An LP-based blind Model for Restoring Bone-conducted Speech

Thang tat Vu, Masashi Unoki, and Masato Akagi
School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
Email: {vu-thang, unoki, akagi}@jaist.ac.jp

Abstract—**Due to the stability against the external noise, bone-conducted (BC) speech seems better to be used instead of noisy air-conducted speech in an extremely noisy environment. However the quality of bone-conducted speech is very low and restoring bone-conducted speech is a challenged topic in speech signal processing field. As the main issue to improve the BC speech, many studies try to model and resolve the degradation when the signal is conducted through bone transduction. In previous study, we proposed a linear prediction (LP) based blind-restoration model. In this paper, we therefore completely evaluated the proposed model in comparison with other models to find out whether our proposed model could adequately improve voice quality and the intelligibility of BC speech, using objective measures (LSD, MCD, and LCD) and carrying out Japanese word-intelligibility tests (JWITs), Vietnamese word-intelligibility tests (VWITs) and Modified Rhyme Tests (MRTs) for English. The results of experiments on different languages, i.e. Japanese, English and Vietnamese proved the practicability of blind-BC restoration.**

Key words- Speech intelligibility, Bone-conducted (BC) speech, Simple recurrent network (SRN), Blind restoration.

I. INTRODUCTION

It is required for safe and secure speech communication while the interfering noises in extremely noisy environments bring significant difficulties for the commutations of human and also automatic speech recognition (ASR) systems. This problem is from the low sound-quality and low intelligibility of speech, due to the influence of the transmission environment. As a solution, many different complex models and algorithms have been used to cancel or reduce these noisy affections but have been only efficient at low- and medium-noise levels. When the noise levels are extremely high, this solution seems to perform ineffectively.

Another possible solution is to use a special microphone to record the speech signals transmitted through the speaker's head and face [1-5]. This recorded signal is referred to as "bone-conducted (BC) speech". Its stability against interfering noise from noisy environments makes BC speech more advantageous than noisy air-conducted (AC) speech.

There are two main drawbacks of BC speech, the degradation due to bone-conduction and the changing of speaker's pronunciations due to surrounding noise referred as Lombard effect. While the Lombard effect is a usual problem as the same as AC speech in noisy environments, the other is critical affection to the speech quality. When the signal is transmitted through bone-conduction, it is complexly affected at a loss of sound quality and speech intelligibility. The degradation varies for different pick-up points (BC microphone positions), speakers, and pronounced syllables. This is because the characteristics of bone-conduction vary for different measuring positions and the distribution of frequency components varies with speakers who pronounce syllables differently. Regarding this main issue, we show a possibility of a blind restoration method for restoring quality and intelligibility of BC speech.

There are several studies on BC speech for applications such as human-hearing aids and machine-hearing systems but the results are still limited. A Gaussian mixture model (GMM)-based voice conversion model was applied to restore body-transmitted speech, which is like BC speech [6]. Due to the difficulty of dealing with F0 features that might cause synthesis problems, this approach has only been applied to unvoiced speech such as whispered speech. In other approaches such as when air-and-bone conductive micro-phones are used, BC speech has been used for ASR systems [7]. However, it has just been an additional source and helps to reduce external noise from noisy AC speech.

Our approach here is to improve BC speech signal and apply directly the restored signals in speech applications in noisy environments with greater efficiency instead of using noisy AC speech. This is even more extremely challenging to blindly restore the signals of BC speech without using any other information of AC speech.

Information on AC speech is usually needed to construct the inverse filtering in restoration models [1-3] and this is a serious drawback in practice when we have no information on AC speech. To construct its associated inverse filtering, the cross-spectrum and long-/short-term Fourier transform methods [1, 2] depends on the AC spectrum, the MTF-based model depends on the gain of power envelope of AC speech and the LP-based model depends on AC-LP coefficients [3]. Averaged gains or averaged AC-LP coefficients can be chosen for constructing an averaged filtering but the models are difficult to adapt to BC speech signal.

Since LP-based model only depends on a few unknown LP coefficients of AC speech (AC-LP coefficients), we proposed a blind restoration model [4], a machine learning method was applied to predict various LP coefficients of AC speech signal from LP coefficients of BC speech signal. The drawback of

1-4244-2426-9/08/$20.00 ©2008 IEEE

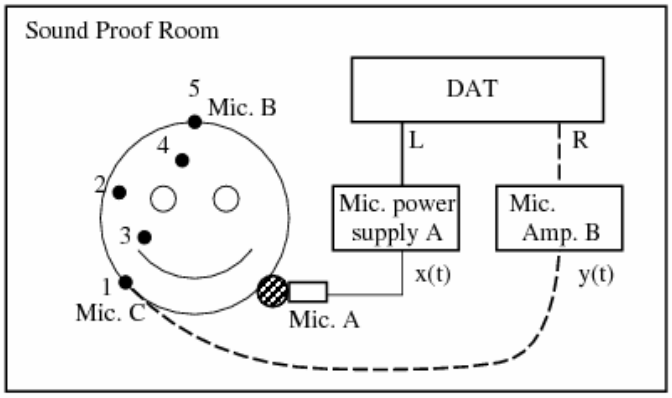

Figure 1. Environment for recording AC/BC speech.

this model was that the LP coefficients were not suitable to enable prediction with statistical models due to the different roles played by LP coefficients and their relatively large dynamic ranges. Besides, the inverse filtering of this model was not adapted to the short-term changing of BC speech signal. However, although the model suffered from above limitations, we obtained reasonable results. These results revealed the existing useful relationship between the AC-LP and BC-LP coefficients for restoring BC speech.

Therefore, we improved the previous LP-based model by (1) extending long-term to frame-based processing, (2) using LSF coefficients on LP representations, and (3) predicting LSF parameters on a frame-by-frame basis via a recurrent neural network. Since LSF coefficients play the same role in the presentation of the spectrum envelope and their values are limited within a range $(0, \pi)$, these coefficients could help to alleviate the limitations with LP coefficients in prediction using statistical methods. The processes of restoration on a frame-by-frame basis could also be adapted to inverse filtering in real time. In addition, since the restoration of neighboring frames should be related, a recurrent network was applied to predict BC-LSF coefficients to complete the blind restoration system.

In this paper, we completely evaluated the proposed model in comparison with other models to find out whether our proposed model could adequately improve voice quality and the intelligibility of BC speech, using objective measures (LSD, MCD, and LCD) and carrying out Japanese word-intelligibility tests (JWITs), Vietnamese word-intelligibility (VWIT) and Modified Rhyme Tests (MRTs) for English. The results of experiments on different language, i.e. Japanese, English and Vietnamese proved the practicability of blind-BC restoration.

The rest of this paper is organized as follows. the next section briefly describes AC/BC speech databases that we constructed for English, Japanese and Vietnamese languages. In section 3, we describe the restoration modes based on LP methods and in details about our proposed LP-based model (LSF model). The incorporating of a simple recurrent network (SRN) into the LSF model brings us the ability of blind restoration for BC speech. In section 4, we completely evaluate the proposed models in comparison with other models. Both objective and subjective measures are carried out for evaluating the restoration ability of models on different BC databases.

TABLE I. LIST OF EQUIPMENTS

Measurement site	Soundproof room
Number of Pickup points	5
Number of speakers	10 (Jp), 6 (En), 10(Vie)
Recorder	MARANZ, PMD671
Format	16 bits PCM
Sample rate	48 kHz
Number of channels	2 (Left:AC, Right: BC)
Mic. A for AC speech	SONY, C536P
Mic. power supply A	SONY, AC148F
Mic. B for BC speech	TEMCO, HG-17
Mic. C for BC speech	TEMCO, SK1
Mic. amp. B and C	Handmade

Finally, section 6 concludes with a summary and mentions future work.

II. AC/BC SPEECH DATABASE

We assumed that there were existing relationships between AC and BC speech that were significant in restoring BC speech. Therefore, a database was essential for analyzing the relationships and differences between BC speech and clean AC speech signals before any models were used to restore BC speech. We constructed large-scale databases of English, Japanese and Vietnamese, containing pairs of BC and clean AC speech signals recorded simultaneously.

Figure 1 and Table I show the environment and equipments used to construct databases. The BC speech was collected at five different pick-up points on the head and face: the (1) mandibular angle, (2) temple, (3) philtrum, (4) forehead, and (5) calvaria. Microphone B was used at the pick-up point (5) and microphone C was used at the others.

Six speakers (three males, three females) participated in the recording of 300 English words in the MRT list.

In the Japanese database, ten speakers (five males and five females) participated in the recording of 100 words and all 101 syllables. The 100 Japanese words were chosen by the degree of familiarity in NTT-AT 2003 database [3], 25 words for each familiarity range: R1 (1.0,2.5) - low, R2 (2.5,4.0) - medium low, R3 (4.0,5.5) - medium high, and R4 (5.5,7.0) - high. It is known that there is a complementary relationship between familiarity and intelligibility [5]. The selection of words in different familiarity ranges would help us to carry out better JWIT.

In the Vietnamese database, ten speakers (five males and five females) participated in the recording of 100 words which were chosen by the frequencies indexes and whether the word type is mono-syllabic or duo-syllabic. There are 30 mono-syllable words for each level of frequencies indexes: C1 (mono, low), C2 (mono, high), and also 20 duo-syllabic words for each of two level: C3 (duo, low), C4 (duo, high). It is known that there is a complementary relationship between frequency index and intelligibility [3]. Beside, it is usually more difficult to understand a mono-syllabic word than a duo-syllabic word even for a Vietnamese. The selection of words in different frequency ranges would help us carry out better VWITs.

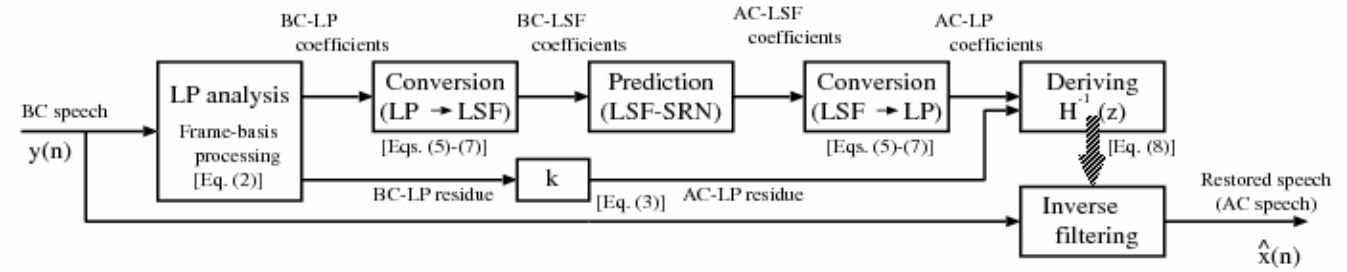

Figure 2. Block diagram of proposed model - LSF-SRN

III. RESTORATION APPROACHES

A. LP Representation

Let *x(n)* and y(n) be AC and its associated BC speech. The signals are represented by LP model in the *z*-domain as:

$$-G_x(z) = X(z)\sum_{i=0}^{P} a_x(i)z^{-i}, \quad a_x(0) = -1, \tag{1}$$

$$-G_y(z) = Y(z)\sum_{i=0}^{Q} a_y(i)z^{-i}, \quad a_y(0) = -1, \tag{2}$$

where *X(z)* and *Y(z)* are the *z*-transforms of *x(n)* and *y(n)*, *P* and *Q* are LP orders, and $a_x(i)$ and $a_y(i)$ are *i*-th LP coefficients. Here, Gx(z) and Gy(z) are the z-transforms of the LP residues $g_x(n)$ and $g_y(n)$.

Since the LP residues are related to the source information (glottal information), we can approximately assume the residue ratio as a constant factor, *k*, as [3]

$$G_y(z)/G_x(z) = k \quad \text{(const.)}. \tag{3}$$

The inverse filter $H^{-1}(z)$ to restore BC to AC speech is derived from Eqs. (1) – (3), simply as

$$H^{-1}(z) = \frac{X(z)}{Y(z)} = k \cdot \frac{\sum_{i=0}^{Q} a_y(i)z^{-i}}{\sum_{i=0}^{P} a_x(i)z^{-i}}. \tag{4}$$

The inverse filtering $H^{-1}(z)$ can be decomposed into two parts. In the first part, the constant value, *k*, can be chosen manually and used to control the magnitude of restored speech. The second part primarily depends on the LP coefficients of signals. Although unknown AC-LP coefficients could be predicted from BC-LP coefficients with some reasonable results [4], they were inappropriate for statistical models in prediction because they played different roles and had a relatively wide dynamic range. LSF coefficients should be a better choice. They have both a well-behaved dynamic range and can be used to encode LP spectral information more efficiently than any other parameters [5]. The almost equivalent roles of LSF coefficients, on the other hand, would be suitable for statistical models.

B. LSF Representation

Let *A(z)* be a general LP filter on an LP representation. The LSF coefficients, ω_i and θ_i, can be derived from a symmetric polynomial *U(z)* and an anti-symmetric polynomial *V(z)*, as the phase of conjugated zeros.

$$A(z) = \sum_{i=0}^{P} a(i)z^{-i}, \quad a(0) = 1 \tag{5}$$

$$U(z) = A(z) + z^{-(P+1)}A(z^{-1}), \tag{6}$$

$$V(z) = A(z) - z^{-(P+1)}A(z^{-1}). \tag{7}$$

Substituting Eqs. (5)-(7) into Eq. (4), we can obtain the equation for the inverse filtering which depends on the LSF coefficients instead of the LP coefficients

$$H^{-1}(z) = k \cdot \frac{U_y(z) + V_y(z)}{U_x(z) + V_x(z)} \tag{8}$$

C. Blind BC restoration model

Figure 2 shows a block diagram of the LP-based blind restoration model. In this section, we explain how to predict AC-LSF coefficients from BC-LSF coefficients.

The problem: Let VY be the observed vector of BC-LSF coefficients V_Y ($l_y(1)$, $l_y(2)$, ... , $l_y(q)$), and let V_X be the associated vector of AC-LSF coefficients V_X ($l_x(1)$, $l_x(2)$, ... , $l_x(p)$). We need to predict approximately the best match series of output vector V_X from a series of input vector V_Y. As the characteristic of LSF coefficients, the LSF coefficients in vectors V_X and V_Y have to agree with the following equations:

$$0 < l_x(1) < l_x(2) < \ldots < l_x(p) < \pi, \tag{9}$$

$$0 < l_y(1) < l_y(2) < \ldots < l_y(q) < \pi. \tag{10}$$

LSF differentials have positive values in the range of (0, π). Using LSF differentials can help simplify the problem. Let V_Y' be the observed vector of BC-LSF differences V_Y' ($\delta_y(1)$, $\delta_y(2)$,..., $\delta_y(Q)$) where $\delta_y(1)=l(1)$, $\delta_y(i) = l_y(i) - l_y(i\text{-}1)$, $2 \le i \le Q$, and let V_X' be the predicted vector of AC-LSF differential $V_X'(\delta_x(1), \delta_x(2),..., \delta_x(P))$ where $\delta_x(1)=l_x(1)$, $\delta_x(i)=l_x(i) - l_x(i\text{-}1)$, $2 \le i \le P$. We need to obtain a prediction model *M* that approximate the best match series of output vector V_X' from a series of input vector V_Y' as: $\{V_X'\} \leftarrow M(\{V_Y'\})$.

Since the overlap between every two neighbors in the series of speech frames, their restorations should be related. We propose the application of an Elman – a simple recurrent network (SRN) — to the prediction problem. The function learnt by this network depends on not only the current inputs but also previous states of the network and this should be a good choice for automatically predicting AC-LSF coefficients.

In this paper, we chose k = 1 and set both LP-orders as P = Q = 20. There were 20 nodes for each layer: input layer, output layer and hidden layer. The length of frames is 250 ms, and the overlap of two neighbors is 125 ms. These values are to keep the frame-length short enough, but also reduce the number of training vectors for a small prediction model.

IV. EVALUATION

The aim of this evaluation was to investigate whether the proposed models could adequately restore BC speech to attain better voice-quality and speech intelligibility and whether this could work well blindly. Using both objective and subjective measurements, we evaluated a previous long-term Fourier transform model [2] and the two proposed LP-based models (one is non-blind and the other is blind model). Then, there were two un-blind models: (1) LTF: the long-term Fourier transform and (2) LSF: LP-based models using LSF coefficients and frame basis processing. The proposed blind restoration model were (3) LSF-SRN: LP-based blind restoration - apply SRN to LSF.

A. Objective measurements

We used LSD (log-spectrum distortion), LCD (LP distance), and MCD (MFCC distance) to evaluate methods. These three objective measurements were computed as follows:

$$\mathrm{LSD} = \sqrt{\frac{1}{W}\sum_{\omega}^{W}\left[20\log_{10}\left(\left|S(\omega)\right|\Big/\left|\hat{S}(\omega)\right|\right)\right]^2}\ , \qquad (12)$$

$$\mathrm{LCD} = \sqrt{\frac{1}{P}\cdot\sum_{i=1}^{P}\left(a_x(i)-a_y(i)\right)^2}\ , \qquad (13)$$

$$\mathrm{MCD} = \sum_{i=0}^{12}(c_{x,i}-c_{y,i})^2\ . \qquad (14)$$

where W is the upper frequency (8 kHz in this case), the amplitude spectra getting by 1024-points FFT calculation of 25-ms frames, 15-ms overlapping. $a_x(i)$ and $a_y(i)$ are the i-th LP coefficients of signals with the LP order being set $P=20$, and $c_{x,i}$ and $c_{y,i}$ are the i-th MFCC of the signals. Table II shows the distances between clean AC speech signal and the signals (the observed BC speech and the restored speech signals). In general, the LSF model is the best model for every objective measurement. LSF-SRN is the following model even it blindly restores BC speech.

B. Subjective evaluation

The modified rhyme tests (MRTs) were carried out on English database with six subjects, the JWITs were carried out with forty Japanese subjects, and the VWITs were carried out with fifteen Vietnamese subjects. All the selected subjects have normal hearing.

1) Modified rhyme test (MRT)

The MRTs are carried out on English database. Listeners are shown six-word lists and then asked to identify which of the six is spoken. There are 50 six-word lists of rhyming or similar sounding mono-syllabic English words. Every word is in C-V-C sound sequence, and the six words in each list differ only in the initial or final consonant sound. The result by MRT

TABLE II. OBJECTIVE DISTANCES TO AC SPEECH SIGNALS

Language	Objective Measure	BC	Non-blind		Blind
			LTF	LSF	LSF-SRN
English	LSD	14.28	13.57	**8.92**	*9.64*
	MCD	21.72	18.15	**12.55**	*15.96*
	LCD	3.04	2.91	**1.79**	*2.43*
Japanese	LSD	12.08	11.33	**10.38**	*11.21*
	MCD	20.52	19.37	**17.53**	*19.39*
	LCD	2.79	2.51	**1.83**	*2.58*
Vietnamese	LSD	11.91	11.50	**9.57**	*11.37*
	MCD	26.93	17.24	**14.24**	*20.45*
	LCD	2.22	1.86	**1.28**	*1. 84*

TABLE III. MODIFIED RHYME TEST, CORRECTION (%).

BC	LTF	LSF	LSF-SRN	AC
69.7	76.3	88.0	82.5	95.7

indicate errors in discrimination of both initial and final consonant sounds, and also show the improvement in intelligibility of restored speech [8]. Table III shows the correct score of MRT. As the same as objective measurements, the LSF model gain the best result, following is the LSF-SRN.

2) Japanese word-intelligibility test (JWIT)

In JWITs, the speech signals of eighty Japanese words were played in random order. The Japanese subjects, who did not know these words previously, were requested to listen each word only one time and write down what they got in hiragana. We intend to evaluate the word intelligibility of these signals on 4 familiarity ranges. Since each subject should listen to a word only one time, we divide 40 subjects into five listening group A, B, C, D and F to listen 400 stimuli. Table IV shows us the way to arrange 400 stimuli for five listening group. In generally, the intelligibility could be evaluated by the average recognition accuracy which is scored by all subjects. Table VI lists the recognition accuracy scores of JWITs.

3) Vietnamese word-intelligibility test (VWIT)

In VWITs, the speech signals of forty words were played in random order. The Vietnamese subjects, were requested to listen each word only one time and write down what they got. We intend to evaluate the word intelligibility of these signals on 4 different categories C1-C4, with different types of words (mono or duo syllabic) and levels of frequency indexes (low or high frequency index). As the same as in JWITs, we divide 15 subjects into five listening groups A', B', C', D' and F' to listen 200 stimuli. Table V shows us the way to arrange 200 stimuli for five listening groups. Table VII lists the score results of VWITs.

C. Discussion

By evaluation results in Table II, III, VI and VII, the non-blind LP-based model LSF and the blind LP-based model LSF-SRN showed the significant ability to restore BC speech, both intelligibility (LSD, MRT, JWIT and VWIT) and spectral distance (LCD, MCD).

As JWIT scores, LSF model improved 36.5% of average recognition accuracy, LSF-SRN model follows with an expressed result 20% greater. It is the same result with VWIT, whereas LTF model improved 30% of average recognition

TABLE IV. JAPANESE WORD INTELLIGIBILITY STIMULI

Word Index		BC	LTF	LSF	LSF SRN	AC
R1 (1.0–2.5) Low familiarity	1–4	A	B	C	D	E
	5–8	E	A	B	C	D
	9–12	D	E	A	B	C
	13–16	C	D	E	A	B
	17–20	B	C	D	E	A
R2 (2.5–4.5) Medium Low familiarity	21–24	A	B	C	D	E
	25–28	E	A	B	C	D
	29–32	D	E	A	B	C
	33–36	C	D	E	A	B
	37–40	B	C	D	E	A
R3 (4.5–5.5) Medium High familiarity	41–44	A	B	C	D	E
	45–48	E	A	B	C	D
	49–52	D	E	A	B	C
	53–56	C	D	E	A	B
	57–60	B	C	D	E	A
R4 (5.5–7.0) High familiarity	61–64	A	B	C	D	E
	65–68	E	A	B	C	D
	69–72	D	E	A	B	C
	73–76	C	D	E	A	B
	77–80	B	C	D	E	A

TABLE V. VIETNAMESE WORD INTELLIGIBILITY STIMULI

Word Index		BC	LTF	LSF	LSF SRN	AC
C1 Mono syl. Low freq.	1–2	A’	B’	C’	D’	E’
	3–4	E’	A’	B’	C’	D’
	5–6	D’	E’	A’	B’	C’
	7–8	C’	D’	E’	A’	B’
	9–10	B’	C’	D’	E’	A’
C2 Mono syl. High freq.	11–12	A’	B’	C’	D’	E’
	13–14	E’	A’	B’	C’	D’
	15–16	D’	E’	A’	B’	C’
	17–18	C’	D’	E’	A’	B’
	19–20	B’	C’	D’	E’	A’
C3 Duo syl. Low freq.	21–22	A’	B’	C’	D’	E’
	23–24	E’	A’	B’	C’	D’
	25–26	D’	E’	A’	B’	C’
	27–28	C’	D’	E’	A’	B’
	29–30	B’	C’	D’	E’	A’
C4 Duo syl. High freq.	31–32	A’	B’	C’	D’	E’
	33–34	E’	A’	B’	C’	D’
	35–36	D’	E’	A’	B’	C’
	37–38	C’	D’	E’	A’	B’
	39–40	B’	C’	D’	E’	A’

TABLE VI. RESULTS OF JAPANESE-WORD INTELLIGIBILITY TESTS (%)

Familiarity	BC	LTF	LSF	LSF-SRN	AC
R1 (1.0–2.5)	3.5	3.5	26.0	14.5	66.0
R2 (2.5–4.5)	3.0	3.0	37.0	19.0	63.0
R2 (4.5–5.5)	13.0	21.0	58.0	43.0	71.5
R4 (5.5–7.0)	20.5	36.0	64.5	43.5	77.5
Avg (1.0-7.0)	10.0	15.9	46.4	30.0	69.5

TABLE VII. RESULTS OF VIETNAMESE-WORD INTELLIGIBILITY TESTS (%)

Category	BC	LTF	LSF	LSF SRN	AC
C1 (mono, low)	43.3	63.3	73.3	50.0	76.7
C2 (mono, high)	60.0	70.0	90.0	80.0	90.0
C3 (duo, low)	60.0	53.3	90.0	70.0	100.0
C4 (duo, high)	50.0	40.0	70.0	56.7	100.0
Avg	53.3	56.7	80.8	64.2	91.7

accuracy and, LSF-SRN model follows with a result 11% greater. In general, we found that, it is more difficult to restore BC speech at low familiarity. At ranges R1 and R2, LTF model even gain no improvement. The improvement result increases quickly with the higher familiarity. LSF model even improve the average recognition accuracy with 40% greater in high familiarity ranges R3 and R4. At these familiarity ranges, LSF-SRN improved the BC speech up to almost the same scores 43% as LSF. We also found that, Even it is a blind model, LSF-SRN showed expressed ability to improve the voice quality and intelligibility of BC speech signal. The intelligibility improvement seems to be independent of languages with expressed results for English (MRTs), Japanese (JWITs) and also Vietnamese (VWITs). It means the SRN was trained adequately for predicting AC-LSF coefficients and help LSF-SRN model achieve good restoration.

V. CONCLUSION

We improved the LP-based model by (1) extending long-term to frame-based processing, (2) using LSF coefficients on LP representations, and (3) predicting LSF parameters on a frame basis via an Elman network. We entirely evaluated the developed model in comparison with previous models to find out whether the developed model can adequately improve voice quality and the intelligibility of BC speech, using three objective measures and three subjective tests.

The results of experiments showed that the LP-based model proved the significant practicability of blind-BC restoration. Especially, the model can be applied to improve the intelligibility of BC speech on different languages.

The next challenge is to improve the spectral distances since the blind restoration model we proposed is still limited in this regard. Significant improvements in both intelligibility and spectral distances remain problems that need to be solved to construct a blind restoration model that will be as good as the LSF model. The factor k is currently assumed as constant in all LP-based models and can be even further improved by considering its change. We should consider the problem of Lombard effect in future work. The same methodology concepts of LP-based models can be applied. Noisy BC speech can be regarded as original BC speech but the SRN of blind LP-based models should be re-trained.

ACKNOWLEDGMENTS

This work was supported by the YAZAKI Memorial Foundation for Science and Technology and a scheme for the ``21st Century COE Program" in Special Coordination Funds for the Promotion of Science and Technology made available by the Ministry of Education, Culture, Sports, Science, and Technology in Japan. This was also partially supported by a Grant Program by the SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

REFERENCES

[1] Kitamori, S. and Takizawa, M. “An Analysis of Bone Conducted Speech Signal by Articulation Tests,” IEICE Trans. J72-A (11), 1764–1771, Nov. 1989.

[2] Tamiya, T. and Shimamura, T. “Reconstruct Filter Design for Bone-Conducted Speech,” Proc. ICSLP2004, II, 1085–1088, Oct. 2004.

[3] Thang, V. T., Kimura, K., Unoki, M., and Akagi, M. “A study on restoration of bone-conducted speech with MTFbased and LP-based model,” J. Signal Processing, 10(6), 407–417, Nov. 2006.

[4] Thang, V. T., Unoki, M., and Akagi, M. “A study on an LP-based model for restoring bone-conducted speech,” Proc. ICCE’2006, 294–299, Hanoi, Vietnam, Oct. 2006.

[5] Thang, V. T., Seide, G., Unoki, M., and Akagi, M., “Method of LP-based blind restoration for improving intelligibility of bone-conducted speech,” Proc. Interspeech2007, 966-969, Antwerp, Belgium, August 2007.

[6] Nakagiri, M., Toda, T., Kashioka, H., and Shikano, K. “Improving Body Transmitted Unvoiced Speech with Statistical Voice Conversion,” Proc. ICSLP2006, 2270–2273, Sept. 2006.

[7] Subramanya, A., Zhang, Z., Liu, Z., Droppo, J., and Acero, A. “A Graphical Model for Multi-Sensory Speech Processing in Airb-and-Bone Conductive Microphones,” Proc. Eurospeech2005, 2361-2364, Lisbon, Portugal, Sept. 2005.

[8] Brungart, D. S. “Evaluation of speech intelligibility with the coordinate response measure,” J. Acoust. Soc. Am., 109(5), 2276–2279, May, 2001.