A NEW DOA ESTIMATION METHOD USING A CIRCULAR MICROPHONE ARRAY

Amin Karbasi and Akihiko Sugiyama†

School of Comp. and Commun. Sciences, Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
†Common Platform Software Research Labs., NEC Corporation
Kawasaki 211-8666, JAPAN

ABSTRACT

This paper proposes a new DOA (direction of arrival) estimation method based on circular microphone array. For an arbitrary number of microphones, it is analytically shown that DOA estimation reduces to an efficient non-linear optimization problem. Simulation results demonstrate that deviation of the estimation error for 20 *and* 10 *dB SNR is smaller than* 0.7 *degree which is comparable to high resolution DOA estimation methods. A larger number of microphones provide a more omni-directional spatial resolution.*

1. INTRODUCTION

Direction of arrival (DOA) estimation of a sound source by using microphone arrays has been an active research topic since the early 1990's [1]. It has important applications in human computer interfaces such as video conferencing [2], speech enhancement and speech recognition [3]. The fundamental principle behind DOA estimation is to capture the phase information present in signals picked up by microphones. There are three main categories in DOA estimation methods. The first group is based on the output power of steered beamformer, such as MVB (maximum variance beamformer) and DSB (delay and sum beamformer) [4]. The multiple source localization problem can be resolved by those based on high-resolution subspace techniques, such as MUSIC and ESPRIT [5]. Simple and widely used methods for realtime source localization form a group of TDOA (the time delay of arrival) methods [6]. They utilize the TDOA between signals at a pair of microphones and the microphone arrangement information.

Generally, the accuracy of the estimated DOA is higher with a larger number of microphones, which naturally impose heavy computation. From this point of view, linear arrays with a small number of microphones have been studied intensively in DOA estimation [8]. However, it is basically not possible to discriminate the front and the rear by a linear array [7]. Moreover, the array size rapidly increases with the number of microphones. To avoid these problems, a circular microphone array [2] based on a TDOA method is desirable.

This paper proposes a new DOA estimation method based on TDOA with a circular omni-directional microphone array. Use of TDOA needs relatively small number of computations and a circular microphone array does not have the front-rear discrimination problem. In the following section, linear microphone arrays are reviewed. The proposed DOA estimation is developed in Section 3. Finally, Section 4 presents the performance of the proposed method by computer simulations.

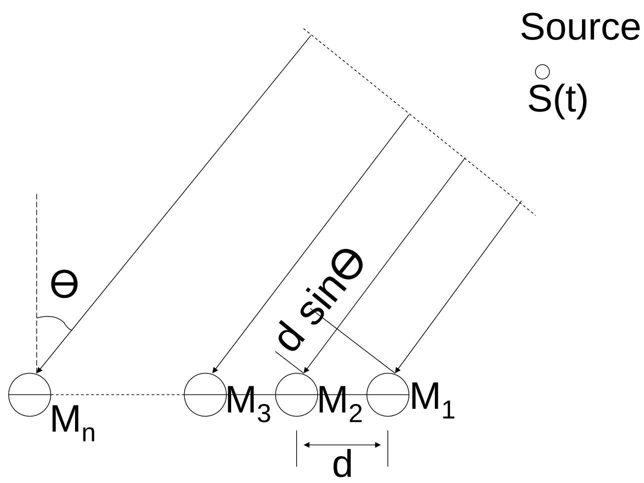

Figure 1: *Linear microphone array.*

2. UNIFORM LINEAR MICROPHONE ARRAY

Figure 1 shows an n-element uniform linear microphone array and a far-field sound source. The microphones are placed in a straight line with a uniform distance, d. The signal from the source reaches the microphones with different delays due to different distances that the sound wave has to travel. The time delay $\tau_{m_i m_j}$ between two signals $x_i(n)$ and $x_j(n)$ at microphones M_i and M_j is

$$\tau_{m_i m_j} = (i-j)\frac{d\sin\theta}{c}, \tag{1}$$

where c is the sound velocity. The signal at microphone i is given by

$$x_i(t) = s(t-\tau_{m_i m_1}) + n_i(t) \quad (1 \le i \le n). \tag{2}$$

where $n_i(t)$ is the noise at the microphone.

Equation 1 shows two important properties of uniform linear microphone array. First, the delay factor is a function of the source DOA θ. Thus, based on $\tau_{m_i m_j}$, the DOA θ could be obtained. Second, the directivity of a uniform linear microphone array is symmetrical about the array surface. It is why this method can not discriminate whether the speaker is located in the front or the rear (the front-rear problem).

3. DEVELOPMENT OF THE PROPOSED METHOD

Figure 2 shows a circular microphone array. In account of circulating around a circle, from here after the following convention is used: it is said that the microphones M_i and M_j are

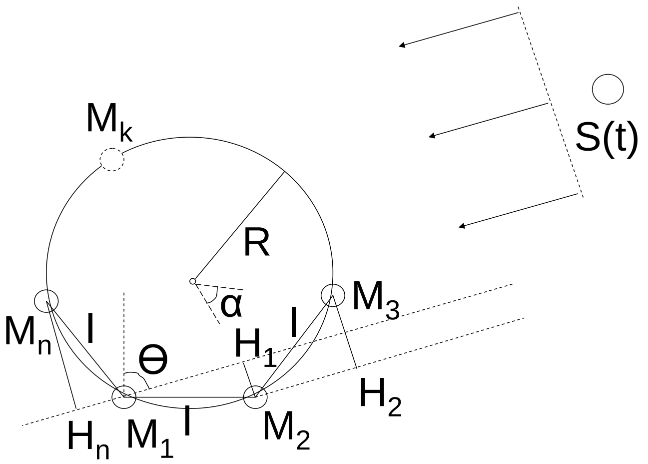

Figure 2: *Circular microphone array.*

one and the same if and only if $i \equiv j \pmod n$. Let us assume that l be the distance between the two adjacent microphones (e.g M_1M_2) and R the distance between a microphone to the array center o, *i.e.* the circumradius of the microphone array. Then, with n being the number of microphones, the following eqations hold.

$$\angle M_i o M_{i+1} = \frac{2\pi}{n} \quad (1 \le i \le n). \tag{3}$$

$$l = 2R\sin(\frac{\pi}{n}). \tag{4}$$

Note that in Eq. (3) by our convention, $(n+1)$-th microphone is nothing but the first one. The value of $\alpha = 2\pi/n$ is fixed in Fig. 2. A speaker in direction θ, which is observed with respect to the line perpendicular to $\overline{M1M2}$, generates speech $s(t)$. Therefore, the following equality can be obtained from Fig. 2:

$$M_iH_i = l\sin(\theta + (i-1)\alpha) \quad (1 \le i \le n). \tag{5}$$

According to Fig. 2, the relative delay between two signals at each pair of adjacent microphones is given by

$$\tau_{m_i m_{i-1}}(\theta) = l\sin(\theta + (i-1)\alpha)/c \quad (1 \le i \le n). \tag{6}$$

Thus, the signal $x_i(t)$ picked up by the i-th microphone is given by the following recursive equation

$$x_i(t) = x_{i-1}(t - \tau_{m_i m_{i-1}}) + n_i(t) \quad (1 \le i \le n), \tag{7}$$

where $n_i(t)$ is the ambient noise. Suppose that $\Phi^{(\omega)}_{m_i m_j}$ is the cross correlation in the frequency domain between $x_i(t)$ and $x_j(t)$ as

$$\Phi^{(\omega)}_{m_i m_j} = E[X_i(\omega)X_j^*(\omega)], \tag{8}$$

where $X_i(w)$ and $X_j(w)$ are the Fourier transform of $x_i(t)$ and $x_j(t)$, respectively and ω is the radial frequency. Therefore, the cross correlations between n signals of adjacent microphones are defined as follows:

$$\Phi^{(\omega)}_{m_i m_{i-1}}(\theta) = \Phi_{ss}(\omega)e^{-j\omega\tau_{m_i m_{i-1}}(\theta)} \quad (1 \le i \le n), \tag{9}$$

where $\Phi_{ss}(\omega)$ is the power spectral density of $s(n)$.

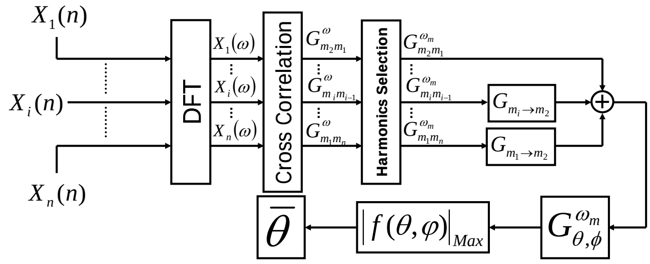

Figure 3: *Block diagram of the proposed method.*

Let us consider the difference $\tau_{m_i \rightharpoonup m_2}(\phi)$ in the delay between two cross correlations for a signal propagating from a direction ϕ.

$$\tau_{m_i \rightharpoonup m_2}(\phi) \triangleq \tau_{m_2 m_1}(\phi) - \tau_{m_i m_{i-1}}(\phi) \quad (1 \le i \le n). \tag{10}$$

Then, the following *phase rotation factors* are defined as

$$G^{(w)}_{m_i \rightharpoonup m_2}(\phi) \triangleq e^{-j\omega\tau_{m_i \rightharpoonup m_2}(\phi)} \quad (1 \le i \le n). \tag{11}$$

Using these phase rotation factors, the following *circular integrated cross spectrum* (CICS) is defined as

$$G^{(w)}_{\phi,\theta} \triangleq \sum_{i=1}^{n} G^{(w)}_{m_i \rightharpoonup m_2}(\phi) G^{(w)}_{m_i m_{i-1}}(\theta) \tag{12}$$

where

$$G^{(w)}_{m_i m_{i-1}}(\theta) = \frac{\Phi^{(w)}_{m_i m_{i-1}}(\theta)}{|\Phi^{(w)}_{m_i m_{i-1}}(\theta)|} \quad (1 \le i \le n). \tag{13}$$

Theorem 1. *In general, for any n-element circular microphone array, the CICS satisfies the following inequality*

$$\left|G^{(w)}_{\phi,\theta}\right| \le n.$$

(A proof of this theorem is given in Appendix B.)

Theorem 2. *For any n-element circular microphone array, the following equality is satisfied.*

$$\left|G^{(w)}_{\phi,\theta}\right| = n,$$

if and only if $\theta = \phi$.

(A proof of this theorem is given in Appendix C.)

In other words, Theorem 2 states that the DOA estimation problem can be reduced to an optimization problem. The value which maximizes the amplitude of CICS is the estimated value of DOA θ. In addition, the proposed method can discriminate the signal from omni-direction, because the preceding theorem holds for $0 \le \theta \le 2\pi$. The block diagram of the proposed method is shown in Fig. 3.

3.1 Noise Sensitivity

According to Eq. (6), the DOA θ is given by

$$\theta = \arcsin\left(\frac{\tau_{m_i m_{i-1}} \cdot c}{l}\right) - (i-1)\alpha \quad (1 \le i \le n), \tag{14}$$

which means any error on τ causes an error on θ. Due to noise in the real environments, the estimated value of τ always contains an error. In order to measure the noise sensitivity of DOA estimation, the *noise robustness factor (NRF)* is defined as

$$I(\theta) = \left| \frac{d\theta}{d\hat{\tau}} \right|, \tag{15}$$

where $I(\theta)$ indicates NRF and $\hat{\tau}$ is the estimated value of τ. In the case of two-microphone array, NRF is reduced to

$$I_2(\theta) = |\cos\theta|. \tag{16}$$

When the array has n microphones, the NRF is the average of n NRFs from n adjacent pairs of microphones as

$$I_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \left| \cos\left(\theta - \frac{(i-1)2\pi}{n}\beta\right) \right|. \tag{17}$$

3.2 Suppressing Effects on Reverberation

The integrated use of CICS in the proposed method is expected to suppress the influence of reverberation. Reverberation is known to be spatially deffuse due to multiple reflection paths. On the other hand, each component of CICS is generated by the data acquired at a different spatial position. From these facts, the reverberation components in each term of (12) are mutually uncorrelated. Therefore, the CICS can be expected to have an anti-reverberation nature.

4. EVALUATIONS

4.1 Conditions

The proposed DOA estimation method was evaluated by numerical simulations. The microphones were virtually located at the vertices of a regular polygon as shown in Fig. 2. Two real signals, male voice and female voice, were used as the source signals. The microphone array input signals were generated by delaying the source signal with an appropriate samples according to θ and mixed with an additive white Gaussian noise (AWGN). The effect of reverberation was not considered. It is important to note that the maximum distance between each pair of adjacent microphones is determined by the spatial sampling frequency.

$$D_{max} = \frac{c}{f_s}, \tag{18}$$

where D_{max} is the maximum distance, c is the sound velocity and f_s is the sampling frequency. In the case of 16 kHz sampling, the maximum distance between two microphones is about 0.021 m. Parameters of simulation are summarized in Table 1.

4.2 Circular Integrated Cross Spectrum

The behavior of CICS gives us some useful information. Figure 4 shows the amplitude of CICS for differen number of microphones, *i.e.* 13, 11, 9, 7 5, and 3 when the source signal was virtually fixed at a DOA of 100 degrees. A larger number of microphones lead to a sharp peak in CICS, resulting in more accurate DOA estimation.

Table 1: *Parameters of Simulation.*

Sound Velocity	340 m/s
Sampling Frequency	16000 Hz
Input SNR	20 dB
Data Length	1024 Samples
Window	Hamming
Overlap	50%
l	0.021 m
Number of Mics	8

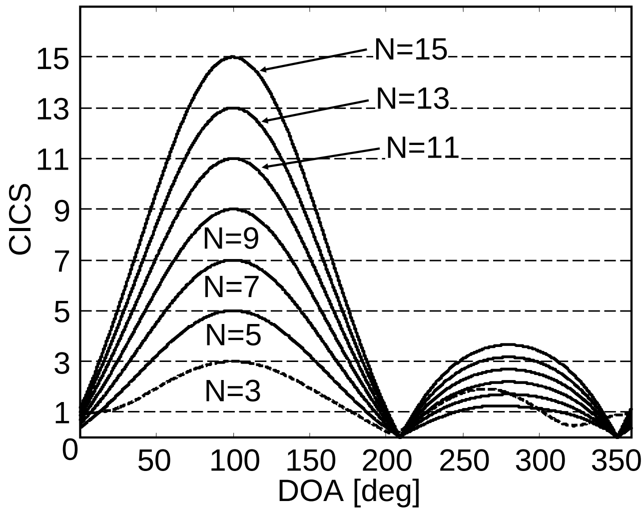

Figure 4: *Circular integrated cross spectrum.*

4.3 DOA Estimation in Noisy Environment

Fig. 5 depicts the DOA estimation results for SNRs (signal-to-noise ratio) of ∞ dB, 20 dB, 10 dB and 0 dB. The source was virtually fixed at a DOA of 60 degrees and parameters are from Table 1. Fig. 5 naturally shows more widely distributed results of DOA estimation for lower SNRs. However, an SNR of 20 dB or higher provides an accuracy smaller than 1 degree.

4.4 Noise Robustness Factor

Figure 6 depicts the noise robustness factor for an array of 2, 4, 6, and 8 microphones. As Fig. 6 illustrates when the number of microphones increases, the The NRF pattern with a larger number of microphones becomes omni-directional and approaches a circle. The DOA estimation is robust independent of the actual DOA.

4.5 Spatial Resolution

The spatial resolution was evaluated by the deviation of estimation error (DEE), which is given by

$$DEE = \sqrt{\frac{1}{K}\sum_{i=1}^{K} |\ \hat{\theta}_i - \theta\ |^2}, \tag{19}$$

where $\hat{\theta}_i$ is the estimated DOA in the i-th iteration, and θ is the true DOA. Parameters were basically the same as in Table 1. K and SNRs were set to 100, 20 dB, 10 dB, and 0

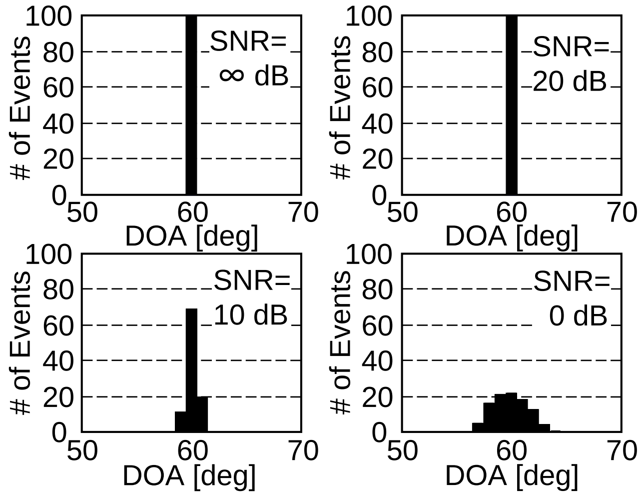

Figure 5: *The effect of SNR on DOA.*

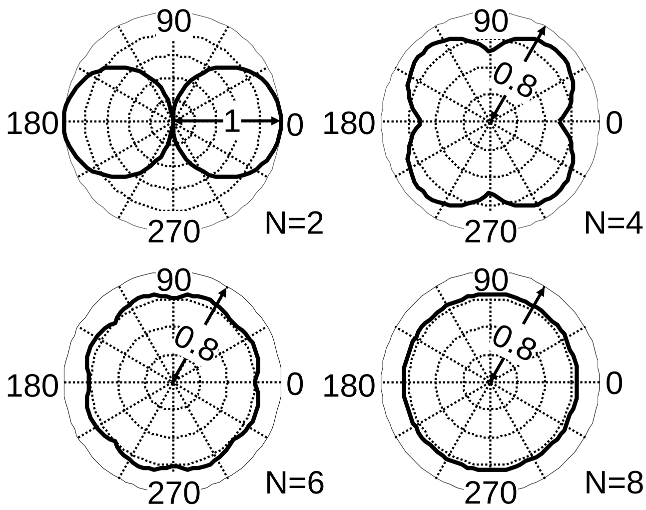

Figure 6: *Noise robustness factor.*

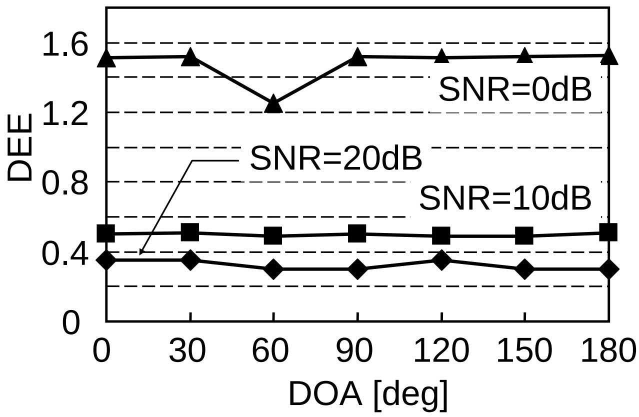

Figure 7: *Deviation of estimation error in different SNRs*

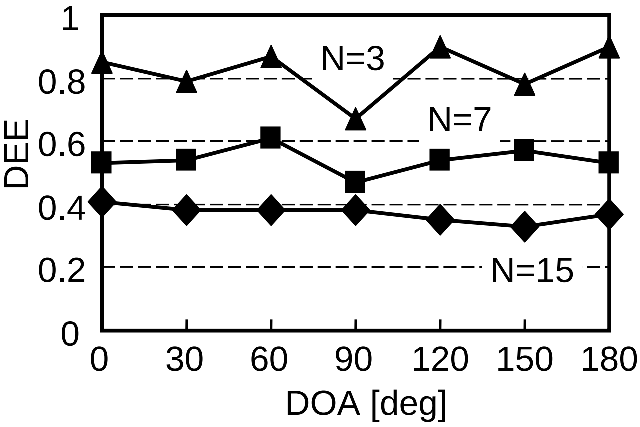

Figure 8: *Resolution for different number of microphones.*

dB. Fig. 7 indicates that the DEE in 20 and 10 dB SNR is less than 1 degree which is better than high resolution DOA methods like MUSIC [5].

4.6 Number of Microphones and Resolution

Fig. 8 illustrates the DEE for 3, 5, and 7 microphones with DOAs at an SNR of 10 dB. All parameters did not change from previous evaluations. It is apparent that a larger number of microphones provides amaller DEE, leading to higher accuracy.

5. CONCLUSION

A new DOA estimation method based on circular microphone array has been presented. It has been analytically shown that DOA estimation reduces to a non-linear optimization problem leading to computationally efficient implementation. There is no restriction on the number of microphones. An omni-directional spatial resolution has been demonstrated with a sufficiently large number of microphones. The deviation of the estimation error has reached a comparable value to that of high resolution DOA estimation methods.

REFERENCES

[1] M. Brandstein and H. Silverman, ”A practical methodology for speech localization with microphone arrays,” Technical Report, Brown University, November 13, 1996.

[2] S. Birchfield and D. Gillmor, “Acoustic source direction by hemisphere sampling,” Proc. of ICASSP2001, pp.3053–3056, May 2001.

[3] M. S. Brandstein and S. M. Griebel, “Nonlinear, model-based microphone array speech enhancement,” Acoustic Signal Processing for Telecommunications, Kluwer Academic Publishers, 2000.

[4] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, “Microphone Arrays, Signal Processing Techniques and Applications,” Springer Verlag, 2001.

[5] R. Roy and T. Kailath, “ESPRIT – estimation of signal parameters via rotational invariance techniques,” IEEE Trans. ASSP, vol. 37, no. 7, pp. 984・95, July 1989.

[6] C. H. Knapp, G. C. Carter, ”The generalized correla-

tion method for estimation of time delay," IEEE. Trans. ASSP Vol.24, pp. 320–327, Aug 1976.
"Robust Localization in Reverberant

[7] M. Sato, A. Sugiyama, O. Houshuyama, N. Yamashita, Y. Fujita, "Near-field sound-source localization based on a signed binary code," IEICE Trans. Fundamentals, vol. E88, No. 8, August 2005.

[8] J. Kleban, "Combined acoustic and visual processing for video conferencing systems," MS Thesis, Rutgers University, 2000.

[9] M. Abramowitz, and I. A. Stegun (Eds.), "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables," 9th printing. New York: Dover, p. 79, 1972.

A. SOME MATHEMATICAL OBSERVATIONS

Lemma 3. *[9] Suppose that $z_1, z_2, \ldots, z_n$ are complex numbers. In general,*

$$|z_1 + z_2 + \ldots + z_n| \le |z_1| + |z_2| + \cdots + |z_n|, \tag{20}$$

and the equality is satisfied if and only if

$$\arg(z_1) = \arg(z_2) = \cdots = \arg(z_n).$$

B. DERIVATION OF THEOREM 1

Proof. From (12), CICS is

$$\begin{aligned} G^{(w)}_{\phi,\theta} &\triangleq \sum_{i=1}^{n} G^{(w)}_{m_i \rightharpoonup m_2}(\phi) G^{(w)}_{m_i m_{i-1}}(\theta), \\ &= \sum_{i=1}^{n} e^{-j\omega\tau_{m_i \rightharpoonup m_2}(\phi)} e^{-j\omega\tau_{m_i m_{i-1}}(\theta)}, \\ &\le \sum_{i=1}^{n} \left| e^{-j\omega\tau_{m_i \rightharpoonup m_2}(\phi)} e^{-j\omega\tau_{m_i m_{i-1}}(\theta)} \right|, \qquad (21) \\ &= n. \end{aligned}$$

In (21), Lemma 3 is used. □

C. DERIVATION OF THEOREM 2

Proof. For proving sufficiency, let us assume that $\theta = \phi$. Then,

$$\arg\left(G^{(w)}_{m_i \rightharpoonup m_2}(\phi) G^{(w)}_{m_i m_{i-1}}(\theta)\right) = \omega \frac{l \sin\theta}{c} \qquad (1 \le i \le n). \tag{22}$$

Using Lemma 3, it is clear that equality will be satisfied.

For proving necessity, by Lemma 3 it can be supposed that all individual terms in the summation are in phase, i.e. have the same complex argument. The goal is to show that $\theta = \phi$. Thus,

$$\arg\left(G^{(w)}_{m_1 \rightharpoonup m_2}(\phi) G^{(w)}_{m_1 m_n}(\theta)\right) = \arg\left(G^{(w)}_{m_3 \rightharpoonup m_2}(\phi) G^{(w)}_{m_3 m_2}(\theta)\right). \tag{23}$$

By some simplifications, the following equations can be written

$$-\tau_{m_3 m_2}(\phi) + \tau_{m_3 m_2}(\theta) = -\tau_{m_1 m_n}(\phi) + \tau_{m_1 m_n}(\theta), \tag{24}$$

$$\sin\left(\theta + \frac{2\pi}{n}\right) - \sin\left(\theta - \frac{2\pi}{n}\right) = \sin\left(\phi + \frac{2\pi}{n}\right) - \sin\left(\phi - \frac{2\pi}{n}\right), \tag{25}$$

$$\cos(\theta) = \cos(\phi). \tag{26}$$

Therefore, the following relation between θ and ϕ will be obtained

$$\phi = \theta \quad \text{or} \quad \phi = -\theta. \tag{27}$$

Let us use again Lemma 3 and write down the following equality

$$\arg\left(G^{(w)}_{m_1 \rightharpoonup m_2}(\phi) G^{(w)}_{m_1 m_n}(\theta)\right) = \arg\left(G^{(w)}_{m_3 \rightharpoonup m_2}(\phi) G^{(w)}_{m_3 m_2}(\theta)\right). \tag{28}$$

After a number of simplifications

$$\tau_{m_2 m_1}(\phi) - \tau_{m_3 m_2}(\phi) = \tau_{m_2 m_1}(\theta) + \tau_{m_3 m_2}(\theta), \tag{29}$$

$$\sin(\phi) - \sin\left(\phi + \frac{2\pi}{n}\right) = \sin(\theta) - \sin\left(\theta + \frac{2\pi}{n}\right), \tag{30}$$

$$\cos\left(\phi + \frac{\pi}{n}\right) = \cos\left(\theta + \frac{\pi}{n}\right). \tag{31}$$

Thus, another relation between θ and ϕ will be obtained.

$$\phi = \theta \quad \text{or} \quad \phi + \frac{\pi}{n} = -\theta - \frac{\pi}{n}. \tag{32}$$

By combining (27), (32) and the fact that the number of microphones is larger than two ($n \ge 3$), it will be clear that ϕ is equal to θ. □