International Journal of Hybrid Information Technology
Vol. 9, No.9 (2016), pp. 103-110
http://dx.doi.org/10.14257/ijhit.2016.9.9.11

A Study of Radio Voice Signal Based on the Time Delay Estimation

Xinrui Liu, Jinxiang Chen, Xianbo He, Wei Li and Gangyuan Zhang

Computer School, China West Normal University, Nanchong 637000 ,China
516105211@qq.com

Abstract

The wave of voice signal is narrow and unstable. Besides, it has a non-continuity in time domain and an inconsistent decline of the radio channel, especially the short wave channel, which makes estimating the time delay of the radio's voice signal very difficult. To solve such problems, the authors creatively combine the voice activity detection with the short-segment processing technology. Meanwhile, by using Weighted Least Squares Phase Fitting method to work out the time delay, we conduct a research in the time delay estimation of radio's voice signal. The laboratory finding finally shows that this method is far better than the typical Generalized Correlation Method for estimation of time delay.

Keywords: *time delay estimation, radio voice signal, short-time analyzing, voice activity detection, passive location*

1. Introduction

In modern signal processing, time delay estimation is an important part of signal detection and parameter extraction, which is widely used in detection, communication, bio-medical science, geophysics and many other fields [1-8]. In particular, the TDOA estimation is the core part of positioning and detecting aided with opportunity transmitter (such as radar and sonar), and its degree of accuracy directly determines that of the target positioning. Now, the time delay estimation has become a mature technology with many studies both in scope and in depth. Among these methods of TDOA estimation, the Generalized Correlation Method [1] tends to be a typical one with the most extensive application. Here is the basic idea of it: first, do the whitening process (*i.e*. pre-filteration) on both two signals; then, lay the time shift on one of the two signals against the other and compare the similarity between the two signals through the Generalized Correlation Method so as to figure out the displacement when the similarity is at its maximum. With the whitening process on the receipt signal ,this method makes the correlation wave crest of the two signals sharper ,thus achieving the improvement in the resolution and reliability of the measured TDOA. Under the assumption that both the signal and the noise should follow the stationary Gaussian process, the Generalized Correlation Method is the best for time delay estimation as the estimated variance gradually approaches the Cramer-Rao lower bound.

Radio's voice signal, short-waved and ultrashort-waved radio broadcasting voice signal, for instance, is one of the common signals in the radio-communication. The time delay estimation is mainly applied to these following fields:

(1) The wide-area diversity reception of the voice signal in distant (up to thousands of miles) short-wave communication. Usually, the stability of such communication is rather poor because of the unsteadiness of the ionosphere. And a practical approach to solve this problem is to use several receptors which are of wide-area distribution to receive the same short-wave signal. It is worth mentioning that receptors should be far enough, hundreds of miles for instance, away from each other to make sure that every channel the receptor corresponds to is independent from each other. Then, the communication effectiveness

ISSN: 1738-9968 IJHIT

would be improved by the diversity combining of all the receptors' output. During this process, we need to estimate the relative time delay of the signals put out by different receptors so as to combine them.

(2) Radio voice signal's source tracking and positioning. This is mostly used in the reconnaissance and location of radio stations, which is a vital part of electronic reconnaissance in electronic countermeasure.

However, in practical application, the Generalized Correlation Method doesn't work out very well in estimating the radio voice signal's time delay. There are three reasons accounting for it. The first one is that the voice signal is quite narrow and its energy mainly consists in the range of 50Hz ~ 2kHz; The second reason is that the voice signal is neither continuous in the time domain nor is it stationary; And the third one is that the channel fading's nonuniform of the real radio wave, especially the short wave radio, causes the correlation among the outputs of different receptors decreases greatly. To get rid of these demerits, we put forward a new method to do the TDOA estimation. It combines the voice activity detection with the short segmentation processing technology. Meanwhile, we choose Weighted Least Squares Phase Fitting Method to work out the time delay. Specifically, because the voice signal is discrete in time domain,we draw the voice fragments from signals through the voice activity detection to diminish the noise interference and increase the reliability of the time delay estimation. Then, for the instability of the voice signal and the uniformity of the channel ,we adopt the short segmentation processing technology to get a better estimation of the two signals' cross correlation spectrum linear phase, which remarkably represses the outliers on the phase. Finally, based on the cross-correlation linear phase estimation we get, the TDOA estimation is quickly figured out through the Weighted Least Squares Phase Fitting method. Now, the author will present the comparative experiment between the Generalized Correlation Method and the new one to prove the effectiveness of the later.

2. The Signal Model and the General Correlation Method

We assume there is neither multipath nor Doppler Frequency Shift and ignore the frequency difference during the demodulation. Then, the baseband voice signals received by two receptors can be expressed in the following way:

$$\begin{aligned} x_1(t) &= A_1(t)s(t) + n_1(t) \\ x_2(t) &= A_2(t)s(t-\tau) + n_2(t) \end{aligned} \tag{1}$$

In the formulas above, $s(t)$ stands for the source of the baseband voice signal ; τ stands for the relative TDOA of the signal at the two receptors ; $n_1(t)$ and $n_1(t)$ stand for mutual independent zero-mean White Gaussian Noise separately ,both independent from $s(t)$;besides, $A_1(t)$ and $A_2(t)$ are the amplitude of fading function of the two channels, in addition ,we assume the change of $A_1(t)$ and $A_2(t)$ is slow, *i.e.* we can assume that during a short period , $A_1(t) = A_1$ and $A_2(t) = A_2$, A_1 and A_2 are both constants.

In this way, the function of $x_i(t)$'s Fourier Transform can be presented as follows:

$$X_i(\omega_k) = \int_0^T x_i(t) e^{-j\omega_k t} dt \qquad \omega_k = \frac{2\pi K}{T} \tag{2}$$

$i = 1, 2$, $k = 1, \cdots K$. According to the Generalized Correlation Method, the time delay estimation is calculated by working out the maximum value of the following formula:

$$R(\tau) = \sum_{\omega_k} W_1(\omega_k) W_2^*(\omega_k) X_1(\omega_k) X_2^*(\omega_k) e^{j\omega\tau} \tag{3}$$

There into, $W_i(\omega_k)$ is the weighted function related to the recorded information about signals and noise. In practical project, the prior information about voice or signals are generally unknown, hence the estimated value is often used to substitute the recorded theological value of the signals and noise.

2.2. New Method for TDOA Estimation

The new method makes a corporation of the voice activity detection and the short segmentation processing technology. Also, the Weighted Least Squares Phase Fitting method is adopted to work out the time delay estimation. Here are the details:

2.3. The Pre-Processing of the Voice Signal Activation Detection

Figure 1 is about working out the voice signal activity detection ,also called speech/non-speech detection[9-11], is chiefly applied to detect the voice among voice signals, widely used in speech recognition, speech coding and speech enhancement. Through it, we can estimate the time delay after splicing the voice fragments, thus reducing the interference of the noise when the voice signal is absent.

We suppose the noise be uncorrelated White Gaussian Noise, and each segment follow the two assumptions below:

$$\begin{aligned} H_0(non-voice): X &= N \\ H_1(voice): X &= N - S \end{aligned} \tag{4}$$

Additionally, S, N ,X are respectively the K dimensional discrete Fourier coefficient vector of the voice signal ,the noise and the received segment ,then the KTH ones are $S(\omega_k)$, $N(\omega_k)$, $X(\omega_k)$ accordingly. Under the assumption that the voice signal and the noise is in accordance with Gaussian Processes, the final decisive can be concluded as the expression below in the sense of likelihood ratio:

$$\frac{1}{K}\sum_{k=0}^{K-1}\{\gamma_k - \log\gamma_k - 1\} \underset{H_0}{\overset{H_1}{>}} \eta \tag{5}$$

In this expression, η indicates the likelihood ratio decision threshold , $\gamma_k = |X(\omega_k)|^2/\lambda_N(k)$, the posteriori SNR and $\lambda_N(k)$, the variance of $N(\omega_k)$ ——noise spectrum

The algorithm takes advantage of the difference between the voice signal and the noise in spectrum distribution. Under the assumption of Gaussian distribution, we calculated the likelihood ratio of the two scenarios mentioned in the expression 4 to decide the voice signal and the noise. As the TDOA information only exists in the voice activation period, to use this period only turns out to be quite effective in reducing noise's impact on the time delay estimation.

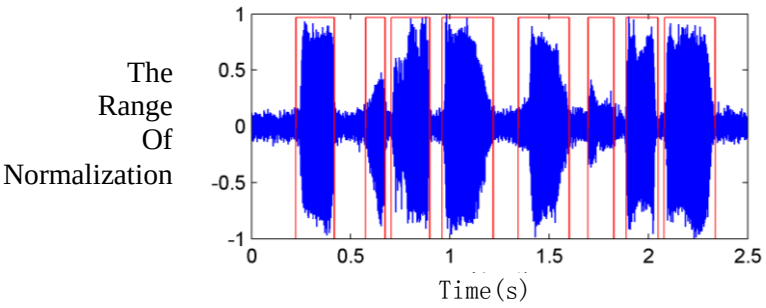

Figure 1. The VAD Output of A Voice Signal

2.4. The Short-Segment Processing

Since the voice signal is unstable, the approach applied to these stable ones doesn't fit it. Some researches indicate that the generating of voice is related to the change of muscle's movement---a relatively slow movement ,as it were ,within a short period ,we can view the feature of the voice remains the same(the voice signal remains to be relatively steady during this period). So, based on this idea, the Short-segment Processing technology are generally used in voice signal processing.

In practical use, this method can effectively avoid the interference and the outliers of the noise. On the foundation of the voice signal activation detection's output, we use the average short-segment processing method to estimate the frequency coherent spectrum of the two channels' signals. Specifically, let's assume that the length of the data that receptor i outputs is N , $i=1,2$. Then, we divide this data into L sections, so every section has a length of N/L . Finally , by an Fourier transform on all the sections, we obtain the data of the frequency domain—— $X_{i,l}(\omega_k)$, $l=1,...,L$

$$X_{i,l}(\omega_k)=\sum_{m=0}^{N-1}x_i(m)w(n-m)e^{-j\frac{2\pi k}{N}m}, \qquad 0\le k\le N-1 \tag{6}$$

There into, $w(n-m)$ is the window function. Afterwards, we work out the cross-correlation spectroscopy of the corresponding section's data of the two channels' $G_{X_1X_2,l}(\omega_k)=X_{1,l}(\omega_k)X_{2,l}^*(\omega_k)$, $l=1,...,L$. And in the end, we take the average of the L sections' cross-correlation spectroscopy $\{G_{X_1X_2,l}(\omega_k)\}_{l=1,...,L}$ and then obtain the average cross-correlation spectroscopy $\bar{G}_{X_1X_2}(\omega_k)$.

Figure 2 is about working out the two signals' (voice activation pre-processed joint-data) average phase difference value (*i.e.* the average phase of Cross-correlation spectroscopy), and more details about the experiment could be seen in the third part. Through the graph, we can conclude that the averaging short-segment processing method can effectively avoid the interference of the outliers, making the two channels' signals phase difference approach the linear phase.

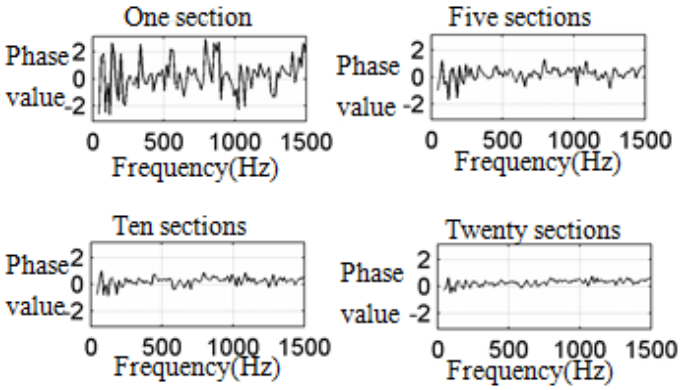

Figure 2. Two Channels' Signals' Sub-Sectional Average Phase Difference (Average of Mid-Value, Every Section 60ms)

2.5. Weighted Least Squares Phase Fitting Method

It is a simple and practical method to estimate the time delay through measuring the two signals' Cross-correlation spectroscopy linear phase. In the ideal situation where the noise is statistically independent and the signal source is non-scattering, this method can achieve the same precision as other TDOA estimating methods. Meanwhile, in the non-ideal case, it has the following potential methods to estimate the time delay through the phase. Namely, when the received noise is correlative or coherent, the adoption of this method can reduce the estimating bias through choosing the proper frequency.

Here we assume that the function of the channel fading to be a constant within a short period, so to speak, $A_1(t) = A_1$ and $A_2(t) = A_2$,then from signal model(1) ,we can theoretically get the cross-correlation spectroscopy theory of two channels' signal as the equation below:

$$G_{X_1X_2}(\omega_k) = | A_1 |^2 | A_2 |^2 | S(\omega_k) |^2 \ e^{j\omega_k\tau + jc} \tag{7}$$

Pick the phase angle of $G_{X_1X_2}(\omega_k)$ to obtain the phase sequence $\mathbf{p} = [\omega_0 \ \ \omega_1 \ \cdots \ \omega_{K-1}]^T \tau + c\mathbf{e}$. Here, c is a constant, and e a vector of matrix laboratory . As the phase sequence p and the to-be-estimated time delay t are linearly related, we can use the least square method to fit the vector p. Let $\boldsymbol{\omega} = [\omega_0 \ \ \omega_1 \ \cdots \ \omega_{K-1}]^T$,and $\mathbf{W}$ also the weighting matrix(like Cross-correlation spectroscopy amplitude weighted), so that the this least square problem can be presented in the following way:

$$\hat{\tau} = \arg\min_{\tilde{\tau},\tilde{c}} \| \mathbf{W}(\mathbf{p} - \boldsymbol{\omega}\tilde{\tau} - \tilde{c}\mathbf{e}) \|_F^2 \tag{8}$$

Then here the least square closed-form solution to the equation is：

$$\hat{\tau} = \frac{1}{[\boldsymbol{\omega} \ \ \mathbf{e}]^T \mathbf{W} [\boldsymbol{\omega} \ \ \mathbf{e}]} [\boldsymbol{\omega} \ \ \mathbf{e}]^T \mathbf{W}\mathbf{p} \tag{9}$$

If $\mathbf{W} = \mathbf{I}$, and $\mathbf{I}$ is a unit matrix, the formula above would be a standard least square estimation. Over the practical processing, the phase statistic is obtained through the Short-segment average Processing method. Comparing with the generalized correlation method, this one stands out in that it figures out the time delay estimation directly from the 9th formula, free from the spectral peak search.

3. The Result

In this part, we present the result of the processing of the signals through our new method, and also the result through the typical Generalized Correlation Method . In addition, we evaluate the effects of the both. Here the implementing steps are concluded as below.

First, choose the signal that has a bigger signal-noise ratio to receive the pre-process of the voice activity detection technology after estimating the SNR of both two channels' signal, and concatenate the corresponding voice segments extracted from the two channels according to the output. Second, use the short-segment average method (introduced in part2.2) to estimate the cross-correlation spectroscopy based on the output of last step. Third, after obtaining the cross-correlation spectroscopy, we detect its phase to estimate the time delay with the Weighted Least Squares Phase Fitting Method. The length of one section in the new method is 60 milliseconds. While in the Generalized Correlation Method, the searching gap is 40 milliseconds, and the second-order spline interpolation around the searched peak gets a better resolution ratio.

The outdoor experiment captures the interphone amplitude modulation voice signal, which carries the wave of 400MH, and the baseband's signal sampling frequency is 100kHz after demodulating. With the time service of the GPS, the two receptors sample simultaneously. The distance between the two receiving antenna is 300 meters, the distance between the interphone and one antenna is 240 meters longer than the interphone to another one. Thus the the time delay caused by the difference of transmitting distance is about 800ns.

Figure 3 presents the signals of the two channels (the signal beyond 2 kHz filtered), and the red wireframes are outputs of the voice activation. Figure 4 shows the effect of the time delay estimation through our new method. It can be seen that when the cumulative voice signal is longer than 0.6 second, the new estimated time delay is within the error range of only ± 100 ns, and the longer, the steadier. Meanwhile, the result of the Generalized Correlation Method is 321.78ns. By contrasting, we can easily see the advantage of the new method.

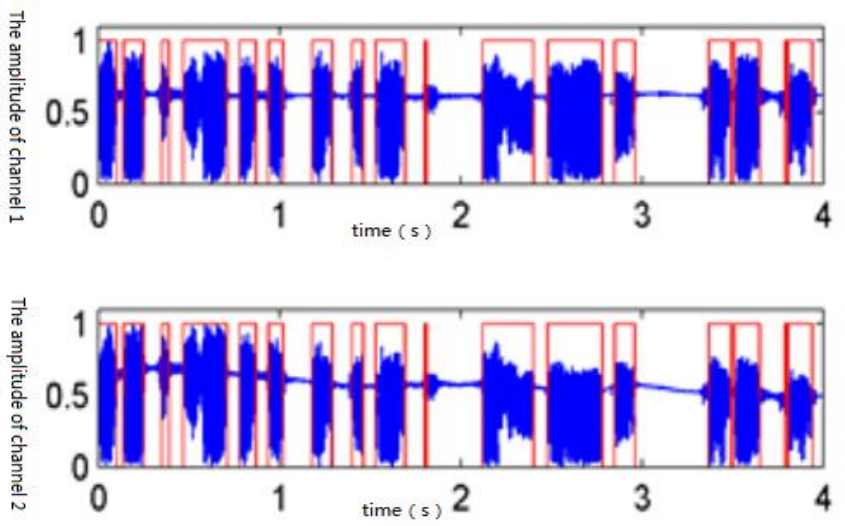

Figure 3. The Output of the Outdoor-Captured Voice Signals Activation Diction

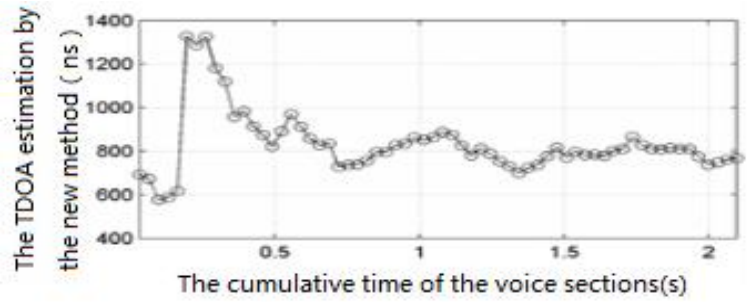

Figure 4. The New Method's Laboratory Time Delay Estimation (60ms Per Section)

4. Conclusion

The difficulty in estimating the time delay of the radio voice signal lies in that the wave of voice signal is narrow and un-stationary. Besides, it has a non-continuity in time domain and an inconsistent decline of the radio channel (especially the short wave channel). However, in spite of these difficulties, the creative method we put forward can apply to such signal's diversity reception, tracking and positioning. It is a scientific combination of the voice activity detection, the short segmentation processing technology and the Weighted Least Squares Phase Fitting method, making the time delay estimation more accurate. The result of the outdoor experiment proves that it is a better method comparing with the Generalized Correlation Method.

Acknowledgements

This work has been Supported by Application of Sichuan Provincial Office of Science and Technology Program (No. 2013SZ0056) of China, Sichuan Provincial Office of Science and Technology Support Program (No. 2014SZ0104) of China, and Educational Reform Project of China West Normal University (No. JGXMQN1329).

References

[1] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for estimation of time delay", IEEE Trans. Acount.Speech, Signal Process., vol. 24, no. 4, **(1976)**, pp. 320–327.
[2] G. C. Carter, "Coherence and time delay estimation", IEEE Proc., vol. 75, **(1987)**, pp. 236–255.
[3] J P. Ianniell, "Time delay estimation via cross-correlation in the presence of large estimation errors", IEEE Trans. Acount. Speech, Signal Process. vol. 30, no. 6, **(1982)**, pp. 998–1003.
[4] A. H. Quazi, "An overview on the time delay estimate in active and passive systems for target localization", IEEE Trans. Acount. Speech, Signal Process., vol. 29, no. 3, **(1981)**, pp. 527–533.
[5] F. Wen and Q. Wan, "Robust time delay estimation for speech signals using information theory: A comparison study", EURASIP Journal on Audio, Speech, and Music Processing, **(2011)**, http://asmp.eurasipjournals.com/content/2011/1/3.
[6] X. Tian, R. Li and W. Wang, "An Efficient Time Delay Estimation Algorithm for Multipath Signal of Distance Signal in TACAN System", Journal of Electronics & Information Technology, vol. 32, no. 9, **(2010)**, pp. 2273–2276.
[7] W. Ren and D. Hu, "Eliminating TDOA Location Ambiguity of High PRF Signal Based on Direction Information Acquired", Journal of Electronics & Information Technology, vol. 32, no. 12, **(2010)**, pp. 3003–3007.
[8] Q. Wan, Y.-P. Du and Z.-J. Lü, "Toa Location Algorithm Using Virtual Matrix", Journal of University of Electronic Science and Technology of China, vol. 38, no. 1, **(2009)**, pp. 43-46.
[9] J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detection", IEEE Signal Process. Let, vol. 6, no. 1, **(1999)**, pp. 1–3.
[10] J. W. Shina, J. Changb and N. S. Kim, "Voice activity detection based on statistical models and machine learning approaches", Computer Speech & Language, vol. 24, no. 3, **(2010)**, pp. 515–530.
[11] P. K. Ghosh, A. Tsiartas and S. Narayanan, "Robust Voice Activity Detection Using Long-Term Signal Variability", IEEE Transactions on Audio Speech and Language Processing, **(2011)**, vol. 19, no. 3, pp. 600–613.

Authors

Xinrui Liu, She is a graduate student at computer school of china west normal university, her research interests include Artificial Intelligence, tracking signal trip.

Jinxiang Chen, He was born in March 1971. In 1995 graduated from the chengdu sports college sports education professional, Now he is working in china west normal university.

Xianbo He, He received the Ph.D. degree in Computer science and technology from college of computer science, Sichuan University, chengdu, China, in 2008. From 2012 to now, He was a professor. His research interests include principles of computer operating system Professor, data structure, the Linux operating system, the embedded software development technology and so on.

Wei Li, She was born in February, 1982. Now she is working in computer school of china west normal university. As the main courses are "XML based", "data structure", *etc*. She mainly engaged in pattern recognition, embedded technology. She has published around 10 papers.

Gangyuan Zhang, He was born in November, 1962. Now he is working in computer school of china west normal university. From 2011 to now, He was a associate professor. His research interests include principles of computer operating system, C language programming, database applications and systems development, PASCAL programming, computer basic teaching method, *etc*.