, and the behavior of the dominant cepstral peak in relation to the pathology in the larynx was explored. From these studies it was found that dominant cepstral peak usually coincides with the peak representing the fundamental period on the quefrency axis, and that the height of the dominant peak correspond well with the affected voice quality[2].

It was found also, however, that in some limited number of patients the dominant peak was a second peak from the left side (low quefrency side) on the quefrency axis. That is, the peak representing the 1st sub-harmonic was taller than the peak indicating the fundamental period in some instances, though the difference in height between the two peaks was small. This phenomenon made the interpretation of the behavior of the dominant peak difficult. A further analysis of the data obtained from these patients was undertaken, in order to clarify the mechanism of this multiplication of the dominant cepstral peak.

Method

The acoustic speech signals of sustained vowel /a/, produced by 90 patients with various laryngeal diseases, were recorded on a digital audio tape-recorder (Sony TCD-D10). A stable portion of the vowel of about 1 second in duration was adopted for analysis. These signals were reproduced and fed into an A/D converter (μ DASBOX-12) connected to a computer (PDP 11/73), and were digitized at a sampling rate of 10KHz. The data were then analyzed with a cepstrum analyzer implemented on the computer. A Hamming window having a duration of 102 msec was employed.

The samples revealing a taller second peak than the first peak were selected, and the detailed behavior of the cepstral peaks in reference to the quefrency axis was studied on the chosen samples. The height of the cepstral peaks and the locations of the peaks on the quefrency axis were printed out, in order to investigate the above mentioned behavior, in addition to the conventional graphic display of the cepstrum. The pitch perturbation quotient (PPQ) and the amplitude perturbation quotient (APQ)[3] were also computed for these samples, in order to examine the effect of both pitch and amplitude perturbations on the behavior of the cepstral peaks.

A sustained vowel /a/ with certain pitch-period perturbation, and one with some amplitude perturbation, were synthesized with a D. Klatt-type formant synthesizer implemented on the PDP 11/73 computer, to further confirm the influence of period and amplitude perturbations on the cepstral peaks. The formant frequencies employed were 0.8KHz, 1.2KHz, 2.3KHz, 3.5KHz and 4.5KHz respectively.

The fundamental period of the synthesized vowel for pitch perturbation study was shifted from 5.0 msec to 5.1 msec alternately, while other parameters were kept constant. The peak amplitude of the synthesized vowel for amplitude perturbation study was alternately decremented by 10%, while other parameters including the fundamental period were kept constant. These synthesized speech samples were then analyzed with the cepstrum analyzer, exactly as the natural speech samples derived from the patients with laryngeal diseases.

Results

A cepstrum made from the acoustic speech of a normal female subjects is shown in Figure 1. An apparent dominant peak can be observed on the left side of the cepstrum. This peak has a quefrency of 4.0 msec, and coincide with the fundamental period of the utterance. The subsequent cepstral peaks are observed respectively at the locations of 8.0 msec, 12.0 msec, 16.0 msec, etc., i. e., at the integral multiples of the fundamental period, on the quefrency scale. These locations apparently coincide with the positions of the sub-harmonics.

A typical multiplication of the dominant cepstral peak is illustrated in Figure 2. The cepstrum on the left side is based on the speech signal recorded before the surgical operation for laryngeal granuloma, and the cepstrum on the right side is drawn from the acoustic speech after the operation. An apparent two peaks can be observed in the cepstrum on the right side. The second peak from the left side of this cepstrum is seen to be somewhat taller than the leftmost peak on the quefrency axis. A similar multiplication was observed in six cases out of the total 90 patients, and the data from these 6 cases were further examined.

The numerical values of the height of the relevant cepstral peaks are listed on the rightmost column of Figure 3. The 2nd peak is apparently higher than the 1st peak. It is interesting to note that the relationship between the locations of the two peaks is different from that for normal subject shown before. Although it is difficult to recognize in the graph, it is found from the numerical values listed here, that the 1st cepstral peak consists of 2 peaks, which are located respectively at 4.5 msec and at 4.6 msec on the quefrency axis. Apparently, this utterance has two fundamental periods, which are slightly different from each other. The second large cepstral peak is located at 9.1 msec, which is the sum of 4.5 msec and 4.6 msec, and not at an integral multiple of either fundamental period. From these data it is presumed that the multiplication of the dominant peak in this cepstrum is due to a regular perturbation in the fundamental period of this utterance.

A similar phenomenon can be observed in the cepstrum in Figure 4, which has been derived from the acoustic speech produced by a patient with unilateral recurrent nerve paralysis. Although in this case the dominant cepstral peak is not quite explicit, the 2nd peak is somewhat taller than the 1st peak. The location of the 2nd peak is seen to be at 19.5 msec. on the quefrency axis, which is equal to the sum of 9.7 msec and 9.8 msec, which are presumed to be the two fundamental periods of this utterance. In two cases out of the six patients, however, the location of the second peak could not be accounted for that clearly.

The cepstrum shown in Figure 5 has been obtained from an utterance of a patient with vocal fold atrophy. A high cepstral value is seen at the location of 6.9 msec and at 7.0 msec, forming a rather wide 1st cepstral peak. The second peak is seen to locate at 13.7 msec, which is different from the sum of 6.9 and 7.0 (6.9+7.0=13.9), even though the cepstral value of 1.260 at 13.9 msec also is considerably high. A different mechanism, such as an amplitude modulation, may have to be assumed in this case, in order to account for the elongation of the 2nd cepstral peak.

Figure 6 shows a cepstrum of a synthetic speech with a type of pitch perturbation. As known from the values listed on the right column of this Figure, the synthesized speech adopted here has two fundamental periods of 5.1 msec and 5.0 msec, which appear alternately. That is, a combination, consisting of a long pitch period followed by a short pitch period, has been repeated. A remarkable elongation of the 2nd cepstral peak can be observed at the location of 10.1 msec, which is the sum of 5.1 msec and 5.0 msec. This fact evidently confirms that perturbation in the fundamental period can be a cause of multiplication of the dominant cepstral peak.

Figure 7 demonstrates a cepstrum of a synthetic speech with an amplitude perturbation. The fundamental period of this sample was kept constant. The peak amplitude corresponding to each fundamental period was modulated alternately by a factor of 0.9. An apparent elongation of the 2nd cepstral peak can be observed in this cepstrum. The peak representing the fundamental period is seen at the position of 5.0 msec, and the subsequent peaks are found at the locations corresponding to the integral multiples of the fundamental period. It is apparent that an amplitude modulation also can produce the elongation of the 2nd cepstral peak, even without any perturbation in the fundamental period.

Discussion

From the data presented above, it is evident that perturbation in the fundamental period, as well as in the peak amplitude, of acoustic speech can cause an elongation of the 2nd cepstral peak, revealing multiple dominant peaks in the cepstrum. Since in the real speech both the fundamental period and the peak amplitude may be perturbed in a complex manner, it may not be always easy to dermine which of these two elements is the real cause of the elongation. Though in the most cases examined in the present study the elongation of the 2nd cepstral peak was judged to be due to a perturbation in the fundamental period, there were certain cases with an elongation of undetermined origin. Although it can be presumed that in these cases the elongation of the 2nd peak is due to some perturbation in the amplitude, it is not quite easy to prove this in the real speech. A further study may be needed, in order to completely understand the detailed behavior of the cepstral peaks in relation to the acoustic characteristics of speech.

When an elongation of the 2nd cepstral peak takes place, the 2nd peak becomes taller than the first peak. The term "dominant peak" may imply the 2nd peak in such an instance. This makes the interpretation of the term "dominant peak" unclear. It may be wise, therefore, to adopt a term "fundamental cepstral peak" to indicate the cepstral peak representing the fundamental period, even though in most cases the dominant peak may coincide with the fundamental peak.

The physiological mechanism for revealing perturbations in the acoustic speech signal is not quite clear yet. Although it was suggested in some literature[3] that physical characteristics of the both vocal folds are playing an important role, detailed mechanism of yielding acoustic perturbation has to be further investigated. The vibratory behavior of the folds in the pathologic larynges, in particular, needs further study, since certain acoustic features, such as an elongation of the 2nd cepstral peak for example, become explicit when a pathology takes place in the larynx. This type of physiological study, in turn, should provide a firm basis for the interpretation of acoustic characteristics of speech.

Summary

Cepstrum analysis was performed on the sustained vowel /a/ samples produced by 90 patients with various laryngeal diseases. A multiplication of the dominant cepstral peak was found in 6 cases. It was found in 4 out of the 6 cases that the 1st cepstral peak consisted of two peaks, corresponding to two slightly different fundamental periods, and that the 2nd cepstral peak located at the quefrency value which is equal to the sum of the fundamental periods.

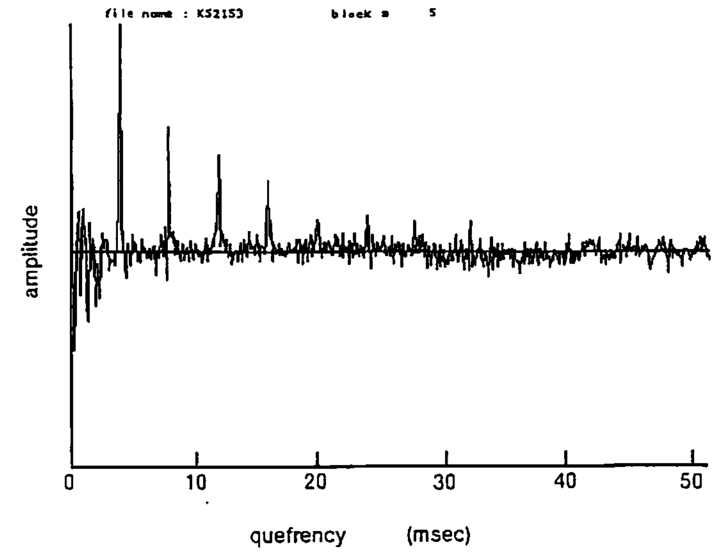
A synthetic speech having a type of perturbation in the fundamental period, and one having an amplitude perturbation, were made with a formant synthesizer implemented on a computer. Cepstrum analyses on these synthetic speech samples revealed an apparent multiplication of the dominant cepstral peak, and suggested that both period perturbation and amplitude perturbation can cause a multiplication of the dominant cepstral peak. The use of the term "fundamental cepstral peak ", instead of "dominant cepstral peak", was proposed.

Reference

1) Koike, Y.: Cepstrum analysis of pathologic voices. J.of Phonetics 14, 501-507, 1986.

2) Koike, Y. and Kouda, J.: The Effect of Laryngeal surgery on the Speech Cepstrum. Ann. Bull. RILP 22, 73-81, 1988.

3) Koike, Y., Takahashi, H. & Calcaterra, T.: Acoustic measures for detecting laryngeal pathology. Acta Oto-Laryngological 84, 105-117,1977.



quefrency (msec)	quepstrum
3.9	0.881
4.0	3.846
4.1	0.259
7.9	0.918
8.0	2.134
8.1	0.263
11.9	0.399
12.0	1.666
12.1	0.867

Figure 1. Normal cepstrum. An apparent dominant cepstral peak can be observed, coinciding with the fundamental period. The abscissa indicates quefrency in msec, and the ordinate the cepstral energy in an arbitrary scale.

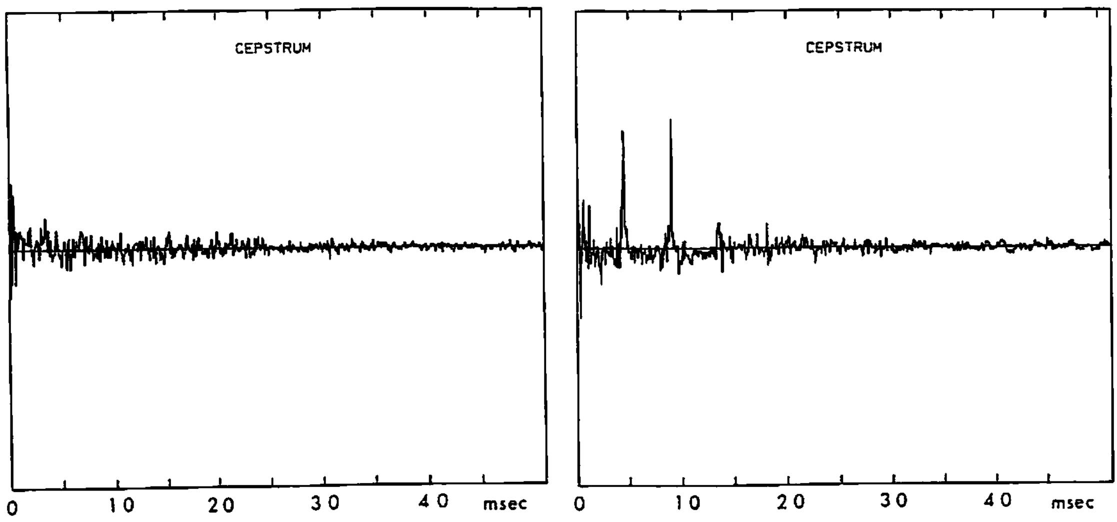
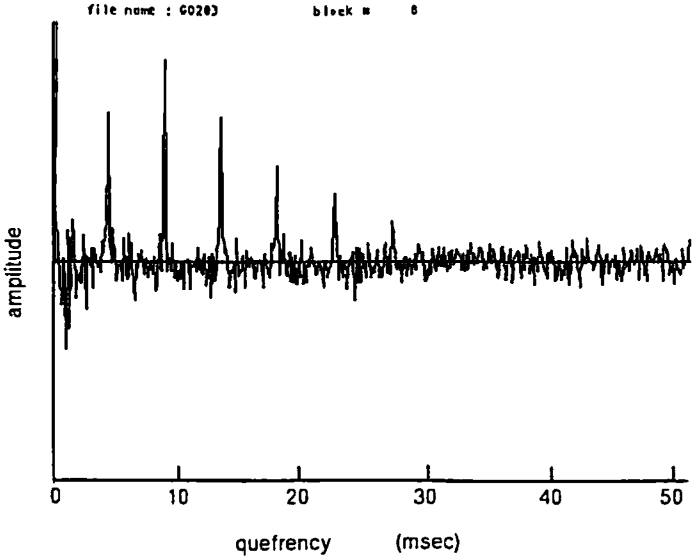
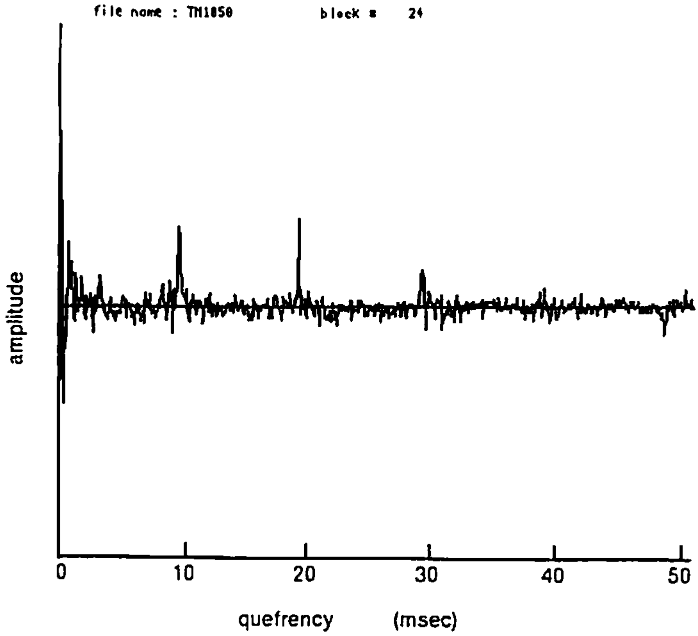


Figure 2. A typical multiplication of the dominant cepstral peak. The tracing on the left side shows a cepstrum derived from the acoustic speech of s patient with laryngeal granuloma before surgery. No apparent peak can be found. The tracing on the right side depicts a cepstrum of the same patient after the surgical operation. Two explicit peaks are found. The second peak from the left is somewhat taller than the first peak.



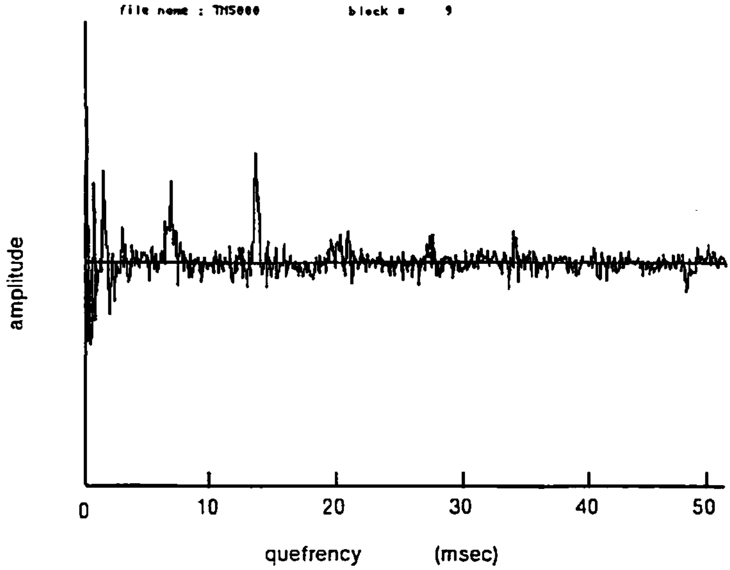
quefrency (msec)	quepstrum
4.4	0.50
4.5	1.49
4.6	1.93
4.7	0.34
9.0	0.68
9.1	2.61
9.2	0.71

Figure 3. A cepstrum with multiplication of the dominant cepstral peak. The first peak consists of two peaks which represent two fundamental periods of 4.5 msec and 4.6 msec respectively. The second cepstral peak is located at 9.1 msec, which is equal to the sum of 4.5 msec and 4.6 msec, on the quefrency axis.



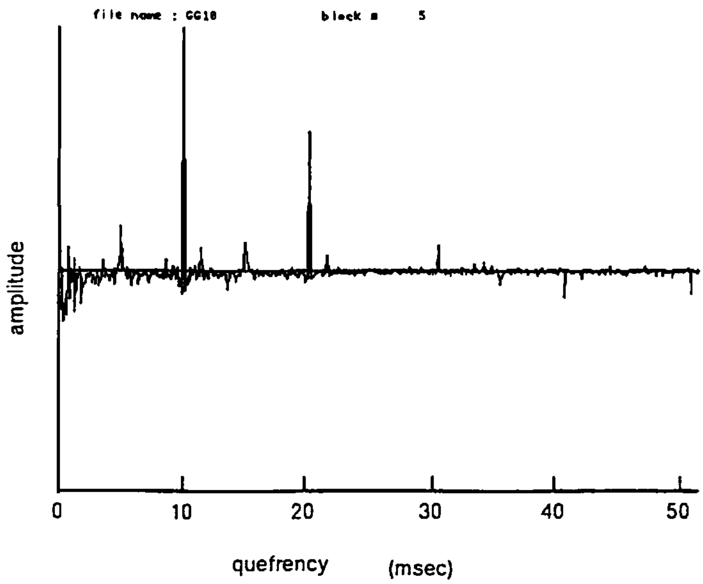
quefrency (msec)	quepstrum
9.6	0.553
9.7	1.703
9.8	1.470
9.9	1.023
10.0	0.312
19.4	0.167
19.5	1.896
19.6	0.923
28.9	0.109
29.0	0.171
29.1	- 0.022

Figure 4. A cepstrum of the acoustic speech produced by a patient with unilateral paralysis. The first peak consists of two peaks having quefrency values of 9.7 msec and 9.8 msec. The second peak is somewhat taller than the first peak.



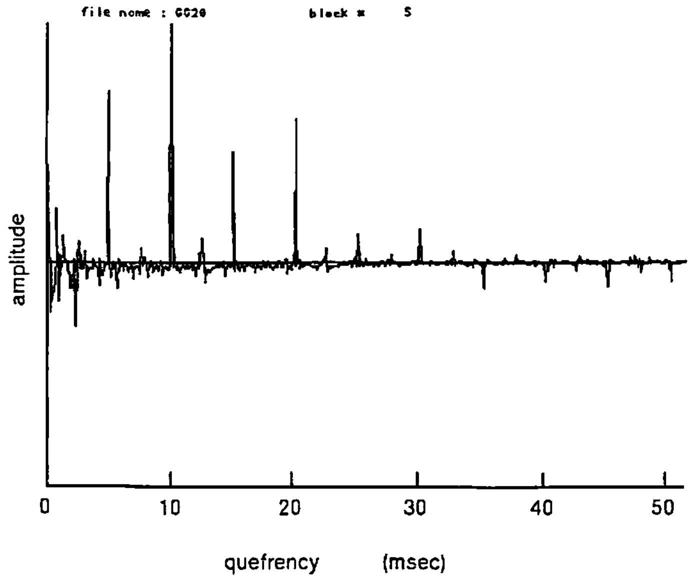
quefrency (msec)	quepstrum
6.8	0.596
6.9	1.549
7.0	1.062
7.1	0.336
13.5	0.560
13.6	1.119
13.7	2.078
13.8	1.169
13.9	1.260
20.9	0.350
21.0	0.591
21.1	0.223
21.2	0.211

Figure 5. A cepstrum derived from an utterance produced by a patient with vocal atrophy. The second cepstral peak is taller than the first peak. The location of the 2nd peak, however, does not coincide with the value of the sum of the fundamental periods.



quefrency (msec)	quepstrum
4.9	0.38
5.0	0.79
5.1	1.78
5.2	0.13
10.0	- 0.89
10.1	9.75
10.2	- 0.77

Figure 6. A cepstrum of a synthetic speech with period-perturbation. A remarkable elongation of the 2nd cepstral peak can be observed at the location of 10.1 msec on the quefrency axis. This synthetic speech has two fundamental periods of 5.0 msec and 5.1 msec, each of which appear alternately. The location of the 2nd peak on the quefrency axis coincide with the sum of these fundamental period values.



quefrency (msec)	quepstrum
4.9	- 0.04
5.0	4.94
5.1	- 0.04
9.9	- 0.11
10.0	6.82
10.1	- 0.12

Figure 7. A cepstrum of a synthetic speech with an amplitude modulation. An apparent elongation of the 2nd cepstral peak can be observed. The peak amplitude of this synthetic speech has been modulated regularly. The cepstral peaks are found at the integral multiples of the fundamental period.