Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 49 (2007) 687–696

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Speech emission control using active cancellation

Kazuhiro Kondo *, Kiyoshi Nakagawa

Department of Electrical Engineering, Faculty of Engineering, Yamagata University, 4-3-16 Jonan, Yonezawa, Yamagata 992-8510, Japan

Received 15 April 2006; received in revised form 8 September 2006; accepted 11 April 2007

Abstract

We investigated on the possibility of an active cancellation system for unnecessary speech radiation control. Some examples of the intended application of this system are cellular speech cancellation and speech input for recognition-based dictation systems. Both of these applications do not require speech to be radiated into surrounding space, but only into the input microphone, and would benefit if global radiation is controlled. We first show that speech cancellation is possible with a secondary source placed in proximity to the mouth generating linear-predicted phase-inverted speech. However, the prediction must also cover the long delay associated with the acoustic to/from electric conversion, as well as A/D, D/A conversions, and all associated processing, which we found could go up to as long as 3 ms.

By using LPC predicted samples recursively to predict further samples, we found that prediction with SNR of about 6 dB is possible, even with this long delay. The prediction coefficient update is suppressed during this recursion. Lowering the sampling frequency in order to lower the number of predicted samples at the cost of reduced bandwidth further enhances prediction accuracy. At a sampling frequency of 8 kHz, speech emission control of about 7 dB for female speech and 4 dB for male speech was found to be possible.

Finally, we experimentally evaluated the proposed active speech control method. Predicted samples of recorded speech was first prepared off line. We then actually played out both the original and the predicted samples simultaneously from two loud speakers. It was found that (1) speech cancellation of up to about 10 dB is possible, but is highly speaker dependent, (2) secondary loud speaker should be oriented in the same direction as the primary source, *i.e.*, the mouth. We plan to investigate further to improve prediction accuracy using prediction coefficient extrapolation. A prototype system implementation using DSPs is also planned.

PACS: 43.50.Ki; 43.72.−p

Keywords: Cellular speech; Active control; Linear prediction

1. Introduction

Cellular phones have become quite ubiquitous in most developed countries. This situation has created new types of problems: we are often bombarded by speech from people chatting away on their cell phones from all directions. This speech clearly is not intended for us, only to the people on the receiving end of the call, and thus is useless after a small portion of it enters the microphones in the handsets. It also creates privacy concerns. Thus it would be beneficial if we could control the radiation of this speech into the surrounding space, or at least if we can mitigate it to some degree.

On the other hand, speech recognition systems have vastly improved these few years. Speech dictation systems with acceptable accuracy have been released, and there is a growing population of regular users. Many of these users will be using these systems in offices, which potentially will have many other users of similar systems, perhaps in neighbouring low partitioned spaces. The speech from the surrounding users would obviously become noise to the dictation system, most likely bringing down the recognition accuracy considerably. It would also undoubtedly create a very user-unfriendly working environment. Thus, it would also be beneficial for this application if we can control the

* Corresponding author. Tel./fax: +81 238 26 3312.
E-mail address: kkondo@yz.yamagata-u.ac.jp (K. Kondo).

0167-6393/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2007.04.010

unnecessary radiation of speech into the surrounding environment.

Active Noise Cancellation (Nelson and Elliott, 1992; Kuo and Morgan, 1996; Elliot, 2001) has received great interest this decade with the advancement of Digital Signal Processors. There have been some successful applications of this technology (Tichy, 2001), *e.g.* control of fan noise radiation through ducts, jet engine radiation control, road noise control in automobiles (Sano et al., 2001) to name just a few. In this paper, we investigated on the possibility of applying similar techniques to the control of speech (Kondo and Nakagawa, 2002). We show that speech cancellation is possible with a secondary source placed in proximity to the mouth generating linearly predicted phase-inverted speech. However, the prediction must also cover the long delay associated with the acoustic to/from electric conversion, as well as A/D, D/A conversions, and all associated processing, which we found could go up to a few milliseconds. We investigated on the possibility of using linear prediction recursively to predict speech covering this long delay. Use of a lower sampling rate to limit the number of samples which need to be predicted ahead at the cost of limiting the effective bandwidth has also been investigated.

We also conducted experiments to further investigate the feasibility of our proposed method (Kondo and Nakagawa, 2005). Linear-predicted phase-inverted speech samples of pre-recorded speech were prepared beforehand. Speech prediction was accomplished using recursive LPC as described in (Kondo and Nakagawa, 2003). We also added an alternative prediction method based on pitch estimation and sample repetition for comparison. This is essentially a forward speech estimation method described in the ITU standard G.711 Appendix I (ITU-T Recommendation G.711 Appendix I, 1999). The purpose of this method was to estimate speech segments lost due to packet loss using previously received speech. Both the pre-recorded speech and the predicted speech were played out from loud speakers placed close to each other. We then measured the speech cancellation level at surrounding positions using a sound level meter.

In the Section 2, the proposed speech cancellation scheme is described. Next, some computer simulations of the proposed scheme is described, followed by a description of the long term linear prediction and its evaluation results. Section 5 describes the setup and results of the experimental evaluations. Finally, the summary and some discussions are given.

2. Active speech cancellation

As stated in the introduction, we will attempt to cancel speech by simply placing a secondary sound source very near the primary sound source, *i.e.* the mouth. Fig. 1 shows the basic arrangement of our assumed system. Here, d, r_p, and r_s are the distance between the primary and the secondary source, the primary source and the observation point, and the secondary source and the observation point respectively. We will assume the following for all our simulations:

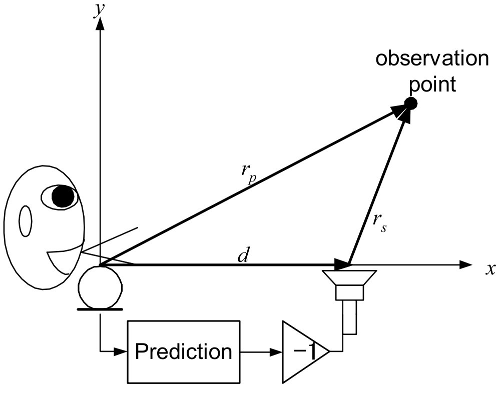

Fig. 1. Active speech control configuration.

(1) All sound sources are simple point sources, radiating sound pressure equally in all directions.
(2) Sound pressure propagates linearly, and can be estimated at any observation point as inversely proportional to the distance from the source, while the transmission delay is simply proportional to the distance.
(3) For now, we will assume that we can obtain pure speech input without contamination from the secondary source at the microphone input.
(4) We assume no "acoustic coupling" between the sources. In other words, the acoustic pressure from the secondary source does not effect pressure radiation from the primary source and vice versa.

One can argue that assumption (1) is too optimistic. However, Flanagan found that simple spherical source model is accurate to within 3 dB within a solid angle of π steradians around the mouth axis for frequencies below 4000 Hz (Flanagan, 1960). We believe this is accurate enough for our purpose for now.

Also for assumption (3), the secondary source would contaminate speech input, and thereby alter the "wanted" speech quality, as well as degrade the cancellation level. However, we believe we should be able to use an alternative speech input to avoid contamination. Bone-conduction vibration pick-up "ear-insert" microphones (Ono, 1977; Black, 1957) are good candidates. These microphones pick-up the internal vibrations and thus are not as affected by external secondary source contamination.

If we generate a sound wave at the origin with pressure $P_0(t)$ (*i.e.* speech), we generate a replica of $P_0(t)$, phase-inverted, at the secondary source, *i.e.*,

$$P_s(t) = -\widehat{P}_0(t) \tag{1}$$

From the above assumptions, at observation point (x, y), we have the sum of P_0 and P_s after it has travelled distances of r_p and r_s respectively, *i.e.*,

$$P_{x,y} = \frac{P_0\left(t - \frac{r_p}{c}\right)}{r_p} + \frac{P_s\left(t - \frac{r_s}{c}\right)}{r_s} \approx \frac{P_0\left(t - \frac{r_p}{c}\right) - \widehat{P}_0\left(t - \frac{r_s}{c}\right)}{r_p} \qquad (2)$$

where c is the sound velocity.

In order to cancel the primary speech, we need to estimate a good replica of the primary speech, and generate the inverted version from the secondary source simultaneously. Thus, we need to predict the next sample from the past primary speech samples. We employed simple linear prediction for this purpose:

$$\hat{x}_n = -1 * \sum_{i=1}^{N} a_{i+1} x_{n-i} \qquad (3)$$

where $\hat{x}_n$ is the predicted sample, x_i, and a_i are input speech samples and the corresponding LPC coefficients respectively.

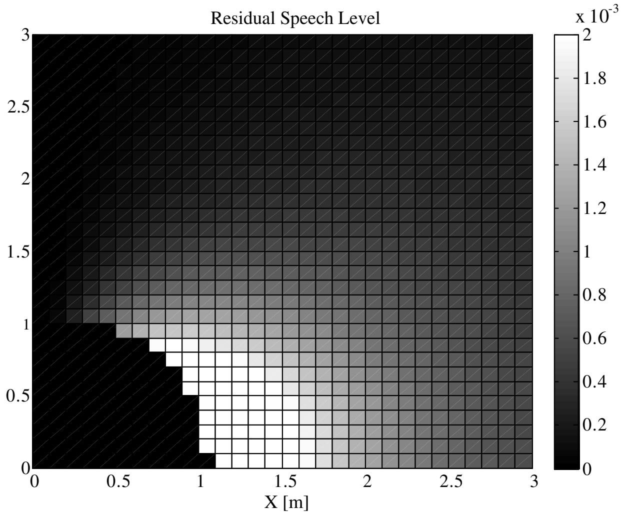

Fig. 3. Speech residual level with cancellation.

3. Simulations

Fig. 2 shows the simulated speech radiation level from the primary source alone, calculated in the 3 × 3 m square observation plane. The level was estimated from the square mean over the whole utterance. The speech sample here is female speech from the SpEAR database (Wan et al.) with the speech "Biblical scholars argue history", approximately 2.4 s in length, sampled at 16 kHz with 16 bits.

Fig. 3 shows the residual speech level with the secondary source at 2 cm left of the primary source on the x-axis, *i.e.* at coordinate (0.02 m, 0 m). A 128th order LPC was used to predict the samples to be played out from the secondary source. LPC coefficients were recalculated for every new sample using the Yule–Walker equation (Haykin, 1996). A block length of 256 samples was used. No windows were used on these samples. Estimates within the circle radius of 1 m within the primary source were not calculated. Notice that the residual level is significant along the x-axis. This is because the difference between the primary and the secondary source tend to be largest along the x-axis, thus yielding largest transmission delay difference between the sources.

Fig. 4 shows the cancellation level, *i.e.* the ratio of the residual level shown in Fig. 3 to the uncontrolled level shown in Fig. 2. Large negative values show larger cancellation of speech radiation. The cancellation is smaller along the x-axis, with cancellation of about −7.7 dB, while it largest along the y-axis, with about −14.1 dB, as discussed previously.

Fig. 5 shows the effect of LPC order on the speech cancellation level. Basically, the block length was twice the length of the prediction order. Again, no windowing was applied. The values are for an observation point at (3 m, 3 m). The order dependency was calculated for four speech samples with approximately the same length, ranging from 2.4 s to 3.5 s, all sampled at 16 kHz with 16 bits. Again, all samples were taken from the SpEAR database (Wan et al.).

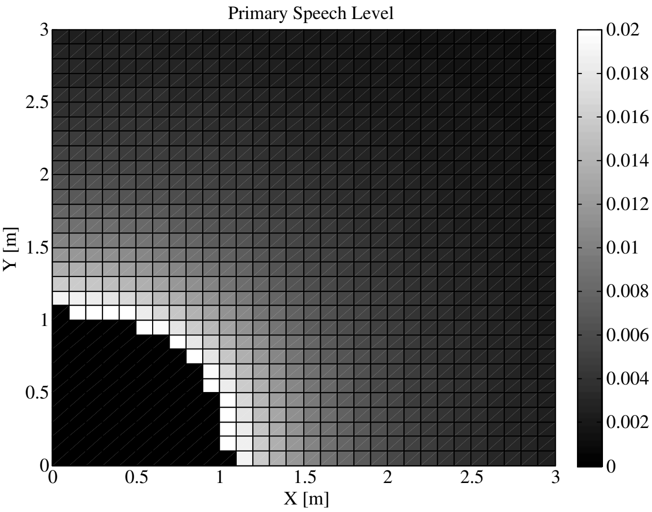

Fig. 2. Speech radiation from primary source.

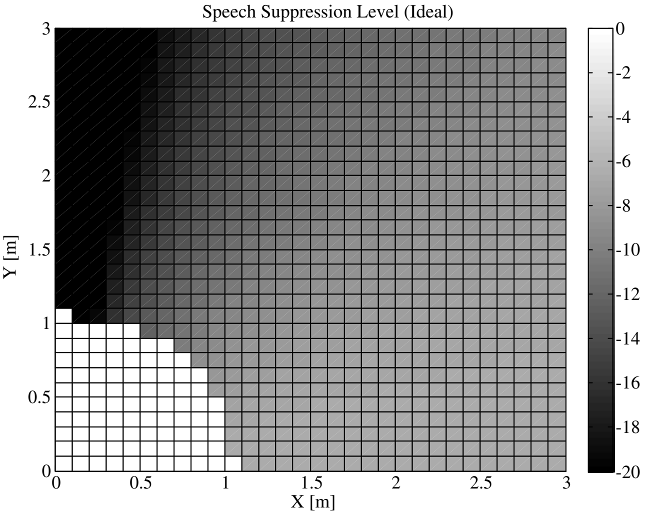

Fig. 4. Speech cancellation.

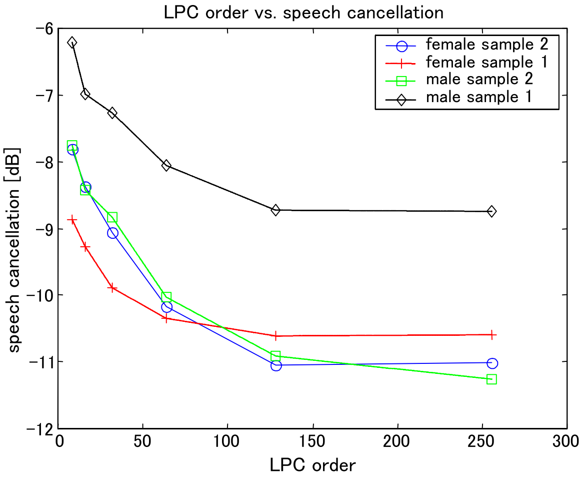

Fig. 5. LPC order vs. speech cancellation.

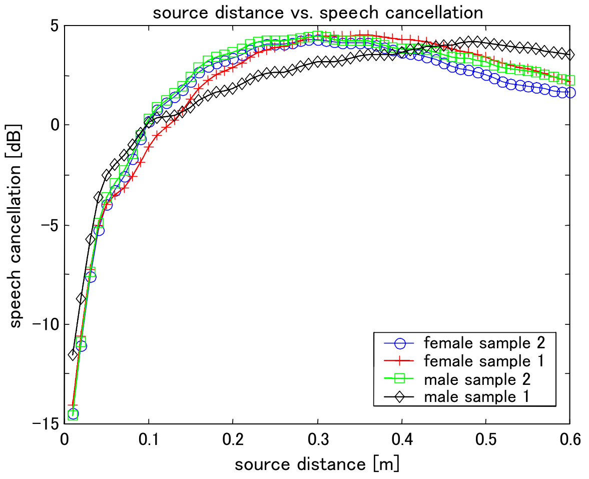

Fig. 6. Primary to secondary source distance vs. speech cancellation.

All samples show similar trends, with the cancellation leveling off at LPC order of 128. One male speech sample shows significantly smaller cancellation saturation level than other samples. This seems to be closely related to the LPC prediction gain.

Table 1 lists the LPC prediction gain for samples used in Fig. 5. The male speech sample with low speech cancellation shows relatively low LPC prediction gain, consequently giving less accurate speech sample estimates.

Fig. 6 shows the effect of primary to secondary source distance d on the speech cancellation. All speech samples from the secondary sources were predicted with 128th order LPC. The cancellation was calculated at coordinate (3 m, 3 m) again.

As can be seen, speech cancellation quickly reduces as the source distance increases, and reaches about even at about 0.1 m. Above this distance, the second source merely adds to the primary source level, thereby “amplifying” the radiation level.

If we assume that the bulk of speech is around 1000 Hz, its wave length λ around 0.3 m, then as stated in (Nelson and Elliott, 1992), the primary to secondary source distance to have effective cancellation, about $\lambda/8$, should be below approximately 0.04 m. This is roughly in line with our observation in Fig. 6.

Table 2 compares the speech cancellation level with LPC predicted speech with orders of 64 and 128 with “ideally” predicted speech, *i.e.* when perfect replica is used (with phase inversion). The speech cancellation especially with LPC order of 128 is very similar to the “ideal” case, as is shown.

Table 1
LPC prediction gain for four speech samples

Speech sample	LPC prediction gain [dB]
Male speech 1	11.61
Male speech 2	15.28
Female speech 1	14.19
Female speech 2	14.60

Table 2
Speech cancellation with various secondary sources

Inter-source distance d (m)	Speech cancellation (dB)		
	Ideal	LPC (64th)	LPC (128th)
0.05	−1.75	−2.21	−2.50
0.02	−9.52	−8.06	−8.72
0.01	−14.83	−10.74	−11.52

4. Long term prediction

4.1. Long term prediction using linear prediction

So far, we have shown that it is possible to cancel speech radiation by placing a secondary speaker close to the mouth, and playing a predicted and phase-inverted speech from the speaker. However, the predicted sample also must incorporate the delay associated with the acoustic to/from electric conversion, as well as A/D, D/A conversions, and all associated digital signal processing. We found this delay could go up to around 3 ms. It is necessary to “predict ahead” of this delay in order to generate a good replica of the sound to be generated from the primary source (“the mouth”) simultaneously from the secondary source.

As we have stated in Section 4.3, we have used conventional linear prediction to predict speech samples. Here, we shall use linear prediction recursively to successively obtain speech samples ahead in time. In other words, we can obtain a speech sample one sampling interval ahead, *i.e.* $\hat{x}_n$ using previously observed samples x_i, where $i = n - 1, n - 2, \ldots, n - N$. Note that the prediction coefficients, a_i, $i = 2, 3, \ldots N + 1$ has been calculated from x_i, $i = n - 1, n - 2, \ldots, n - N$ using the Yule–Walker equation (Haykin, 1996). We will use the same prediction coefficient to predict $\hat{x}_{n+1}$ from $\hat{x}_n$ and x_i, $i = n - 1, n - 2, \ldots, n - N + 1$. $\hat{x}_{n+2}, \hat{x}_{n+3}, \ldots$ can be predicted in a similar manner.

4.2. Long term prediction using pitch repetition

We also tested a well known prediction scheme included in the ITU Recommendation G.711 Annex I (ITU-T Recommendation G.711 Appendix I, 1999) for long term prediction. The described method is used to predict speech segments lost during packet transmission. A number of recent received speech samples are retained in a buffer. To predict samples ahead in time, pitch is estimated by finding the peak in the normalized cross-correlation function of the most recent samples in the pitch buffer. In order to predict n samples ahead, we simply extract the $mod(n,p)$th sample in the last pitch period in the buffer, where $mod()$ is the modulus after division, and p is the estimated pitch period. This is illustrated in Fig. 7.

This pitch-based method does not provide excellent accuracy, but the accuracy remains fairly constant with the increase in the number of predicted samples ahead. The accuracy is speaker dependent to some degree.

4.3. Prediction accuracy using long term prediction

Fig. 8 shows the SNR of predicted speech samples using recursive LPC (denoted "LPC") and the pitch buffer repetition (denoted "pitch"). As stated previously, "LPC" shows generally higher SNRs than "pitch" at smaller predicted time ahead (PTA), but this decreases rapidly as the PTA increases. "Pitch" shows relatively constant SNR, with gradual decrease as PTA increases, eventually showing higher SNR than "LPC". Both "pitch" and "LPC" methods show fairly high speaker dependency.

4.4. Speech cancellation using long term prediction

We now will attempt to simulate the speech cancellation possible using these long prediction schemes. The following is assumed for all simulations:

(1) All sound sources are simple point sources, radiating sound pressure equally in all directions.
(2) Sound pressure propagates linearly, and can be estimated at any observation point as inversely proportional to the distance from the source, while the transmission delay is simply proportional to the distance.
(3) For now, we will assume that we can obtain pure speech input without contamination from the secondary source at the microphone input.

From the sampled speech, linear prediction is used to obtain speech samples 3 ms ahead, phase-inverted, and is played out from the secondary source. Both the primary speech and the predicted sample from the secondary source will travel within the space to an arbitrary point surrounding both sources, and is summed, thereby cancelling each other. Since speech spectrum predominantly occupies low frequency ranges below 1 kHz, it may be possible to lower the sampling rate, allowing less samples to be predicted ahead in order to cover the long delay associated with the microphone-to-secondary source path. For example, if we halve the sampling rate, the number of samples required to cover the delay is also halved. This obviously comes at a cost of narrower operating bandwidth of the cancelling speech from the secondary source. Table 3 shows the cancellation level at an observation coordinate (3 m, 3 m), as well as the maximum cancellation within the 3 m × 3 m plane. The primary source (*i.e.* the mouth) is assumed to be at the origin, and the secondary source is placed at coordinate (0.02 m, 0 m). All other conditions are the same as in the previous section.

As expected, lowering the sampling rate does increase the cancellation level, especially for the male sample. This is expected since the male sample will contain most of the components in the lower frequency range. The cancellation level reaches maximum at 8 kHz sampling, which is a nice integer down sample rate from the original sample rate of 16 kHz. The reason for this may be that the distortion is minimum in the down sampling process to 8 kHz compared to other frequencies, which require a more complex sample rate conversion. Fig. 9 shows the cancellation level, defined as the log ratio between unsuppressed vs. suppressed speech power level, where the larger the cancellation, the more negative the cancellation level. The cancellation level is shown within a 3 m × 3 m plane from

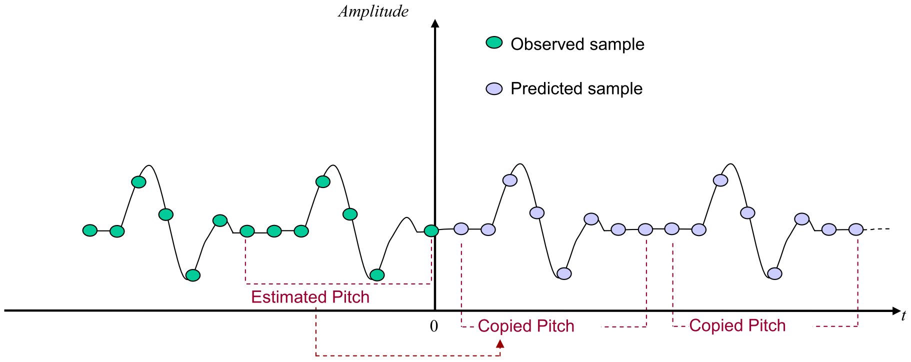

Fig. 7. The pitch repetition scheme.

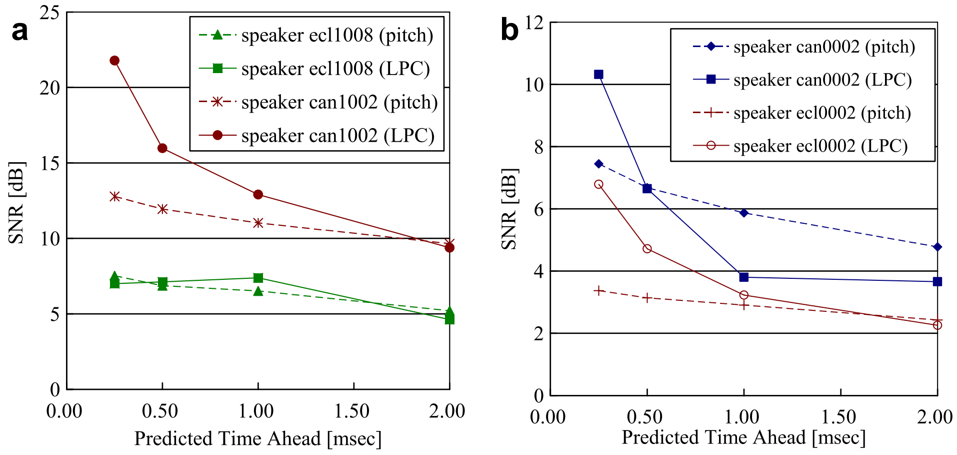

Fig. 8. SNR of predicted speech: (a) female speech and (b) male speech.

Table 3
Speech cancellation vs. sample rate

Sample	Sample rate (kHz)	Predicted samples ahead	Speech cancellation (dB)	
			At Coord. (3 m, 3 m)	Max in plane
Female	16	50	−6.32	−6.39
	10	32	−6.27	−6.33
	8	25	−6.44	−6.55
	4	13	−6.42	−6.52
Male	16	50	−3.38	−3.75
	10	32	−3.42	−3.73
	8	25	−3.98	−4.00
	4	13	−3.86	−3.89

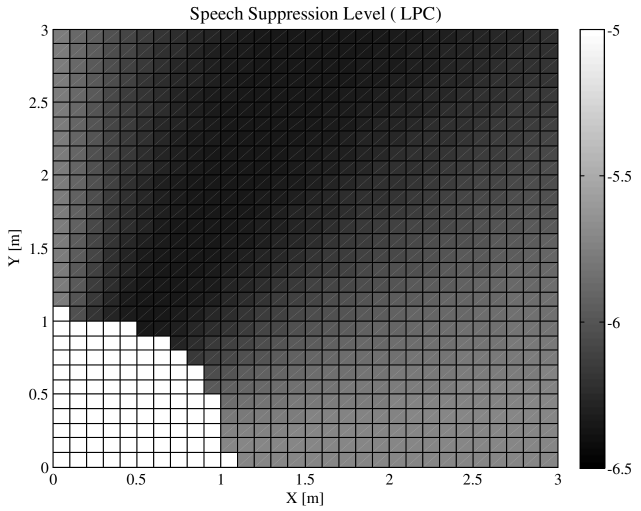

Fig. 9. Speech cancellation level within a 3 m × 3 m plane at 8 kHz sampling.

the primary source for the female sample at 8 kHz sampling. There is apparently a "lobe" with relatively large cancellation in the diagonal direction. The maximum cancellation within this "lobe" was about −6.6 dB. The male sample showed similar diagonal "lobe" albeit at a somewhat lower cancellation level. It was also found that it is possible to "steer" this lobe by introducing a small variable delay in the prediction path.

5. Experimental evaluation

5.1. Setup

We actually tried to generate speech signals and its predicted, phase-inverted signal from two loud speakers placed near each other simultaneously, and measured the amount of cancellation possible. Since the prediction is fairly computationally demanding and difficult to accomplish in real-time with conventional computers, we prepared predicted speech samples beforehand. The speech signal and the phase-inverted predicted samples were played out from two identical loud speakers simultaneously. The loud speakers were 8 cm full range speakers in box enclosures, and were mounted on boom stands using ball heads for camera mounts.

We tested two loud speaker orientations as shown in Fig. 10. One orientation (A), shown in Fig. 10a, was with loud speakers facing the same direction. The physical dimension of the speakers and its enclosures limits the distance between the loud speaker centers, denoted d in the figure, to 12 cm. With the other orientation, as shown in Fig. 10b, where the loud speakers face each other, there is no such limit, and we tested distances of $d = 2$ cm and $d = 10$ cm. Fig. 11 shows photos of the above settings. The sound pressure level was measured by averaging the peak within an utterance measured with a sound level meter (Ono Sokki LA-5111) with flat frequency weighting. Five peak measurements were averaged. All loud speakers and the sound meter were positioned 1 m above the floor. As shown in Fig. 10, the primary speaker which played out the speech signal was placed on the origin, while the observation points (the sound level meter) were placed 3 m from the origin at angles of 0°, 45°, and 90°, respectively. The secondary loud speaker, which generated the

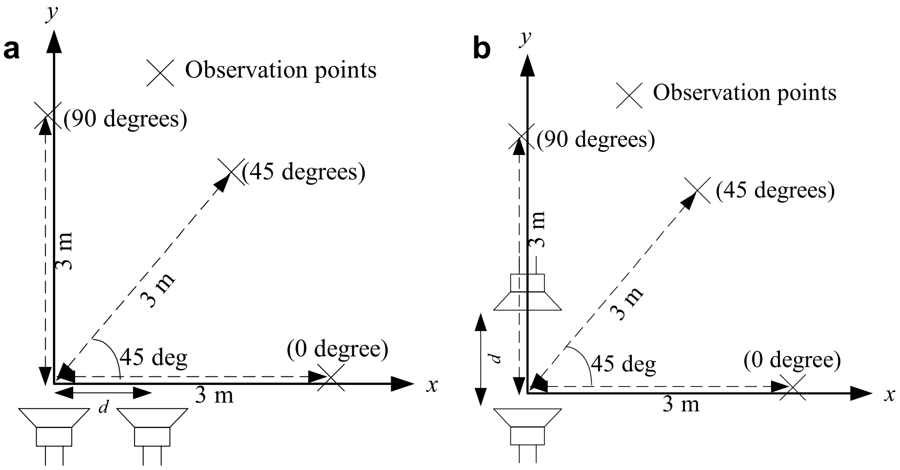

Fig. 10. Loud speaker orientations: (a) orientation A and (b) orientation B.

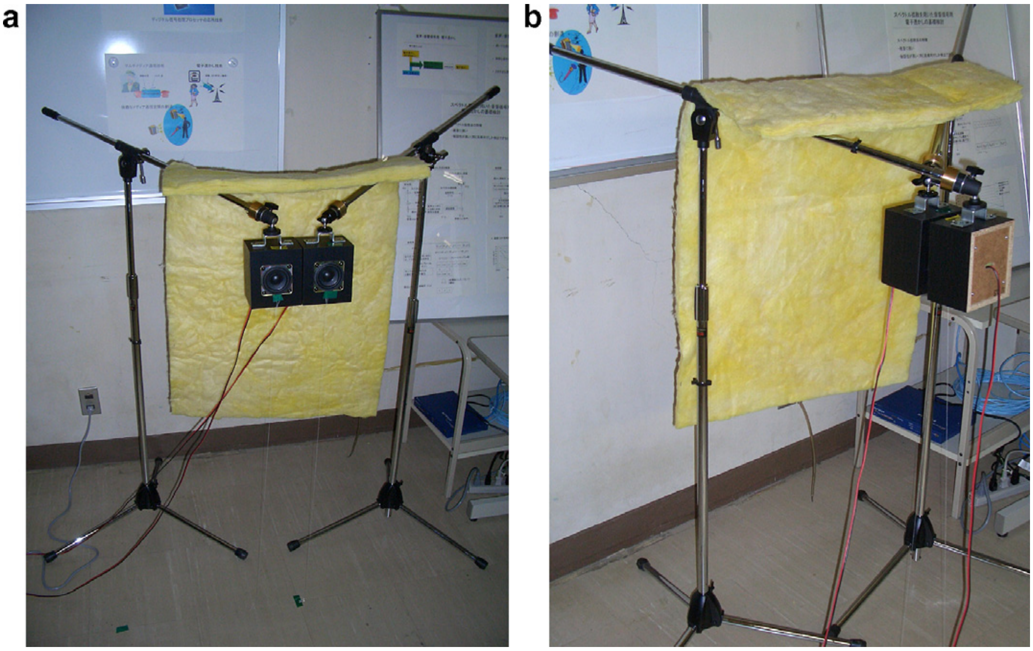

Fig. 11. Photos of loud speaker orientations: (a) orientation A and (b) orientation B.

phase-inverted predicted speech, was placed either on the *x*-axis (0°) in the orientation shown in Fig. 10a, or placed on the *y*-axis (90°) as in Fig. 10b.

We measured the round trip delay in the electric-acoustic to electric-acoustic loop, and found it to range from 180 to 260 μs depending on the loud speaker and microphone used. With the A/D and the D/A conversion, the delay varied widely depending on the implementation, from 750 μs to over 3 ms. However, with optimum design, it should be possible to bring this delay to close to the bare analog loop delay described above. Accordingly, we prepared speech samples predicted from 250 μs to 2 ms ahead to cover this delay.

For speech samples, we used read Japanese sentences from the ASJ Speech Corpus (Information Processing Development Corporation, 1991) down-sampled to 8 kHz. We randomly chose two male and two female speakers reading the same short sentence.

5.2. Results

Fig. 12 shows the speech cancellation level for loud speaker orientation shown in Fig. 10a, or orientation A. Speech cancellation was calculated as the ratio of the average sound pressure level with speech from the both the primary and the secondary source (the phase-inverted predicted speech) to the sound level without the secondary source. Again, more negative values show larger cancellation of speech radiation. The distance *d* between the primary and secondary loud speakers was set to 12 cm. The sound level meter was placed on the *x*-axis, 3 m from the origin.

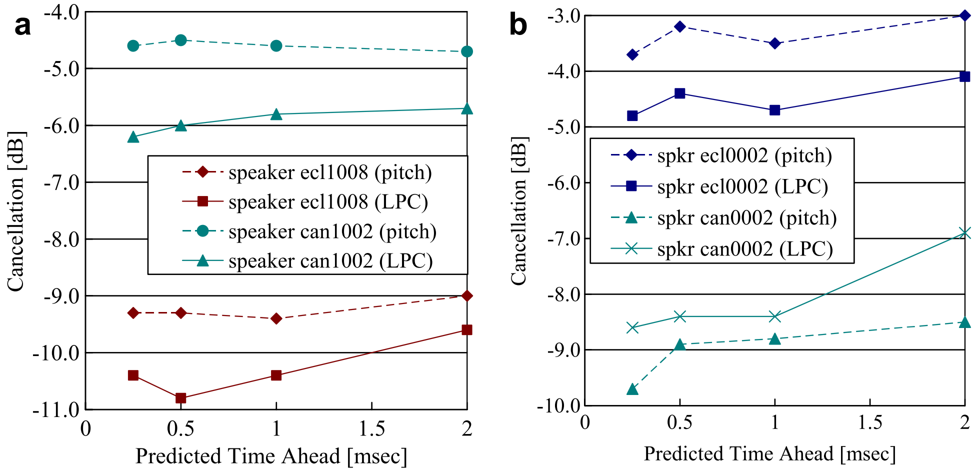

Fig. 12. Cancellation level for orientation A at 0°: (a) female speech and (b) male speech.

Speech cancellation of over 10 dB (nominal values of less than −10 dB) was possible for some speakers, while it was as low as 3 dB for others. Generally, the prediction with recursive LPC outperforms pitch repetition. Speakers with high prediction accuracy do not necessarily show high levels of cancellation. Surprisingly, the predicted time ahead (PTA) did not have significant effect on the cancellation level.

Fig. 13 shows the speech cancellation level at observation positions 0°, 45°, and 90° from the x-axis, all within a radius of 3 m. For speaker ecl1008, observation at 45° showed clearly lower cancellation level than other angles. On the other hand, for speaker ecl0002, observation at 90° showed lowest cancellation. However, for both speakers, observation on the x-axis (0°) showed the best cancellation overall for this loud speaker orientation (A).

Fig. 14 compares the cancellation level with both loud speaker orientations in Fig. 10. As stated before, for the orientation shown in Fig. 10a (orientation A), the physical size limits the inter-loudspeaker distance d to 12 cm. For orientation in Fig. 10b (orientation B), we tested $d = 2$ cm and 10 cm. All measurements were on female speaker ecl1008 with observation on the x-axis at 3 m from the origin.

Overall, orientation A shows higher cancellation than orientation B, even though the inter-speaker distance d was larger. Surprisingly, for orientation B, the cancellation was greater with a larger d of 10 cm. This is contrary to our previous simulation results stated in Section 3, and needs further investigation.

Finally, Fig. 15 shows the power spectrum of the original speech signal, and the residual signal using pitch repetition and recursive LPC prediction. We also included residual signals with “ideal” prediction, where the original speech is simply phase-inverted and played out from the secondary source. This refers to the ideal case where perfect prediction was possible, and shows the upper bound of the proposed method.

As shown in the figure, the “ideal” case shows constant cancellation over all of the bandwidth. The “pitch” and the

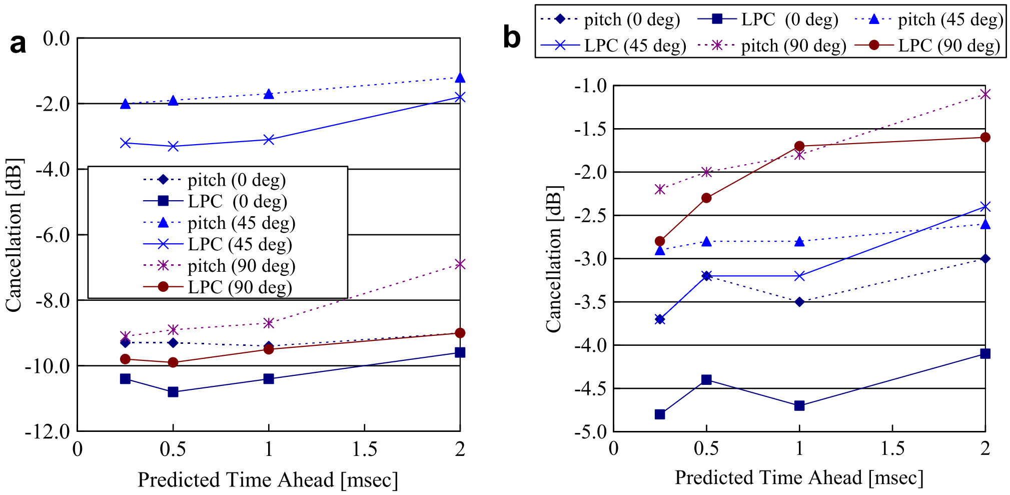

Fig. 13. Cancellation for orientation A at various angles: (a) female speaker ecl1008 and (b) male speaker ecl0002.

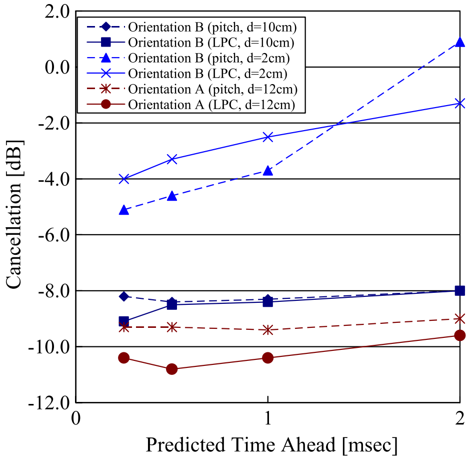

Fig. 14. Speech cancellation vs. loud speaker orientation.

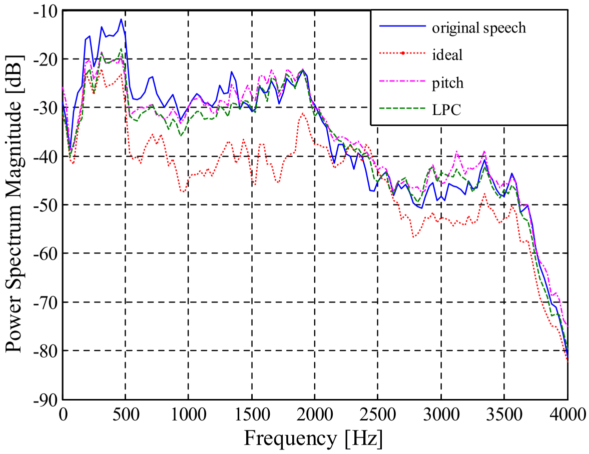

Fig. 15. Power spectrum of residual signal.

“LPC” methods show some cancellation in lower frequencies below 1000 Hz. The “pitch” shows higher level than the original speech, *i.e.* additional noise, in the 2800–3300 Hz range, which was perceived as subjectively annoying high frequency “hissing”.

6. Summary and discussion

We evaluated an active speech control scheme which reduces unnecessary speech radiated into the surrounding space. The proposed method reduces speech by generating phase-inverted predicted speech from a secondary loud speaker. Speech was predicted using LPC recursively to predict samples ahead of the associated processing delay. An alternative method of prediction using pitch estimation and pitch interval repetition was included in this study.

Through simulations, we found that with the proposed method, prediction with SNR of about 6 dB is possible, even with the long delay associated in the path from the speech input microphone to the secondary source output. The prediction coefficient update should be suppressed during this recursion. Lowering the sampling frequency in order to lower the number of predicted samples at the cost of reduced bandwidth was found to further enhance the prediction accuracy. At a sampling frequency of 8 kHz, simulations showed speech emission control of about 7 dB for female speech and 4 dB for male speech.

To evaluate the proposed method experimentally, samples of recorded speech were prepared off line. Then, both the original and the predicted phase-inverted sample were actually played out simultaneously from two loud speakers. The following were the main conclusions and observations of this experiment:

- The prediction accuracy using recursive LPC is fairly high when predicted time ahead (PTA), which compensates for the acoustic–electric–acoustic loop delay, is small. But it decreases rapidly as the PTA increases. Prediction accuracy using pitch repetition is fairly constant, but at somewhat lower level than recursive LPC with small PTA.
- Speech cancellation of up to 10 dB is possible, but this cancellation is highly speaker dependent.
- The PTA and the prediction accuracy do not affect the cancellation level significantly. The primary to secondary loud speaker distance does not affect the cancellation level significantly.
- The secondary source, *i.e.* the loud speaker, should be oriented in the same direction as the primary source, *i.e.*, the mouth. The direction in which the largest cancellation is possible is along the line joining the two sources.

As noted above, the secondary speaker needs to be placed close to the mouth facing the same direction. One possible solution for this set up is depicted in Fig. 16. The speaker is placed on the lower part of a handset, where the microphone would be in a regular handset, facing outward. The speech input can be obtained from an ear-insert microphone, as suggested before. We could use a

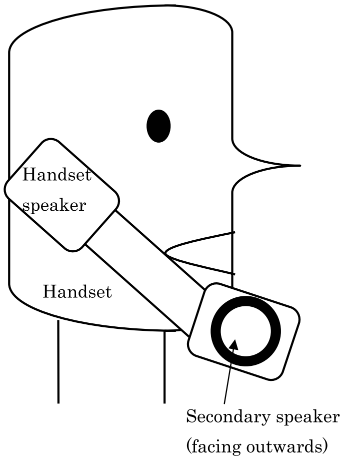

Fig. 16. One possible configuration of the secondary speaker.

piezoelectric vibrator for secondary output for less space and thickness. The vibrator normally shows small delay, which should help increase the prediction accuracy. However, low frequency characteristics are known to be somewhat inferior compared to dynamic speakers. These tradeoffs need to be investigated in detail. Also, the power required to continuously generate the predicted replica speech is another issue that needs to be addressed.

We would further like to improve the speech cancellation level, as well as its speaker dependency. Although not investigated in detail in this paper, the optimum LPC order and analysis block length seems to be speaker dependent. Thus, a training run to obtain the optimum values for each speaker may help improve the prediction accuracy and reduce the speaker dependency. Also, we are currently using fixed LPC coefficients for recursive prediction. The LPC coefficients are fixed at estimated values from the latest available speech samples. Use of extrapolated LPC coefficients may improve the LPC gain to some degree. However, we may need to investigate a more sophisticated speech modeling techniques than conventional LPC for significant improvement.

Finally, we also need to decrease the computation required to predict speech. Accordingly, we would like to implement this method using the state-of-the-art DSPs for real-time operation. The proposed method is still rather expensive to implement in real-time on currently available DPSs. Since we are using recursive prediction, conversion to a non-recursive estimation can reduce the computational complexity significantly. Also, combination of pitch repetition and LPC may reduce the complexity as well.

Acknowledgements

Part of this work was supported by Yamagata University Research Institute (YURI), the Telecommunications Advancement Foundation (TAF), and the Ministry of Education, Culture, Sports, Science and Technology. We thank them for their generous support. The authors also thank members of the Nakagawa Laboratory for their support and discussions.

References

Black, R.D., 1957. Ear-insert Microphone. J. Acoust. Soc. Am. 29, 260–264.

Elliot, S.J., 2001. Signal Processing for Active Control. Academic Press, San Diego.

Flanagan, J.L., 1960. Analog measurements of sound radiation from the mouth. J. Acoust. Soc. Am. 32, 1613–1620.

Haykin, S., 1996. Adaptive Filter Theory. Prentice-Hall, Upper Saddle River.

ITU-T Recommendation G.711 Appendix I, 1999. A high quality low complexity algorithm for packet loss concealment with G.711.

Japan Information Processing Development Corporation, 1991. ASJ continuous speech corpus for research.

Kondo, K., Nakagawa, K., 2002. Active Speech Cancellation for Cellular Speech. In: Proc. International Conference on Spoken Language Processing. Denver, CO.

Kondo, K., Nakagawa, K., 2003. On long term prediction for active cellular speech emission control. In: Proc. Internoise 2003. Seogwipo, Korea.

Kondo, K., Nakagawa, K., 2005. Experimental Evaluation of an Active Speech Control Method. In: Proc. International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA.

Kuo, Sen M., Morgan, Dennis R., 1996. Active Noise Control Systems. Wiley, New York.

Nelson, P.A., Elliott, S.J., 1992. Active Control of Sound. Academic Press, San Diego.

Ono, H., 1977. Improvement and Evaluation of the vibration pick-up type ear microphone and two-way communication device. J. Acoust. Soc. Am. 62, 1613–1620.

Sano, H., Inoue, T., Takahashi, A., Terai, K., Nakamura, Y., 2001. Active control system for low-frequency road noise combined with an audio system. Trans. Speech Audio Process. 9, 755–763.

Tichy, Jiri, 2001. Applications of active noise control. Proc. Internoise 2001. The Hague, Holland.

Wan, E., Nelson, A., Peterson, R., Speech Enhancement Assessment Resource (SpEAR) Database. <http://cslu.ece.ogi.edu/nsel/data/SpEAR_database.html>. Oregon Graduate Institute of Science and Technology.