ICIC Express Letters
Part B: Applications
Volume **7**, Number **9**, September **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **2007–2012**

COMPARISONS OF CONTEXT-DEPENDENT SUB-WORD TRANSFER FUNCTIONS FOR THE SPEECH SUPPORT SYSTEM

Yibing Cheng, Masashi Nakayama and Shunsuke Ishimitsu

Graduate School of Information Sciences
Hiroshima City University
3-4-1 Ozuka-Higashi, Asa-Minami-Ku, Hiroshima 731-3194, Japan
{ masashi; ishimitu }@hiroshima-cu.ac.jp

Received February 2016; accepted May 2016

Abstract. *Speech is an important communication method for people. However, those with speech disorders use a form of substitute speech and are faced with speech communication problems because the substitute speech does not have a clear sound quality at loud volumes. Owing to this problem, these substitute forms are stressful and disappointing to use for communication in daily life. To deal with these problems, we proposed and experimented with the development of the speech support system using transfer functions as a solution. The system aims to help people communicate by improving the sound quality of speech using synthesis filters developed by previous research. This paper investigates the improvement of sound quality that occurs when analysis and synthesis filters are used in different contexts of sub-words conditioned on the head, middle, and tail parts of a word. Experimental results were obtained to find the best high frequency performance using the linear predictive coding analysis which estimates transfer function at the head part of a word with a pre-emphasis method.*
Keywords: Body-conducted speech, Linear predictive coding, Context-dependent transfer function, Pre-emphasis

1. **Introduction.** Speech communication is very important for many people. However, approximately 42,000 people are affected by impairments of speech, language, and swallowing mastication in Japan [1]. It is difficult to communicate fluently with others, and they become greatly distressed while talking. To address this problem, S. A. Selouani et al. discussed improving the intelligibility of pathologic speech, and making it as natural, and as close to the original voice of the speaker as possible [2]. M. S. Hawley et al. proposed that VIVOCA, the voice-input voice-output communication aid, recognizes the disordered speech of the users and then builds messages, which are converted into synthetic speech [3]. In addition, the authors also proposed and developed a speech support system using body-conducted speech (BCS) for speech disorders [4].

In previous research, an improvement in sound quality of BCS was conducted with linear predictive coding analysis (LPC) estimated spectral envelopes of the formant frequencies [5] and confirmed for specific speech disorders using an LPC sub-word unit transfer function. The transfer function consisted of LPC coefficients using a healthy person's speech [6] as a sub-word unit filter. However, we still need to investigate the effectiveness of context-dependent transfer functions for sub-words, because the speech has time-varying and non-stationary characteristics for each position of the word. We addressed the position of each sub-word on the head, middle and tail in a word.

2. **Speech Support System Using BCS for Speech Disorders.** The common solutions used to deal with such communication disorders are esophagus vocalization and electric throat vocalizations. However, esophagus speech sounds are not a clear by comparison with normal speech, as the disorders cannot provide a sufficient volume of speech

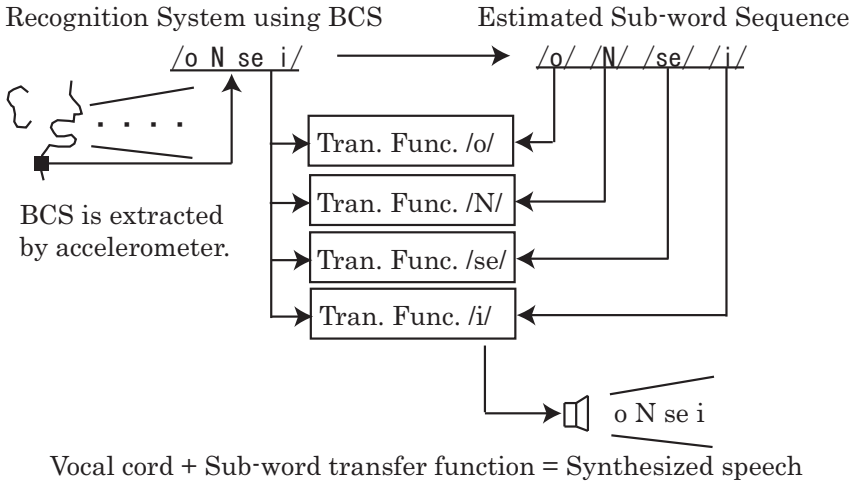


FIGURE 1. Speech support system using BCS for speech disorders [4]

or frequency for use in daily life. However, it is difficult to express the nonverbal information and achieve a smooth conversation for others.

Figure 1 shows the speech support system using BCS for speech disorders. Since BCS is a noise robust signal compared to normal speech, it was not affected in the 98dB SPL and −20dB SNR environments [7]. Though the speech sound quality of disorders is unclear at loud volumes, the use of the BCS as a noise robust signal can solve the problem. For instance, the frequency characteristic of BCS lacks the high frequency elements, etc. Then, the proposed system improves its sound quality using speech recognition and transfer functions for advance. The concepts of the system are based on the source-filter theory in a mechanism of vocalization. These sounds are synthesized via an articulation filter based on frequency responses of the oral nasal cavities [7]. This system can generate clear speech using transfer functions of individual, and include the features of the speech disorders. The procedures of the system are as follows.

1. BCS picks up from the upper lip using an accelerometer.
2. The system estimates the sub-word sequence using the sub-word BCS recognition.
3. The system synthesizes clear speech using the sub-word transfer function according to the sub-word sequence calculated in 2.
4. The system translates into the clear speech during a conversation.

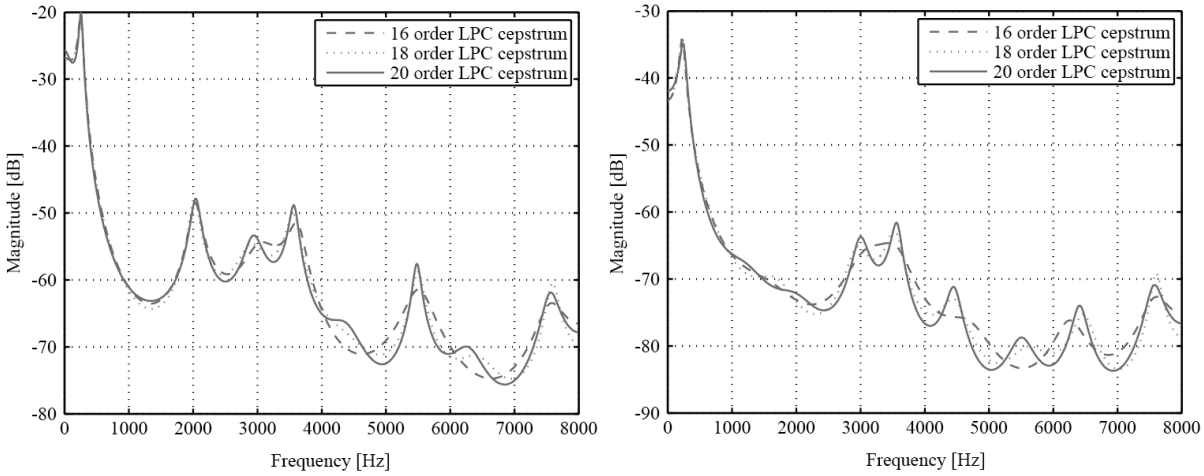
3. **Frequency Characteristics of Context-Dependent Sub-Word Transfer Function.** Previous studies have confirmed for each isolated sub-word transfer function [6] and discussed the context-dependent transfer function using the LPC method. The conventional system is only evaluated using the transfer functions estimated with the context-independent sub-word transfer function. In this paper, the transfer functions estimated with the context-dependent sub-word units are applied to improving its accuracy. Five Japanese vowels, /a/, /i/, /u/, /e/, and /o/ were employed as the context-dependent sub-words and especially the result of /i/ is described in this paper.

3.1. **Experimental setup.** For the experiment, we first prepared the database containing speech and BCS uttered JEIDA 100 Japanese local place names spoken by Japanese speakers who have no speaking disorders. The recognition decoder, Julius, was employed to estimate the boundary of a sub-word in Japanese for syntheses clarifying speeches using transfer functions and vocal code sound from BCS. The recognition demonstrated as supervised recognition, which estimated only the time information of sub-words for synthesis of demonstration. The synthesis sound estimates with analysis filters and vocal code sound. The transfer functions which are coded by 16, 18, and 20 coefficients

using LPC, and vocal code sounds are generated from BCS using analysis filter as transfer function between vocal code and mouth conditioned on context-dependent sub-word. However, the magnitudes of frequency components are dramatically decreasing along the high frequency components. So, we experimented the comparisons of transfer functions with or without using pre-emphasis processing at Vowel /i/.

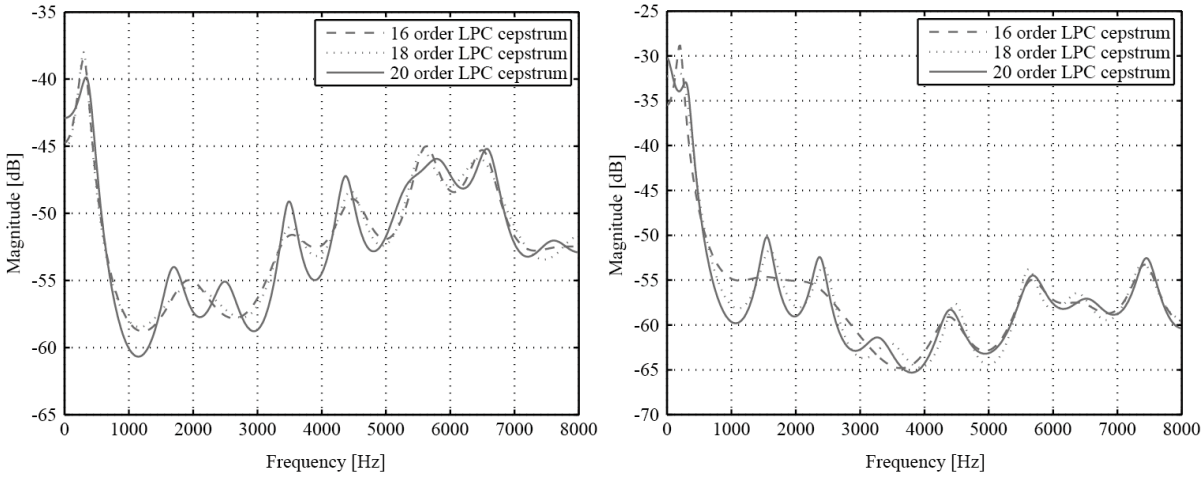
3.2. **Experimental result.** Figures 2 to 4 show the frequency characteristics of LPC of sub-word /i/ in the different word positions. Vowel /i/ is prepared on each position of a word in the recorded database of sounds. From these figures, the transfer function of the sub-word can be seen to have an approximate representation in the LPC when it is from the same position in the word. However, the magnitudes are decreasing along to high frequency direction, because they can be seen to need pre-emphasis processing for them.

Figures 5 to 7 show the frequency characteristics of the sub-word /i/ in the different word positions using pre-emphasis processing. Comparing Figures 5 to 7 with Figures 2 to 4, 20 orders of LPC captured each formant respectively, and 7 to 10 peaks to estimate the frequency characteristics of the sub-word.



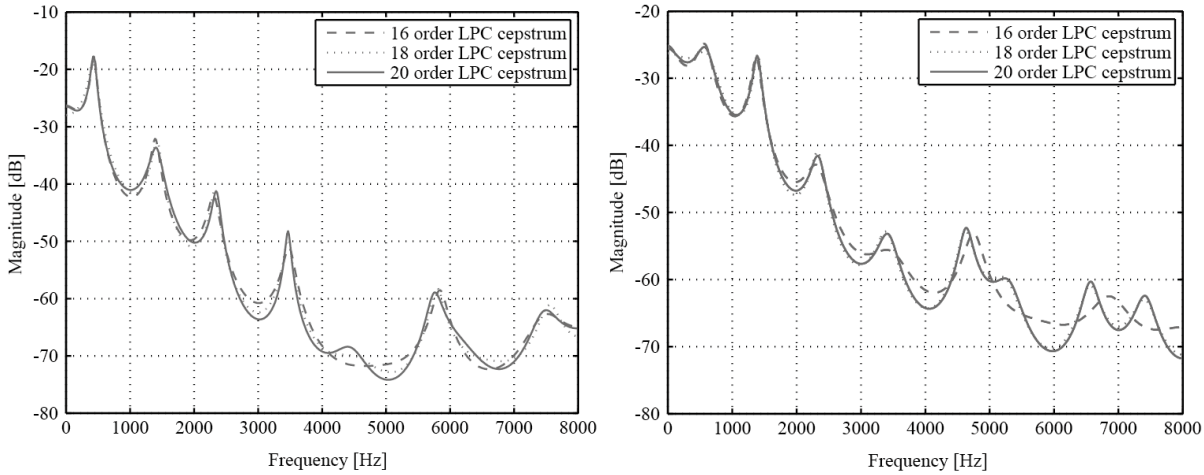
(a) /i/ from word, Ibusuki (b) /i/ from word, Iyo

FIGURE 2. Spectral envelope of sub-word /i/ using LPC at head position



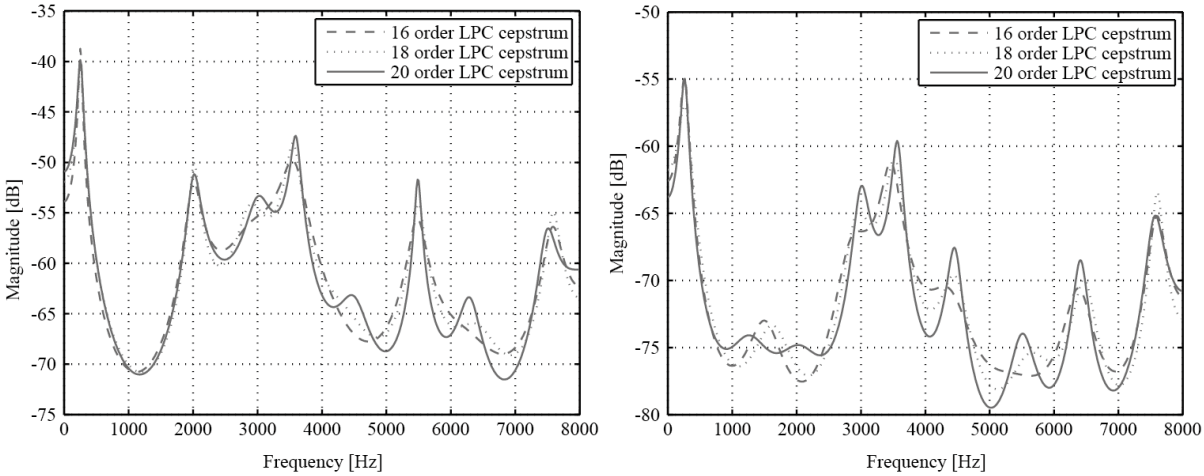
(a) /i/ from word, Kuroiso (b) /i/ from word, Niiza

FIGURE 3. Spectral envelope of sub-word /i/ using LPC at middle position



(a) /i/ from word, Rumoi (b) /i/ from word, Hisai

FIGURE 4. Spectral envelope of sub-word /i/ using LPC at tail position

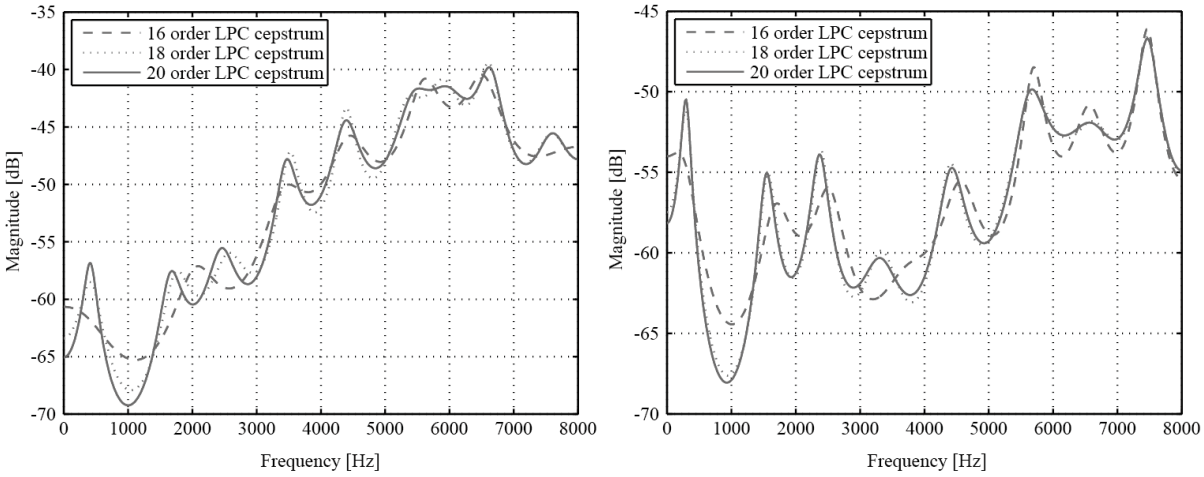


(a) With pre-emphasis, Ibusuki (b) With pre-emphasis, Iyo

FIGURE 5. Spectral envelope of sub-word /i/ using LPC and pre-emphasis at head position

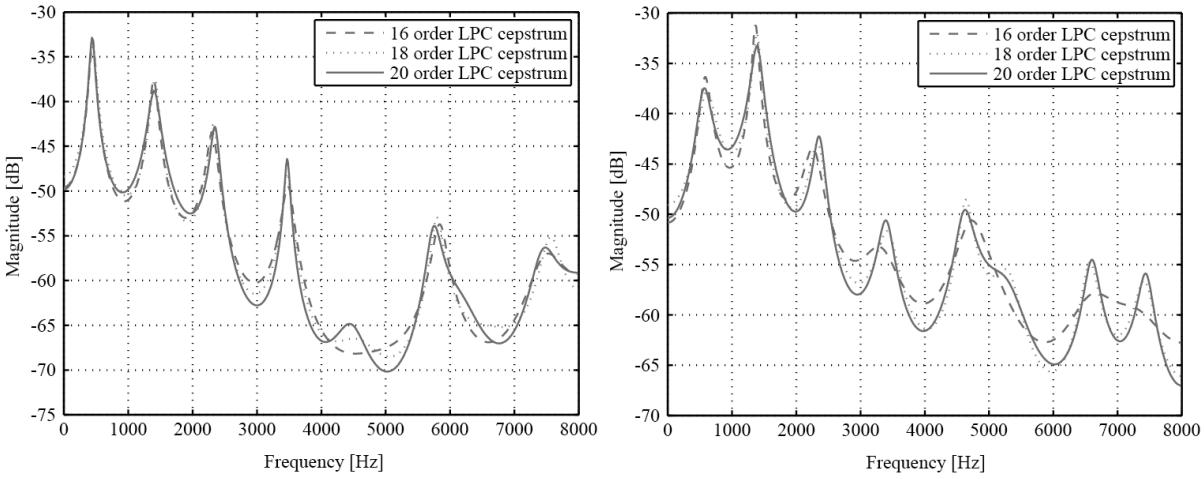
3.3. **Discussion.** The same sub-word shows different frequency characteristics because of the dependency on the sub-word positions in the word. In the spectral envelopes from the LPC method, the 20 LPC orders represented 7 to 10 formants in the frequency characteristics. Additionally, pre-emphasis revealed the different results in this experiment. The sub-word at the head position using pre-emphasis processing showed better performance in the high frequency section rather than at the middle and tail positions.

4. **Conclusions and Future Work.** In this study, we discussed and experimented the sub-word's context-dependency with the comparisons of the LPC coefficients positioned in a word. The results showed that the most appropriate way to estimate the transfer function database in our speech support system is by considering the spectral envelope of sub-word using the LPC method at the head of the word on pre-emphasis processing. In future, we will evaluate many number of sub-words conditioned on the context of the sub-word and further develop the database of transfer functions for the speech support system. Additionally, we also plan to develop the system with user-friendly interface, and will ultimately provide better support for speech disorders.



(a) With pre-emphasis, Kuroiso (b) With pre-emphasis, Niiza

FIGURE 6. Spectral envelope of sub-word /i/ using LPC and pre-emphasis at middle position



(a) With pre-emphasis, Rumoi (b) With pre-emphasis, Hisai

FIGURE 7. Spectral envelope of sub-word /i/ using LPC and pre-emphasis at tail position

REFERENCES

[1] Cabinet Office, Government of Japan, *The White Paper on Disorders on 2011*, http://www8.cao.go.jp/shougai/whitepaper/h23hakusho/zenbun/index.html.

[2] S. A. Selouani, M. S. Yakoub and D. O'Shaughnessy, Alternative speech communication system for persons with severe speech disorders, *EURASIP Journal on Advances in Signal Processing*, no.6, 2009.

[3] M. S. Hawley, S. P. Cunningham, P. D. Green, P. Enderby, R. Palmer, S. Sehgal and P. O'Neill, A voice-input voice-output communication aid for people with severe speech impairment, *Trans. Neural System and Rehabilitation Engineering*, vol.21, no.1, 2013.

[4] M. Nakayama and S. Ishimitsu, Speech support system using body-conducted speech recognition for disorders, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4255-4265, 2009.

[5] K. Oda, S. Ishimitsu, M. Nakayama, K. Makiyama and S. Horihata, The evaluation of speech recognition system for disorders, *The Local Branches of the JSME, The 48th Meeting of Conference, Proc. of the Japan Society of Mechanical Engineers*, pp.359-360, 2010.

[6] T. Yamanaka, S. Ishimitsu and K. Fukui, Study of improving sound quality in support system for speech impaired, *ICIC Express Letters, Part B: Applications*, vol.5, no.2, pp.595-600, 2014.

[7] S. Ishimitsu, M. Nakayama, T. Yoshimi and H. Yanagawa, Noise-robust recognition system making use of body-conducted speech microphone, *AES the 122nd Convention*, Vienna, Austria, 2007.

[8] I. Karl, P. Nordstrom and F. Driessen, Variable pre-emphasis LPC for modeling vocal effort in the singing voice, *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, pp.18-20, 2006.