Adaptive Post Processing of Nonlinear Amplified Sound Signal

Jae-Kyu Lee, Jong-Suk Choi, Cheong-Gyu Seok, and Munsang Kim

Intelligent Robotics Research Center,   *Korea Institute of Science and Technology, Seoul, Korea*

(Tel : +82-2-958-{6816, 5618, 5630,5623}; E-mail: {jklee1965, cjs, robotics36, munsang }@kist.re.kr)

Abstract: We propose a real-time post processing of nonlinear amplified signal to improve voice recognition in remote talk. In the previous research, we have found the nonlinear amplification has unique advantage for both the voice activity detection and the sound localization in remote talk. However, the original signal becomes distorted due to its nonlinear amplification and, as a result, the rest of sequence such as speech recognition show less satisfactorily results. To remedy this problem, we implement a linearization algorithm to recover the voice signal's linear characteristics after the localization has been done.

Keywords: nonlinear amplification, linearization algorithm, post processing, speech recognition

1. INTRODUCTION

Recognizing human speech is a challenging task and has been developed greatly. Recently, sound localization also becomes a technology of interest in the HRI (Human Robot Interaction) area in order that we recognize speech with high confidence [1-2]. The detailed procedures of the localization are to amplify sound signal, to remove the noise, to separate human speech from various sounds by detecting voice activity, and finally to locate the source of speech. Such localization can be also coordinated with the usage of computer vision or other sensors to achieve human-like behaviours. For the localization, first we collect the sound signal by a specific nonlinear amplifier through the microphones.
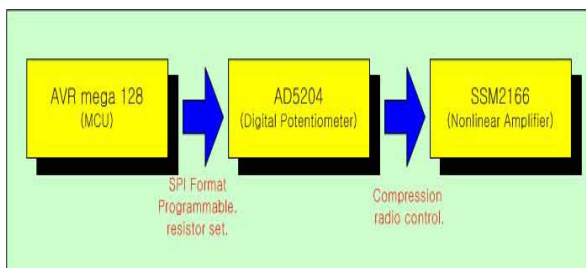

Fig.1 Schematic diagram of programmable SSM 2166 nonlinear amplification board v.2.0

Our previous research [1-2] shows the nonlinear amplification gives a high confidence of voice activity detection (VAD) even in remote talk; however, it has a disadvantage of signal distortion due to its nonlinearity. And such a distorted signal does negative contribution to source separation and speech recognition since many of separation and recognition techniques use linear signals for their applications [3]. To remedy this problem, an algorithm for the linearization of the speech signal has been suggested after the localization. By doing this, we will find a reasonable improve to the requirement of speech recognition without any change in the hardware specification.

2. NON LINEAR AMPLIFICATION FOR LOCALIZATION AND SPEECH RECOGNITION

In this section, we briefly introduce nonlinear amplifier version 2.0. We have made a specific board for the nonlinear amplification using the SSM 2166 [4] and updated it (the version 2.0). The main characteristic of this board is that the compression ratio (CR) can be adjustable via computer programming. Therefore once installed inside of prototype robot, called IROBAA (Intelligent ROBot for Active Audition), one need not to change the board to acquire different amplification ratio. All they must do is change the amplification ratio via connected computer by reprogramming. Figure 1 show the flow diagram of the reprogramming process of the nonlinear amplifier board v2.0 based on SSM 2166, and how to adjust the requested value by the software set up.
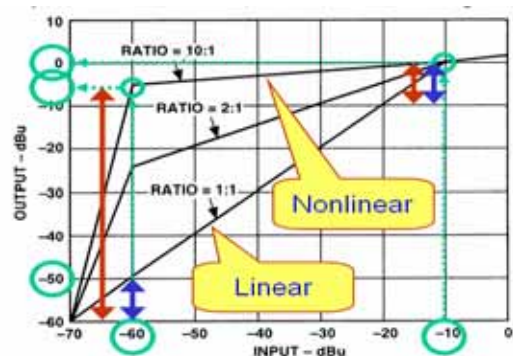

Fig. 2 Nonlinear amplification characteristics of SSM 2166 A/D board

By adjusting a value of resistance by programming, we can change not only the compression ratio but also the amplification ratio.

The other characteristic of this board is it can be used as nonlinear amplification as well as linear amplification. The nonlinear amplification means magnify certain scope of input signal in different scale of amplifying ratio. Figure 2 shows the nonlinear amplification scale and scope scheme of the SSM 2166. The *x*-axis represents the value of signal input and *y*-axis shows its amplified value in nonlinear scale. The nomenclature CR means compression ratio and it can be reset by programming. And through the experiment we set the CR value as 5:1 for nonlinear amplification and it means the

signal within the range of -70dBu through -10dBu will be amplified as nonlinearly like CR is 2:1 or 10:1 as shown in figure 2 and figure 3 shows nonlinear amplifier on single board computer as embedded system.

Fig.3 Programmable nonlinear amplification board on embedded system

Using this A/D board at CR is 5:1, we conduct experiment for localization and recognition, then in the same condition, we conduct the experiment again at CR is 1:1. And then we apply our post processing algorithm to those nonlinearly amplified signals. Since the details of nonlinear amplification and localization algorithm is already discussed thoroughly in our previous papers [1-2], we now mainly discuss post processing for linearization based on localization test using nonlinear amplifier. Experiments took place as shown in figure 4 inside of rather ideally quiet living room. A single male speaker locates in the corner of room and plays the same text, which is "go to the living room" at equal interval toward the robot named IROBAA. This localization test was performed several times, with distance of 0.5m, 1m, 2m, and 3m. L&H engine has been used for recognition. The L&H engine is commercial software and it receives linear signal for recognition in 11 kHz, then gives the recognition results and its confidence value. The test room is size of 130m^2 just like apartment living room.
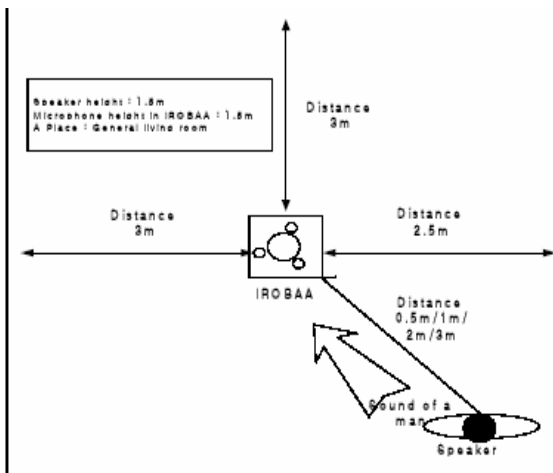

Fig.4 Experimental setup

Figure 5 shows experiment result of localization. The *x*-axis is distance between sound source and robot, and the *y*-axis shows successful rate in percentage rate. The localization results at CR is 5:1 in solid line shows excellent feat up to 3-meter range, whereas that of at CR is 1:1 in dotted line decreases rapidly after 2-meter range. This is mainly due to linear amplifier cannot detect weak signal from distance of 2 meters. However, nonlinear amplification can detect sound signal up to 3 meters.
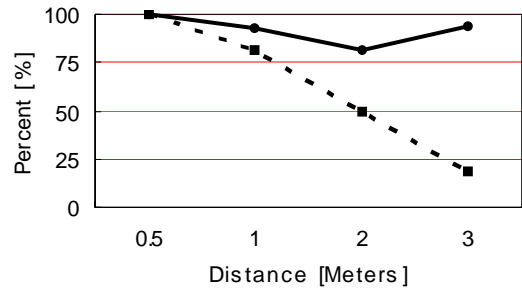

Fig.5 Localization result using nonlinear amplification board

For the purpose of human-robot interaction, it is necessary for performing speech recognition and synthesis as well as detection of sound's direction. In IROBAA, a commercial speech recognition engine performs for speech recognition. The recognition results using nonlinear amplifier will be introduced in section IV along with post processing results.

3. LINEARIZATION USING A POST PROCESSOR

Even though the localization performed successfully by nonlinear amplifier at long distance, it is still necessary to utilize linear signal for recognition. For this purpose, we applied linearization scheme after localization accomplished. To do this, first we look into the nonlinear signal. In figure 6, the linear speech signal is shown in the left and nonlinear signals from SSM 2166 board with CR is 5:1 is shown in the right. It is obvious that the nonlinear signal suffered distortion at the top and bottom of valley and its shape of envelop also has been changed.

Fig. 6 Signal comparison: linear and nonlinear (from left to right)

Therefore, strictly speaking, those shapes of nonlinear signal must come back to linear one. In our application, signals can only be processed after being received; therefore, a post-processor can be applied to linearize such system in figure 7. Here we introduce inversion modeling of the linear behavior of a nonlinear system will be used. Let L_p^{-1} indicate the inverse linear operator obtained by an adaptive linear filter which performs reverse modeling of a physical system as set of polynominal coefficients. Then the relationship of our linear system $u(k)$ and nonlinear system $y(k)$ shows as in equation (1).

$$y(k) = L_p u(k - \delta) \tag{1}$$

where δ is any possible delay by discrete data

And equation (1) can be rewritten as follows:

$$u(k - \delta) = L_p^{-1} y(k) \tag{2}$$
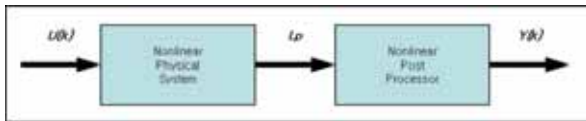
where L_p^{-1} is inverse vector of L_p

Fig.7 Nonlinear post processor is placed at the output side of the nonlinear physical system to post distorted signal

From both equations (1) and (2), if we are able to extract L_p^{-1} then the nonlinear system becomes linear. Since many signals have polynominal characteristics and therefore considered as series [7]. And here we choose Lagrange polynominal form to represent our sound signal for linearization. Equation (3) expresses the shape preserving Lagrange interpolation for any given sample data $f(x_i)$.

$$f_n(x) = \sum_{i=0}^{n} L_p^{-1}{}_i(x) f(x_i) \tag{3}$$

where $L_p^{-1}{}_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}$

In this case, by define and find L_p^{-1} would be the solution of linearization process. Therefore, by adjusting these coefficients, we find the most suitable L_p to make approach the linear signal. The details of procedures are, therefore, first, driving fitting equation from figure 2, second, using equation (1) through (3) organize coefficients vector matrix, and then finally apply those values to the polynominal of nonlinear signal. After we have newly developed signal as linearized one, we calculate cross correlation and compare the results with the original linearly amplified speech signal.

4. EXPERIMENTAL RESULTS

Figure 8 shows two experimental results of our post

processing. The left one is linearized result by a group of coefficients vectors named "group I" after we make efforts to convert nonlinear signal for localization. Compare to nonlinear signal from figure 6, the top and bottom of valley of signal have more round shape but there is still some "bulges" at the top and the bottom of signals. And the right picture in the figure 8, such bulges becomes more smoothed by using coefficients "group II." Based on these results, we conclude such linearization considering sound signals as polynominal type and adjusting their coefficients is the reasonable approach.
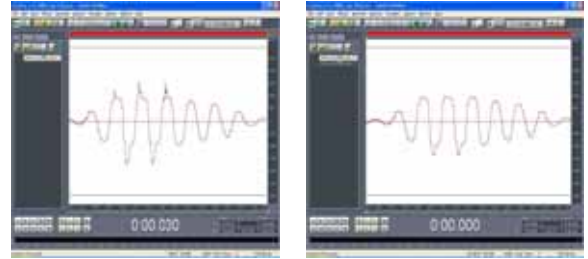

Fig. 8 Signal comparison: linearized I and II (from left to right)

During the post processing, every source of speech signal is the same sine wave of 250 Hz frequency which is the same as human voice signal and distance between sound source and microphone is all the same. In figure 9, we have cross correlation comparison between linear and nonlinear signal with post processing with coefficients group I (solid line), and linear and nonlinear signal without it (dotted line).

As we can see the linearized results perform poorer than the original nonlinear signal at all regions. The x-axis in figure 9 shows numbers of signal frames contains 21 sampled frame values each and the y-axis shows its correlation value.
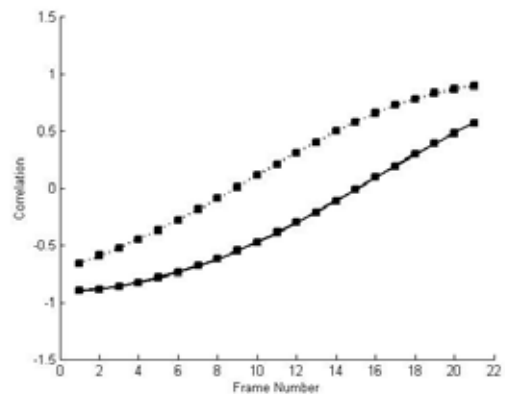

Fig. 9 Numerical results with L_p coefficient group I after post processing

However, figure 10 shows better results. As we can see, more than 18 out of 21 frame, the cross correlation become bigger than without post processing when we apply the coefficients vector group II. And this fact implies our approach to post processing by adjusting coefficients vectors for interpolation and compare those results by calculating

cross correlation has potential and prospect to overcome the distortion problem due to nonlinear amplification. If we discover perfect correlation of 1, it will be the perfect solution. And we are working on iteration algorithm by using cross correlation values to define tolerance.
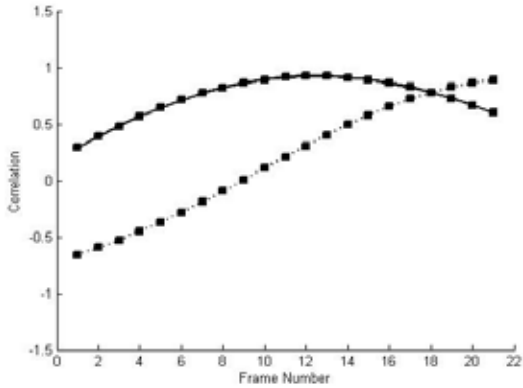
remedy such "bulges" in the signal as shown in figure 8. And figure 11 illustrates the flow process of our implementation.
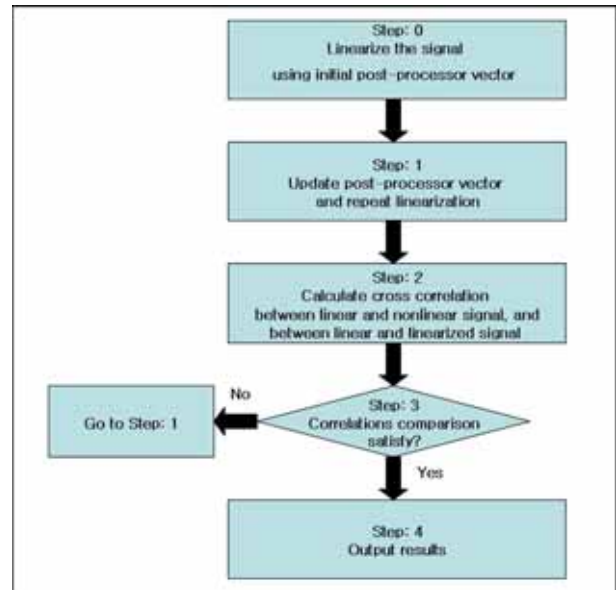

Fig. 10 Numerical results with L_p coefficient group II after post processing

Fig. 11 Flow chart for adaptive linearization using a post processor

We also attempt recognition for sound text using our post processing algorithm and achieved the same successful results as shown in Table I. Table I shows our recognition results of post processing using IROBAA. The experiment conducted 10 times for the same sound text, which is "go to the living room" and the recognition result shows superior performance at any ranges by judging its confidence value. These confidence values are given by L&H (Lernout & Hauspie Speech Product) [8] engine in real time process during the test. Table I shows the rate of recognition success from 1 meter to 3 meters distances along with average confidence values. From Table I, in one-meter range, which is close range, it shows better results in case of without post processing; however, the average confidence value is lower. And, in the distance of 2 meters, the results are become improved with post processing.

5. CONCLUSION AND FUTURE WORKS

We conduct localization and speech recognition experiment using non-linear A/D board version 2.0, and also conduct both numerical and experimental research on post processing for linearization. The post processing algorithm is developed based on the idea of Lagrange polynominal and its coefficients vectors are defined to adjust nonlinear signal as linearized one. And based on those results, we made conclusions as follows:

5.1 Conclusions
1) The non-linear amplifier v.2.0 shows satisfactorily performance for the programmable amplification.
2) Through the implementation results in post processing, we conclude there is possibility of linearization of non-linear signal not by modification of hardware specification but by adjusting Lagrange coefficients numerically.
3) In experiment, we found the same results as numerical analysis, and concluded post processing is useful tool for speech recognition of non-linear signal.
4) Our ultimate goal is to find general algorithm that makes perfect resemblance between non-linear and linear signal by linearization, so that we can take advantage of nonlinear signal for localization in one hand and the other hand, linearized signals for source separation and speech recognition.

TABLE I
RECOGNITION SUCCESS RATE IN EXPERIMENTS WITH & WITHOUT POST PROCESSING

Range	*With* post processing		*Without* post processing	
	Success rate [%]	Confidence value	Success rate [%]	Confidence value
1m	50 %	31.4	70 %	19.14
2m	70 %	17.86	10 %	2
3m	50 %	27.2	30 %	8.3

The post-processing algorithm performed satisfactorily up to 3-meter range and showed better confidence value compare to 2-meter range test. In our case, 3-meter range test has better performance than 2-meter range. Generally speaking, our post processing algorithm performs well in real experiment, however, more modification must be added up to

5.2 Future Works
We achieved acceptable performance in post processing through the simulation as well as experiment; however, we still must explore the following problems:

1) We have to find more reliable measure to determine the confidence value of voice recognition.
2) We will need to establish rigid rules to verify the credits of our own post-processing algorithm especially to make the iteration sequence become converge.
3) Explore the echo effect and geometry effect of test room and verify how they influence the sound signal process during the experiment.

REFERENCES

[1] Kim, H., Choi, J., Kim, M., and Lee, C., "Reliable Detection of Sound's Direction for Human Robot Interaction" Proc. of 2004 IEEE/RJS Int. Conf. on Intelligent Robots and Systems, Sep. 28 ~ Oct. 2, 2004, Sendai, Japan.

[2] Kim, H., Choi, J., Kim, M., and Lee, C., "Sound's Direction Detection and Speech Recognition System for Humanoid Active Audition" in Proc. Int. Conf. On Control, Automation, and Systems, Gyeongju, Korea, Oct. 2003, pp. 633 ~ 638.

[3] Okuno, H., et al, "Assessment of General Applicability of Robot Audition System by Recognizing Three simultaneous Speeches", Proc. of 2004 IEEE/RJS Int. Conf. on Intelligent Robots and Systems, Sep. 28 ~ Oct. 2, 2004, Sendai, Japan.

[4] "SSM2166 Technical Manual", Analog Device, 1999.

[5] "Numerical Methods for Engineers", Chapra, S. C., Candale, R. P., MaGraw-Hill, 1998.

[6] Lemout & Hauspie Speech Products, October 2000

[7] Gao, Xiang Y., Snelgrove, W.M., "Adaptive Linearization Scheme for Weakly Nonlinear systems Using Adaptive Linear and Nonlinear FIR Filters", Circuits and Systems, 1990., Proceedings of the 33rd Midwest Symposium on 12-14, Page(s):9 - 12 vol..1 Aug. 1990

[8] Gao, X. Y., Snelgrove, W. M., "Adaptive Linearization for a Loud Speaker", Acoustics, Speech, and Signal Processing,, 1991. ICASSP-91, International Conference on 14-17 Page(s):3589 - 3592 vol.5 April 1991.

[9] Wei-Bing Gao; Dang-Nan Wu;, "On the method of global linearization and motion control of nonlinear mechanical systems" Industrial Electronics, Control, Instrumentation, and Automation, 1992. 'Power Electronics and Motion Control', Proceedings of the International Conference on 9-13, Page(s):1476 - 1481 vol.3, Nov. 1992.

[10] Gao Weihua; Jianying Xie, "Fuzzy feedback linearization control for nonlinear systems", Intelligent Control and Automation, Proceedings of the 3rd World Congress on Volume 3, Page(s):1536 - 1539 vol.3, 28 June-2 July 2000.