Automotive 3-Microphone Noise Canceller in a Frequently Moving Noise Source Environment

Z. Qi and T. J. Moir

***Abstract*—**A combined three-microphone voice activity detector (VAD) and noise-canceling system is studied to enhance speech recognition in an automobile environment. A previous experiment clearly shows the ability of the composite system to cancel a single noise source outside of a defined zone. This paper investigates the performance of the composite system when there are frequently moving noise sources (noise sources are coming from different locations but are not always presented at the same time) e.g. there is other passenger speech or speech from a radio when a desired speech is presented. To work in a frequently moving noise sources environment, whilst a three-microphone voice activity detector (VAD) detects voice from a "VAD valid zone", the 3-microphone noise canceller uses a "noise canceller valid zone" defined in free-space around the users head. Therefore, a desired voice should be in the intersection of the noise canceller valid zone and VAD valid zone. Thus all noise is suppressed outside this intersection of area. Experiments are shown for a real environment e.g. all results were recorded in a car by omni-directional electret condenser microphones.

***Keywords*—**signal processing, voice activity detection, noise canceller, microphone array beamforming

I. Introduction

THE most challenging of in-car speech recognition problems is picking up a speech signal from a desired source e.g. a driver's voice, rather than mechanical noise and other passenger's speech. The mechanical noise emanates from a number of sources including the engine, road, wind and air-conditioner. Other passenger's speech as well as speech from the radio is also a challenge to speech recognition.[1] Microphone array beamforming is a well known solution to this issue and has been studied for some thirty years. It has applications to such areas as communications [2], hearing aids[3], speech-recognition[4] robotics[5] and hands-free telephony[6]. A real-time beamformer can be used to reduce the effects of noise on a speech signal. A two microphone approach can be used with one microphone near the desired speech and a second microphone near the noise source[7]. The resulting adaptive filter is updated using the least-mean-squares algorithm (LMS)[8]. This approach is successful when the speech signal is far enough away from the noise so that elements of the speech are not picked up by the noise microphone. It is well known that noise cancellation (Widrow noise canceller) works well when the disturbing noise emanates from a point source. It does not work well when the noise is diffusing. [9, 10]

When all mechanical noise and undesired speech come from unknown directions, a microphone array beamformer is used to enhance speech from a geometrical zone and reduce any other speech or noise outside of this zone.[11] In order to improve hands-free speech recognition performance in car environments, a microphone beamforming array has been implemented with a Voice Activity Detector (VAD) which uses time-delay estimation together with magnitude-squared coherence (MSC). [12] This microphone array has been used to form a beamformer with normalized least-mean squares (NLMS) to improve Signal to Noise Ratio (SNR). The experiment clearly shows the ability of the composite system to reduce noise outside of a defined zone. Experiments have been conducted in real-time on a combined three-microphone VAD and noise-canceling system. The VAD assumes that the desired speech falls within a desired geometric zone in free-space which is most appropriate for an automobile environment as it can be defined around the drivers head. The noise-canceling is only required when noise is present during desired speech as the VAD will mute any solo noise-source outside of the zone. The experiment used only pre-recoded phrases. This work clearly demonstrates the ability of the algorithm to cancel speech outside of the zone.

However, in a frequently moving noise sources environment, the noise cancellation needs to suppress the unwanted noise when desired speech is also present. This paper investigates this problem in some detail with real-time experiments clearly showing the performance of the canceller.

II. Algorithm

A. Three-microphone VAD switch

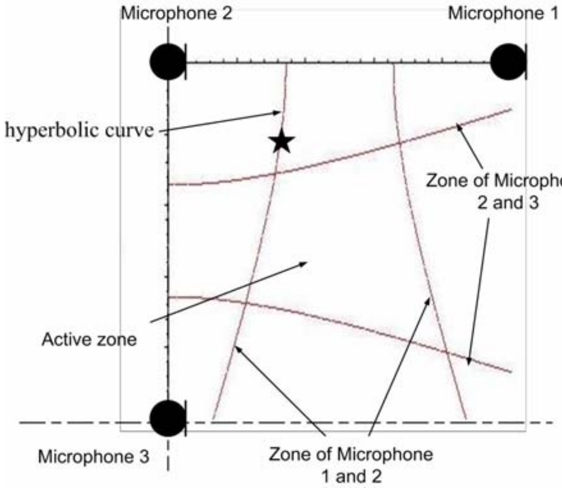
Carter et al.[13] describe a method for estimating the magnitude-squared coherence (MSC) function for two zero-mean wide-sense-stationary random processes. The estimation technique utilizes the weighted overlapped segmentation fast Fourier transform (FFT). Analytical and empirical results for statistics of the estimator are presented. The analytical expressions are limited to the non-overlapped case. Empirical results show a decrease in bias and variance of the estimator with increasing overlap and suggest a 50-percent overlap as

Manuscript received July 17, 2006.

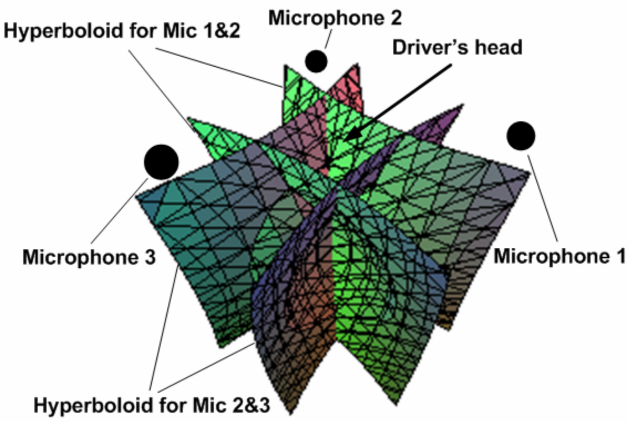
Z. Qi is with the Institute of Information and Mathematic Science, Massey University at Albany, Auckland, New Zealand. (e-mail: tqi@unitec.ac.nz).

T. J. Moir is with Institute of Information and Mathematic Science, Massey University at Albany, Auckland, New Zealand. (e-mail: t.j.moir@massey.ac.nz).

being highly desirable when cosine (Hanning) weighting is used. Once the MSC is found the Generalized Cross-Correlation (GCC) method is used to give a robust estimate of time-delay. A microphone array as shown in Figure 1 is currently located to ensure that there is an intersection. Clearly, a linear array cannot have such an intersection. In Figure 1 three microphones are located as shown and there is 50 cm distance between these microphones. A desired speech source is located 35.4 cm away from Microphone 1, 2 and 3. Therefore, when speech travels to microphones 1, 2 and 3, it has the same distance to travel. The sample rate of Microphone 1, 2 and 3 is 11025 Hz, and the speed of sound in air is around 34600cm/second. Therefore during every sample the speech travels 3.1 cm so that the wave-front of speech arriving at microphones 1, 2 and 3 have no time difference of arrival (TDOA) with respect to one another. When the speaker is away from the desired position, a finite TDOA between Microphone 1, 2 and 3 is expected. For any point on a hyperbolic curve As shown at Figure 1(a), the difference between distances to a pair of microphones (as foci) is constant e.g. speech source from the star point on the hyperbolic curve travels to Microphone 1 has 5 samples intervals delayed with respect to the microphone 2.



(a)



(b)

Fig. 1 Plan view of 3-microphone VAD valid zone (a) 2-D (b) 3-D

The technique can be summarized as follows for three microphones and two estimated time-delays. When the VAD is set to be within some defined number of samples (e.g. 5 samples is typically used), then the estimation of time delay (TDOA)'s from each microphone pair is estimated and compared with some threshold value $d_{\max}$.Therefore an "VAD valid zone" is defined as in Figure 1(a),(b) [12]. In Figure 1(a), the plan view of three microphones 50cm apart is shown. The distance between Microphone 1 and Microphone 3 is 70.7 cm. The microphones need not be at right-angles but are positioned in such a way that the intersection of the two hyperboloids (in 3-D space Figure 1 (b)) form an VAD valid zone around the drivers head [11]. A 3-microphone VAD valid zone can be steered by pre-defined time-difference of arrival (TDOA). For example, the VAD valid zone can be moved towards microphone 1.

B. Normalized Least Mean Square Filter

The Normalise least mean squares (NLMS) algorithm is used extensively in adaptive filtering algorithms due to its simplicity for real-time applications. [14] A NLMS filter block diagram is shown as Figure 2. To define the self learning process the filter uses an adaptive algorithm to reduce the NLMS between the output signal *y(k)* and the desired signal *d(k)*. For stationary (in the statistical sense) signals, when the NLMS performance criteria for the NLMS have achieved its minimum value through the iterations of the adapting algorithm, the adaptive filter is finished and its coefficients have converged to a constant solution. Then the output from the adaptive filter matches closely the desired signal *d(k)*. If the input data characteristics are changed, (sometimes called the filter environment) the filter adapts to the new environment by generating a new set of coefficients for the new data. Notice that when *e(k)* goes to zero and remains there.

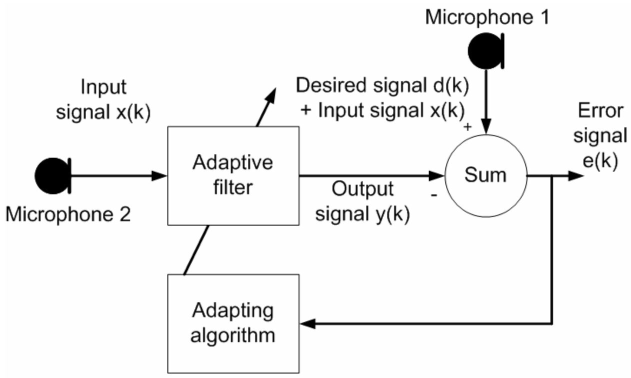


Fig. 2 NLMS adaptive filter as noise canceller block diagram

The NLMS adaptive filter weights are updated accordingly

$$W_{k+1} = W_k + 2\mu e_k X_k \qquad (1)$$

Where the weight vector

$$W_k = [w_{1,k}\ \ w_{2,k}\ \dots\ w_{N,k}]^T \qquad (2)$$

are the coefficients of the adaptive filter at time k,

$$X_k = [x_k \ x_{k-1} \ ... \ x_{k-N+1}]^T \tag{3}$$

are the N samples of the input data in filter memory at time k,

$$e_k = d_k - W_k^T X_k \tag{4}$$

The NLMS algorithm is given by [15]

$$W_{k+1} = W_k + 2\mu e_k \frac{X_k}{\|X_k\|^2 + \delta} \tag{5}$$

where $\delta = 0.0001$ and $\|X_k\|$ is Euclidean norm of X_k and is given by

$$\|X_k\|^2 = x_k^2 + x_{k-1}^2 + ... \ x_{k-N+1}^2 \tag{6}$$

We define Microphone 1 as the primary input and Microphone 2 as the reference input. In Figure 1, experiments [16] show that voice close to the primary input is enhanced while voice close to the reference input is reduced.

C. Three-microphone adaptive Filter

The generalized side-lobe canceller (GSC) with an adaptive blocking matrix is introduced to enhance the desired signal and reduce unwanted signal [17-19]. Similar proposals have been investigated quite extensively. [20, 21]

A three-microphone noise canceller [12] based on Van Compernolle's work [22] is showed as Figure 3. There are four NLMS algorithms in a three-microphone noise canceller. The top path of the beamformer has a summation term which forms the primary input whilst both of the bottom paths have a difference term which forms the reference input. The three microphone signals contain speech as well as noise. The left section of the system serves at improving the noise reference by eliminating speech so that the Voice Activity Detection (VAD) switches this part on when speech energy is dominant. The right section consists of NLMS 2 and NLMS 4, which are only switched on to adapt during the absence of speech (i.e. during noise periods). For these experiments the number of weights used in W1 and W3 were 100 and in W2 and W4, 450. The 3-microphone VAD works so as to switch to freeze or enable the various NLMS algorithms. Also, the VAD switches off (mutes) the signal output when speech does not come from the desired zone. Therefore, while the driver's voice activates the 3-microphone VAD, the 3-microphone noise canceller is expected to reduce a passenger's voice, or for instance the car radio. For this to be successful the noise canceller valid zone must be positioned around the drivers head and be large enough to provide some movement.

In fact, whilst the Enable line in Figure 3 is enabled by the 3-microphone VAD (E = 1) and hence NLMS 1 and 3 are enabled, there are two pairs of microphones acting as NLMS filters: the first pair comprises Microphone 1 and 2 and the second pair comprises microphone 1 and 3. As in Figure 1, experiments [16] show that a voice close to the primary input (e.g. Microphone 1 in Figure 3) is enhanced and a voice close to the reference input (e.g. Microphone 2 or 3) is reduced. Therefore the driver's voice should be close to microphone 1 to be enhanced. In Figure 4, the circled area around microphone 1 is called the "noise canceller valid zone", where a desired voice is treated as a desired source but not noise.

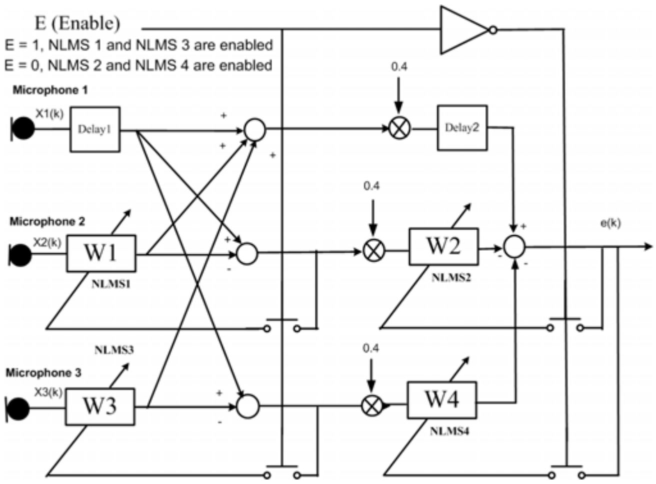


Fig. 3 Three-microphone noise canceller block diagram

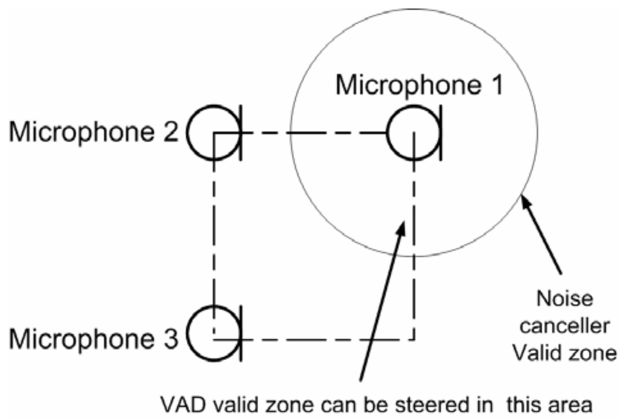


Fig. 4 Definition of a noise canceller valid zone around microphone1

While the 3-microphone VAD active zone is in the square area as shown in Figure 4, a 3-microphone noise canceller valid zone should has an intersection with the VAD valid zone.

III. Experiments

When there are two or more voices present simultaneously it becomes difficult to measure definitively the improvement in SNR of the desired voice. Therefore these series of experiments are specially designed to measure this improvement by timing the second unwanted voice appropriately. For all of the following experiments the distance between microphone 1 and microphone 2 is 50 cm and the distance between microphone 1 and microphone 3 is 70.7 cm.

Microphone 1.2 and 3 are omni-directional electret condenser microphones with the following specifications:

- Sensitivity: -62 +/- 3 dB
- Impedance: <2K Ohm
- Frequency Range: 50-12, 500Hz

A pre-amplifier and an anti-aliasing filter with a 5 kHz cutoff are used. The sample rate of Microphone 1, 2 and 3 is 11025 Hz. A desired source and a noise source were used in the test which was taken from loud-speakers. The enclosures are 21cm x 10 cm x 11 cm. The loudspeakers in the enclosure have a specification:

- 0.5W 3"; External Diameter: 3 inches;
- Frequency response, lower limit: 200Hz; Frequency response, upper limit: 6kHz;
- Impedance: 8 Ohms;

A PC sound card line output is used to drive this speaker. All results were recorded in a car in real time.

A. Experiment 1: The 3-microphone noise canceller in a 2 speech environment.

A 3-microphone noise canceller was tested as shown in Figure 5. When the driver's voice is presented the second voice is also present and is present after the driver's voice has stopped. In this way we can measure by how much the second voice has reduced (otherwise its reduced form overlaps the drivers voice and is difficult to measure). Since the driver's voice is located in the VAD valid zone, the 3-microphone VAD measures a desired voice and enables E (E = 1) as shown in Figure 3. When the driver's voice disappears, the 3-microphone VAD disables E (E = 0).

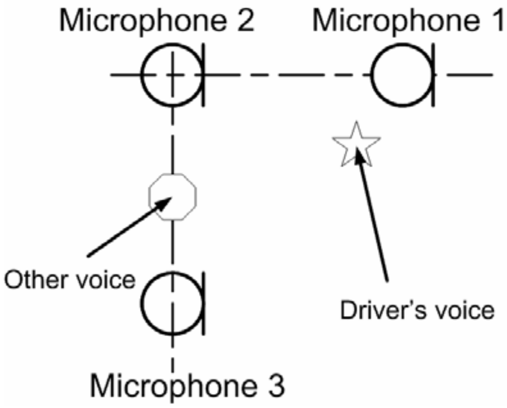


Fig. 5 The 3-microphone noise canceller in a speech and unwanted speech environment

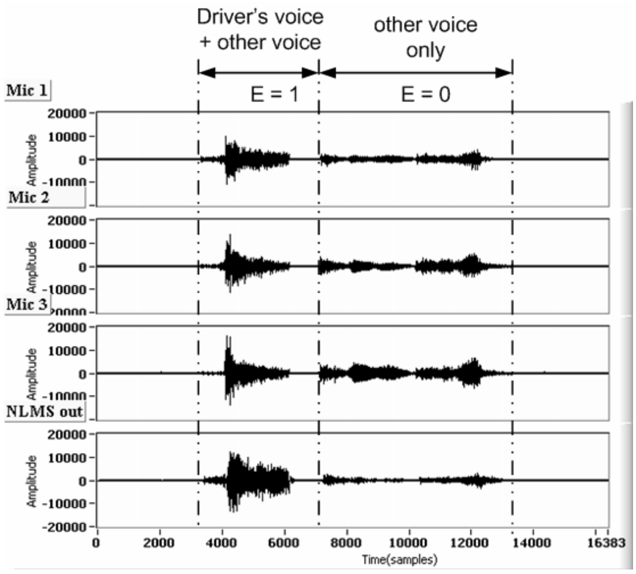


Fig. 6 Refers to Figure 3, driver's voice enabled VAD (E = 1) and then disabled VAD (E = 0)

The result as shown in Figure 6 clearly indicates that when the VAD is enabled (E=1) the driver's voice is enhanced and otherwise (when E=0) the other voice is reduced. A previous experiment [12] clearly shows the Signal to Noise Ratio (SNR) improvement of the 3-microphone noise canceller is more than 6 dB.

When both voices appear simultaneously and when E = 1, the driver's voice and other voice are overlapping and the result as shown in Figure 6 cannot clearly show (other than by listening tests) if the other voice is reduced significantly. Therefore, the following experiments investigate this in more detail.

B. Experiment 2: Definition of Noise canceller valid zone

Whilst E = 1 as shown in Figure 3, NLMS 1 and NLMS 3 are enabled by the VAD. In this experiment white noise comes via test points 7, 6, 5 and so on down to 1. The output waveform of Microphones 1, 2, 3 and the 3-microphone noise canceller error (output) are shown in Figure 8. The waveforms in Figure 8 are clearly indicating that the output of the 3-microphone noise canceller followed and enhanced voice from microphone 1 but microphone 2 and 3 had no effect as each microphone has a valid sensitivity distance of 25 cm. This result indicates a "noise canceller valid zone" of this 3-microphone noise canceller as shown in Figure 7.

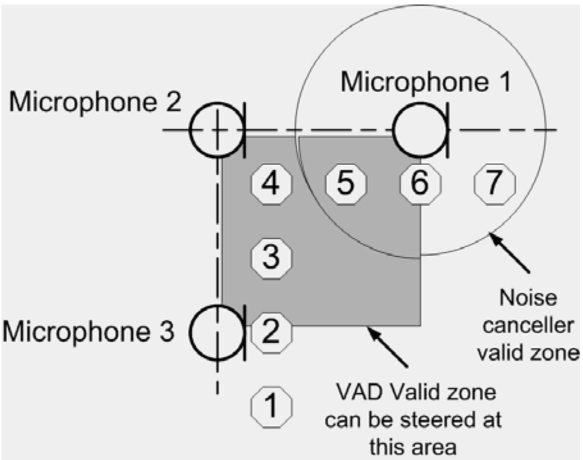


Fig. 7 A test for the noise canceller valid zone

The SNR calculation is defined here as:

$$SNR_i = 10\log_{10}\frac{Power\ of\ noise\ canceller\ output}{Power\ at\ microphone\ i} \quad i = 1,2,3 \qquad (7)$$

where the power of the noise at the noise canceller output and the power of the noise at microphone 1, 2 or 3 are total average power during the periods of block samples 1, 2 or 3 as shown in Figure 8. All results at the sample blocks are shown in Table 1. All block samples are 96ms second duration. The calculated SNR results are shown in Table 2. As shown in Table 2, only sample block 1 at microphone 1 is amplified and has 2.41 dB power. All other SNR(s) are attenuated since they are from outside of the noise canceller valid zone

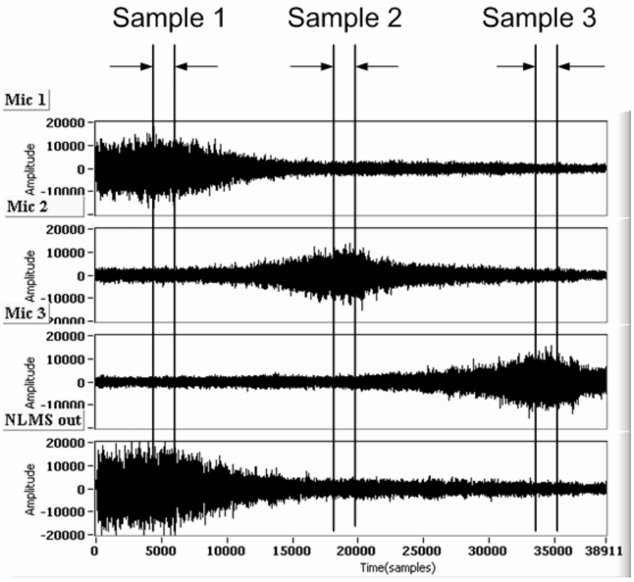


Fig. 8 White noise source testing waveforms

TABLE I
POWER OF MICROPHONES INPUTS AND NOISE CANCELLER'S OUTPUT

	Sample 1 (dB)	Sample 2 (dB)	Sample 3 (dB)
Microphone 1	-13.88	-26.95	-30.03
Microphone 2	-25.91	-15.64	-27.18
Microphone 3	-29.4	-28.04	-14.13
Output	-11.47	-24.15	-26.81

TABLE II
SNR RESULTS OF WHITE NOISE TEST

	SNR_1 (dB)	SNR_2 (dB)	SNR_3 (dB)
Microphone 1	2.41		
Microphone 2		-8.51	
Microphone 3			-12.6

C. Experiment 3: A single noise environment: Driver's voice and a second voice (stationary white noise)

As speech is easily investigated in a mixed graphic of speech plus white noise, a white noise signal is used as a second voice to test the 3-microphone noise canceller in a frequently moving noise sources environment. A 3-microphone VAD valid zone was steered by pre-defined time-difference of arrival (TDOA). So VAD valid zone can be moved towards microphone 1. A driver's voice shown in Figure 9 turns on the 3-microphone VAD whilst at the same time, a white noise source jumps from test points 1, 2, 3 and so on up to test point 18.

Figure 10 shows a test result when a driver's voice is in the VAD valid zone whilst E is held at unity by the VAD. A white noise source comes via test points 1 – 18 as depicted in Figure 9. Note from Fig 10 that the driver's voice has been enhanced but it is difficult to measure the reduction in noise. Therefore Figure 11 shows a test result with the Enable pin in Figure 3 manually enabled whilst white noise comes via test points 1, 2 ... and finally to test point 18 but with no desired speech present. The measurement of the dB noise reduction is now far easier than with Fig 10.

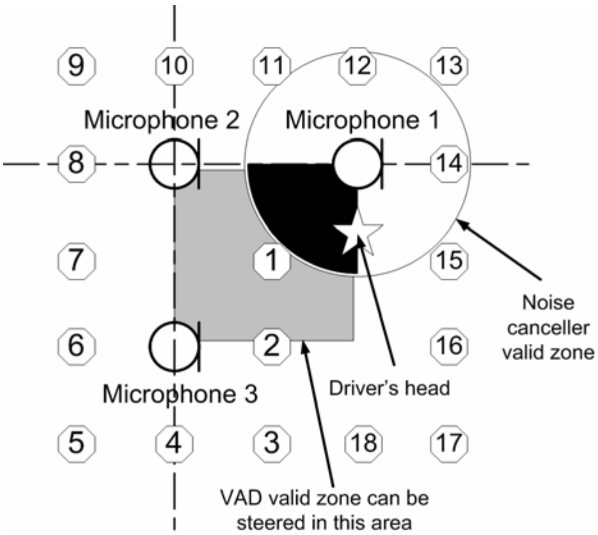


Fig. 9 Numbering the test points in a frequently moving noise sources environment (Driver's voice and a second voice)

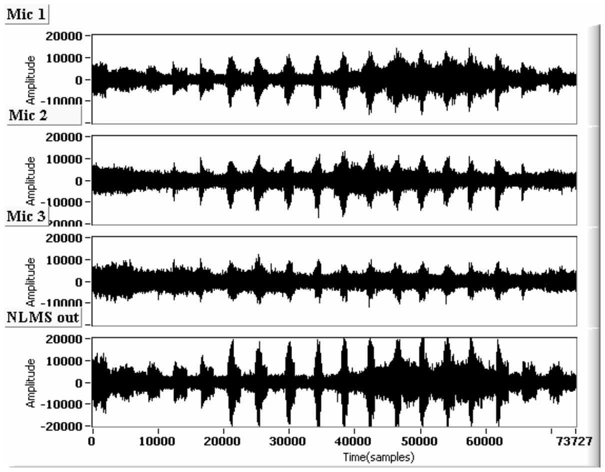


Fig. 10 Voice in the VAD valid zone activates the VAD, whilst a white noise source comes from test points 1, 2 and so on

The previous experiment clearly demonstrates the ability of the algorithm to cancel speech outside of the VAD valid zone for a solo voice and enhance voice from within the VAD valid zone as in Figure 10 the driver's voice is amplified by 6 dB [12]. As shown in Figure 12, when the driver's (desired) voice is detected and a second voice is presented from test point 1 - 18, the signal output of 3-microphone noise canceller follows the signal from Microphone 1 only. This means that any second voice present to Microphone 2 and 3 are ignored. Therefore, if the second voice is far enough away from Microphone 1 (e.g. outside of the noise canceller valid zone), the second voice is reduced. For example, in the case of a second voice moving from test point 12 to test point 7, the output of the 3-microphone noise canceller reduces the second voice by 8.5 dB.

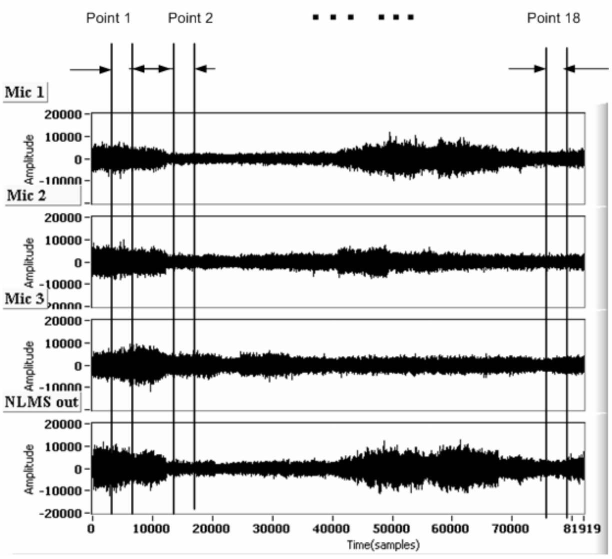


Fig. 11 While manually E = 1 (in Figure 3), no speech and white noise jumps from test points 1, 2 ... finally to 18

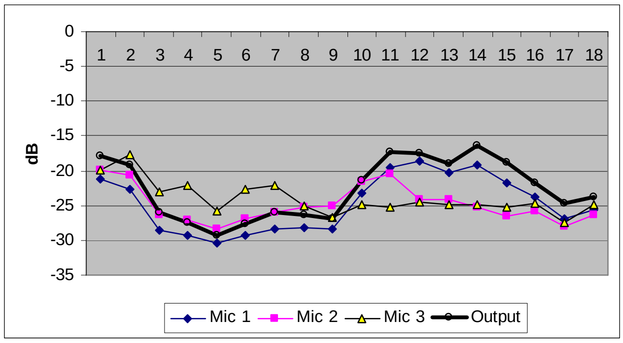


Fig. 12 Moving second voice (white noise) results

D. *Experiment 4: A single noise environment: Driver's voice and a second voice (speech)*

A 3-microphone VAD valid zone was steered by pre-defined time-difference of arrival (TDOA). So this VAD valid zone can be moved towards microphone 1. A driver's voice shown in Figure 13 turns on the 3-microphone VAD whilst at the same time, the second speech comes via test points 1, 2 and 3.

In Figure 14, from the top, the waveforms are Microphone 1, 2and 3 respectively. The waveform of the 3-microphone noise canceller output is shown at the bottom. Comparing input waveforms of Microphone 1 and the 3-microphone noise canceller output, we can clearly see that the 3-microphone noise canceller output follows the output of Microphone 1 and not microphone 2 and 3. Of course the second voice is still present but can easily be nullified by using the VAD.

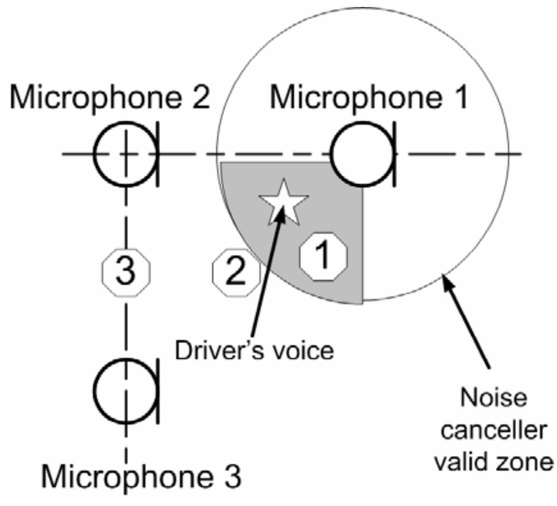


Fig. 13 Numbering the test points in a frequently moving noise sources environment (Driver's voice and a second voice)

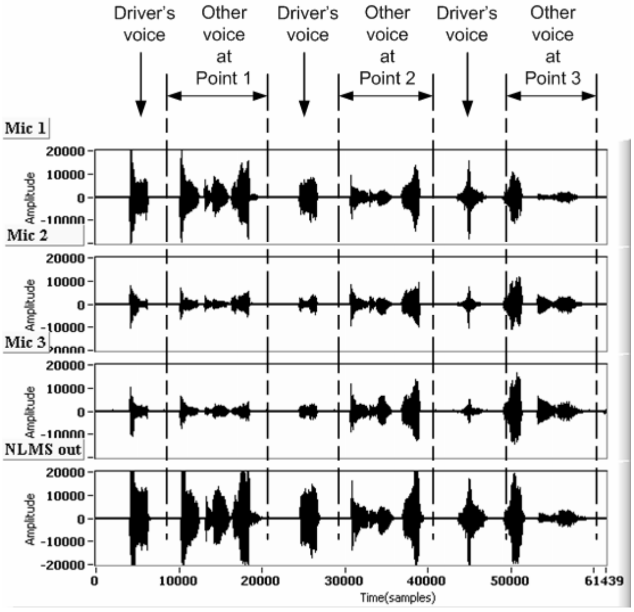


Fig. 14 while E = 1 (as in Figure 3), other speech arrives via test points 1, 2 and 3

IV. Conclusions

In a frequently moving noise sources environment (noise sources are coming from different locations but not always presented at the same time), the 3-microphone noise canceller with geometric VAD has the effect of canceling un-wanted speech or noise from outside of a VAD valid zone. As the same time, there is a 3-microphone noise canceller valid zone defined. In order to enhance desired speech and reduce noise(s), a desired voice should be in the intersection of the noise canceller valid zone and the VAD valid zone. Thus all noise is suppressed outside this intersected area. Experiments performed have verified the improvements given by this method in a real environment.

References

[1] M. Shozakai, S. Nakamura, and K. Shikano, "Robust speech recognition in car environments," presented at Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of the 1998 IEEE International Conference on, 1998.

[2] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *Antennas and Propagation, IEEE Transactions on [legacy, pre - 1988]*, vol. 30, pp. 27-34, 1982.

[3] B. Widrow and F. Luo, "Microphone arrays for hearing aids: An overview," *Speech Communication*, vol. 39, pp. 27-34, 2003.

[4] T. Nishiura, R. Gruhn, and S. Nakamura, "Collaborative steering of microphone array and video camera toward multi-lingual tele-conference through speech-to-speech translation," presented at Automatic Speech Recognition and Understanding, 2001. ASRU '01. IEEE Workshop on, 2001.

[5] S. Stergiopoulos and A. C. Dhanantwari, "Implementation of adaptive processing in integrated active-passive sonars with multi-dimensional arrays," presented at Advances in Digital Filtering and Signal Processing, 1998 IEEE Symposium on, 1998.

[6] G. W. Elko, "Microphone array systems for hands-free telecommunication," *Speech Communication*, vol. 22, pp. 229-240, 1996.

[7] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, J. Eugene Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, pp. 1692-1716, 1975.

[8] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *IRE Wescon Convention Record*, pp. 94-104, 1960.

[9] M. M. Goulding and J. S. Bird, "Speech enhancement for mobile telephony," *Vehicular Technology, IEEE Transactions on*, vol. 39, pp. 316-326, 1990.

[10] W. Armbruester, R. Czarnach, and P. Vary, "Adaptive Noise Cancellation with Reference Input - Possible Applications and Theoretical Limits," in *Signal Processing III: Theories and Applications*, I. T. Young, Ed.: Elsevier, 1986, pp. 391-394.

[11] H. Agaiby and T. J. Moir, "A robust word boundary detection algorithm with application to speech recognition," presented at Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on, 1997.

[12] Z. Qi and T. J. Moir, "An Automotive three-microphone Voice Activity Detector and noise canceller," presented at 2005 International Conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, 2005.

[13] G. Carter, C. Knapp, and A. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," *Audio and Electroacoustics, IEEE Transactions on*, vol. 21, pp. 337-344, 1973.

[14] S. Haykin, *Adaptive Filter Theory*, 4 ed: Prentice Hall, 2002.

[15] G. Barrault, M. H. Costa, J. C. M. Bermudez, and A. Lenzi, "A new analytical model for the NLMS algorithm," presented at Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on, 2005.

[16] C. Rulph, DSP applications using C and the TMS320C6x DSK: J. Wiley, 2002.

[17] W. Herbordt, Sound Capture for Human/machine Interfaces - Practical Aspects of Microphone Array Signal Processing: Springer-Verlag, 2005.

[18] W. Herbordt, T. Horiuchi, M. Fujimoto, T. Jitsuhiro, and S. Nakamura, "Hands-Free Speech Recognition and Communication on PDAS Using Microphone Array Technology," presented at Automatic Speech Recognition and Understanding, 2005 IEEE Workshop on, 2005.

[19] O. Hoshuyama and A. Sugiyama, "Robust Adaptive Beamforming," in *Microphone Arrays: Signal Processing Techniques and Applications (Digital Signal Processing)*, M. Brandstein and Ward, Eds.: Springer-Verlag, 2001.

[20] R. B. Wallace and R. A. Goubran, "Improved tracking adaptive noise canceler for nonstationary environments," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 40, pp. 700-703, 1992.

[21] R. B. Wallace and R. A. Goubran, "Noise cancellation using parallel adaptive filters," Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions on], vol. 39, pp. 239-243, 1992.

[22] D. Van Compernolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," presented at Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on, 1990.