Proceedings of 2008 CACS International Automatic Control Conference
National Cheng Kung University, Tainan, Taiwan, Nov. 21-23, 2008

An Approach to Semi-spherical Microphone Array Based Sound Localization System

Ming-Yuan Shieh*, Chin-Chien Chen, Chien-Yuan Chen

Department of Electrical Engineering,
Southern Taiwan University
Tainan County, Taiwan
**e-mail: myshieh@mail.stut.edu.tw*

***Abstract*—The paper proposes a semi-spherical microphone array based sound localization system for a service robot. The hardware of the proposed system basically contains 12 capacitor microphones disposed in two layers on the semi-sphere of 19 cm diameter. It aims to estimate the degree relationship between the main speaker and the robot and provides the robot useful information for more effective human-robot interactions. The proposed system can determine the location of the voice according to energy information between the main speaker and robot not only in normal environment but also in blatant and/or reverberative space. The experimental results show that the proposed system has obtained satisfactory recognition efficiency, moreover, raised the robotic friendliness and adaptability.**

Keywords*—Semi-spherical microphone array, Sound localization system, Service robot, Human-robot interaction.

I. INTRODUCTION

Sound is the most native medium for communication between humans and nature. How to detect where the sounds are and/or come from is not difficult for a normal human. However, the same problem but always perplexes artificial auditory systems especially in sound localization. The estimation and tracking problems of sound source have discussed and studied widely in the past no matter in sonar, radar, telemetry, communication, security, and defense strategy. How to detect the location of sound source is still obscure, however, there are several effective methods [1-4] proposed that solved problems such as signal filtering, anti-noise, beam former, cross correlation.

Since the ear is very important for human perceiving environmental signals, the significance of the acoustic system is the same for robot. How to let a robot have abilities of vision, hearing, language and expression becomes essential. A fusion of these perceptions is the main task of robotic design.

In this paper, the authors focus on the problem of the localization of the sound source. It needs to consider those factors such as environmental noise, reverberation, echo and reflection. To design microphone array to suppress the environmental noise in the implementation is an effective way. Besides, to adopt suitable filters to reduce the high-frequent echo and reflection can always get satisfactory results. Although, if such schemes are combined with those algorithms having compensatory and suppressant effects, the performance of signal process will be improved, the system becomes complex and obscure simultaneously.

The ways to estimate the localization of the sound source mostly use the well-defined sound modules whatever in frequency or time domains. For those defined in frequency domain, the periodicity of the received sound signal is considered being for improving frequency spectrum. The ways in time domain include that estimates the cycle of vocal shake at the time of the ends. However, most results [5, 6] in recent studies of speaker sound localization show that their proposed schemes can only distinguish the possible location of the speaker with the error of 15 to 30 degrees. The main error lies in these schemes aim to recover the delay relationship to determine sound localization, but in fact, only very small time delay occurred when the sound reaches any pair of neighboring microphones.

Most studies in microphone array adopt plane type rather than three-dimension (3D) type because of their different system complexities. However, a plane type of microphone array just can detect the direction of the sound in a plane, but can not estimate the actual localization of the sound in 3D space. Therefore, some approaches [7-9] have focused on the design of 3D microphone array, but unfortunately, their results almost still are simulation not experimental ones. It indicates that the real applications of similar types of 3D microphone array still need more efforts in the future. Among proposed 3D types of microphone array, we think the spherical type is the most effective one. Moreover, if consider simplification, the semi-spherical type is possibly a well choice.

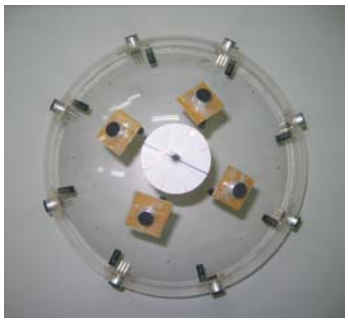
Since human and many animals have exquisite acoustic apparatus which provides clear and accurate perception. However, the microphone arrays or other artificial acoustic systems developed until nowadays are still not useful. For example, the authors propose an artificial hear who intend to simulates human hear construction, however, the proposed system always fails in detecting the sound direction. It usually can not know whether the sound comes from the front or the rear. In order to resolve such problem, in this paper, we adopt a semi-spherical microphone array to receive all directions of the sound around the robot, and a 3D sound localization system to conclude the exact location of the main speaker.

The proposed paper aims to design a configuration of semi-spherical microphone array, the sound detection circuit, and the FSL system. The authors intend to adopt

simple but useful methods without any complex frequency computation. The resultant inferred from the 3D sound localization system will be an exact direction rather than a rough possible area. It is noted that the experimental results demonstrate the feasibility of the proposed scheme.

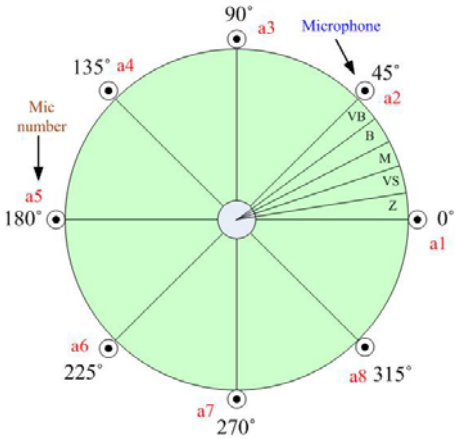
II. 3D Sound Localization System

Fig. 1 introduces the proposed configuration of the semi-spherical microphone array. In the disposal, there consist of 12 capacitor microphones disposed in two layers on the semi-sphere of 19 cm diameter. In which, the upper layer disposes four microphones at every 90 degrees, and the lower layer disposes eight ones at every 45 degrees. The hardware can be mounted upon the head of the robot to determine the 3D relative locality relationship between the sound and the robot. Since the disposal could receive all the sounds come from any direction, it provides omnibearing detection to determine the exact localization of the signal. Based on this data, the robot can perform face-to-face conversations, direct interaction, or/and objective searching.



(a) Top view

(b) Side view



(c) the disposal of microphones in the lower layer

Fig. 1 The configuration of the semi-spherical microphone array

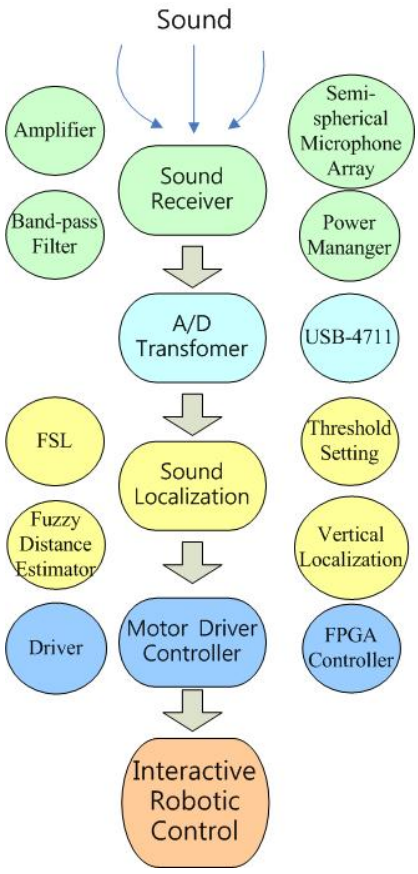


Fig. 2 The system flowchart of the proposed sound localization system

However, there is noise or background existed in the signal of sound. One needs to eliminate them from the signal before signal recognition processes. It is fortunate that the amplitude of the energy of the sound is a simple distinguishable property, if the value is larger than the threshold then the sound must be the voice else not.

Because of the sensitive response of capacitance microphones, to design a pre-filter for signal processes is necessary. In this paper, we adopt a 2^{nd} order amplifier and a band pass filter to improve the original sound. In which, the gain of the amplifier is designed as 100, and the cutoff frequencies of high pass and low pass filters are chosen as 20 Hz and 7 Hz respectively.

In order to reduce the complexity of computation, every acquirement of sound will be executed per 500 ms. Such a acquirement just represents a sound frame. It helps one could directly analyze the signal rather than pre-segment the sound. When the amplitude of a voice or sound is larger than the threshold, the signal will be recorded as a phonetic command. However, in real environments, there are numerous conditions needed to be considered. For example, one has to notice that the disposal of ambient objects, the pattern of the room and the location of microphone arrays all affect the detection of sound. Therefore, in order to reduce the influence of reflective noise, one could only analyze the first data whose amplitude is larger than the threshold. After several tests in our study, the resultant threshold could be chosen as 0.5V.

Table 1. The fuzzy rules of the FSDE

V2 \ V1	VB	B	M	S	VS
VB	C	F	VF	VF	VF
B	N	C	F	VF	VF
M	VN	N	C	F	VF
S	VN	VN	N	C	F
VS	VN	VN	VN	N	C

F: far, VF: very far, C: centered, N: near, VN: very near

The system flowchart of the proposed sound localization system is shown in Fig. 2. It is composed of four subsystems, such as sound receiver, analog to digital (A/D) transformer, sound localization estimator, motor driver controller. In which, every subsystem play individual functions but in a series of steps to estimate accurate sound localization and then to control the robot performing assigned tasks. It is worthy noted that the adopted A/D transformer (USB-4711) is manufactured by Advantech Co., Ltd. The device provides 16 analog input ports, 2 analog output ports, and 8 digital I/O ports effectively for signal acquirement and transform.

In proposed 3D sound localization system, there are three subsystems. The first is the fuzzy sound direction estimator (FSDE), the next is the fuzzy vertical localization estimator (FVLE), and the last is the fuzzy distance estimator (FDE).

A. Fuzzy sound direction estimator (FSDE)

The FSDE considers the two of whose energy summation are the maximal two among all microphone signals as inputs. The maximal one is named as V_1, whose reference angle is denoted as θ_1. The next is named as V_2, whose reference angle is denoted as θ_2. Assume that the total reference angle is named as θ_{ref}, which can be determined by (1). It is noted that the reference angle of which gets the maximal average of sound energy dominates the total reference angle.

The output of the FSDE is the inferred direction of the sound and named as θ_{FSDE}. It represents the angular correction from the direction where occurs the maximal energy of sound. One can follow those fuzzy rules tabulated in Table 1. In the rule table, the "V1" means the variable which has the maximal amplitude among all detected sounds, and the "V2" denotes the second big. If V1 is "B" and V2 is "M," it means the direction of the detected sound near (N) the direction of "V1". Then, the resultant angle θ_{Rot} will be equal to the reference angle θ_{ref} adds or subtracts the inferred angle θ_{FSDE}. The relationship could be described by (2).

$$\theta_{ref} = \theta_1 \tag{1}$$

$$\begin{cases} \theta_{Rot} = \theta_{ref} + \theta_{FSDE} & , when\ \theta_1 < \theta_2 \\ \theta_{Rot} = \theta_{ref} - \theta_{FSDE} & , when\ \theta_1 > \theta_2 \end{cases} \tag{2}$$

B. Fuzzy vertical localization estimator (FVLE)

The FVLE aims to estimate the vertical height of the sound's location over the bottom of the semi-spherical microphone array. As shown in Fig. 3, one can compare the average of the most three voltages of the microphones in the lower layer with the voltage of the responding microphone in the upper layer. If the one in the lower layer is larger than the one in the upper layer, one can say that the vertical location of the estimated sound will be lower, else means higher.

If one considers the average of maximal three or two in the lower or the upper layer as the input. The one in the lower layer is named as V_L, the one in the upper layer is named as V_U. The output of the FVLE is the inferred vertical location of the sound and named as H_{FVLE}. One can follow those fuzzy rules tabulated in Table 2. In the rule table, the "VL" means the variable which has the average of the maximal three amplitudes among all detected sounds in the lower layer, and the "VH" denotes the average of the responding two microphones in the upper layer. For example, if VL is "B" and VH is "M," it means the vertical location of the detected sound near the lower layer (NL).

C. Fuzzy distance estimator (FDE)

Besides, in this paper, a fuzzy distance estimator (FDE) is proposed to estimate the distance between the detected sound and the semi-spherical microphone array. Consider the most voltage as the reference voltage V_x when a defined sound is made of the distance 15 cm away from the semi-spherical microphone array sound. Then, we make the same sound at different distances from 50 cm to 250 cm respectively. In every test, we record the most voltage (V_I) according to every distance between the sound and the microphone array.

Table 2. The fuzzy rules of the FVLE

VH \ VL	VB	B	M	S	VS
VB	C	NH	H	H	H
B	NL	C	NH	H	H
M	L	NL	C	NH	H
S	L	L	NL	C	NH
VS	L	L	L	NL	C

H: higher, NH: near higher, C: centered, NL: near lower, L: lower

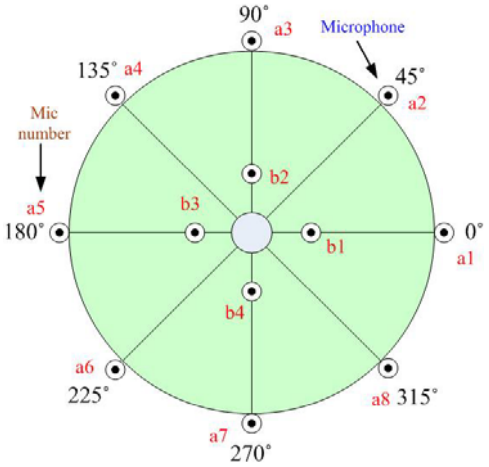


Fig. 3 The disposals of microphones in the lower (a1-a8) and upper (b1-b4) layers

Table 3. The fuzzy rule table of the FDE

V_I	VH	H	A	L	VL
d	VN	N	M	F	VF

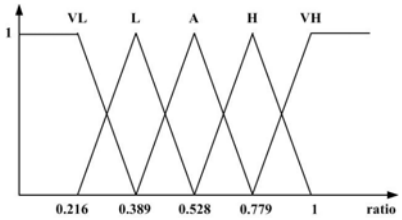


Fig. 4 The membership function of the input (V_I) in the FDE

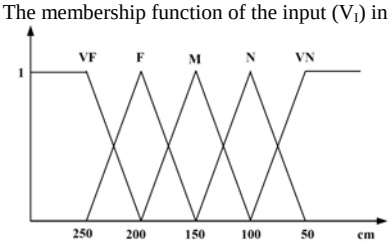


Fig. 5 The membership function of the output (d) in the FDE

The ratio of the voltage V_I over the reference voltage V_X will be as the partition points of fuzzy subsets as shown in Fig. 4. By the same definition, the output membership function can be defined as Fig. 5, which indicates the sound departed from the microphone array every 50 cm. Table. 3 display the responding fuzzy rules. It indicates that the estimated distance will be proportioned to the ratio of voltage.

III. Experimental Results

In order to detect the sounds come from any side of the robot, a semi-spherical configuration of 12 microphones array as shown in Fig. 6 is proposed in this paper. Every microphone is disposed with an optimal interval of 5 cm depart from neighbors after many times experiments. All microphones are connected in series to the sound detection circuit which aims to perform signal processes. It handles not only the amplification of the gain 100 but also the band-pass filtering with the frequency between 20~7k Hz. The proceeded sound will be divided into speech, silence, or noised segments by the algorithm of detecting sound frame. Briefly, one can observe the energy curve of the sound, and then get that the segment is a speech in case of the amplitude higher than the threshold, and a noised or silence sound else.

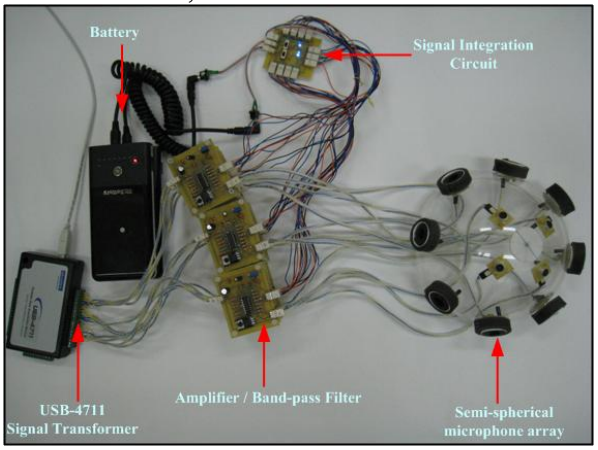


Fig. 6 The actual components of the proposed system

All experiments in our studies are took place in a laboratory of 9.8m*7.6m as shown in Fig. 7. In such room, there are six long tables, four personal computers, three cabinets, a door, a window of two wickets, and an air conditioner. It is noted that the red marks in Fig. 7 consist of a high cabinet and the air conditioner, which are the main noise makers. Besides, the yellow denotes the proposed system which is placed at the center of the room because that it will obtain better performance at such position.

Fig. 8 shows the wave pattern of a ring from a mobile phone which is used as the basic test sound. It is recorded when the mobile phone is apart from the microphone arrays about 50 cm.

First, we intend to test what effect generated when the sound in front (0^o) or in rear (180^o) of the semi-spherical microphone array. Table 6 illustrates the experimental results in 0^o and 180^o respectively. From these data of that average errors are just only 3.3^o and 3.2^o, one can find that the proposed system can really distinguish what location the sound is.

Next, we would like to test whether the distance between the sound and the proposed system could be estimated. The ring of a mobile phone is adopted as the reference sound. Table 5 indicates the results of 10 tests for two different sound distances estimation. It shows both the detection in case of the sound away from the proposed system 100cm or 200cm has average errors within 4.6%, it is near accurate. However, there are still some tests with errors of larger than 5.5%, or even up to 7.6%. It means that the once distance estimation provided by the proposed FDE is not convincible unless the average of over 10 times tests.

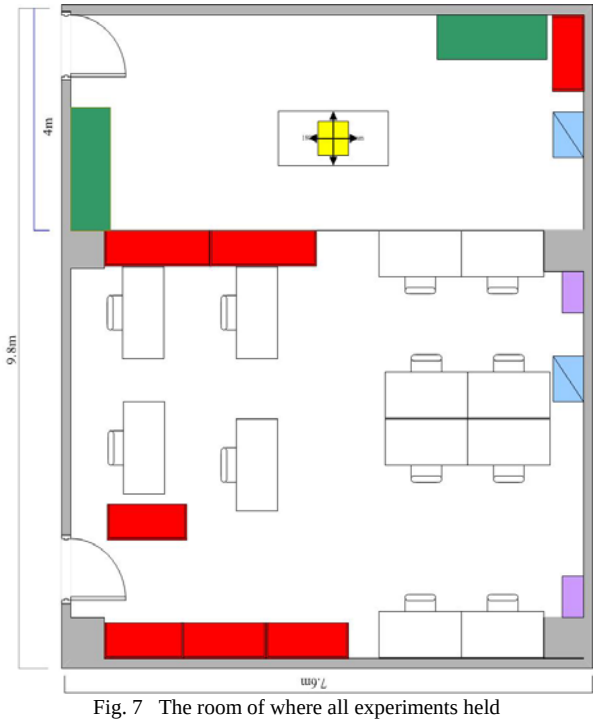


Fig. 7 The room of where all experiments held

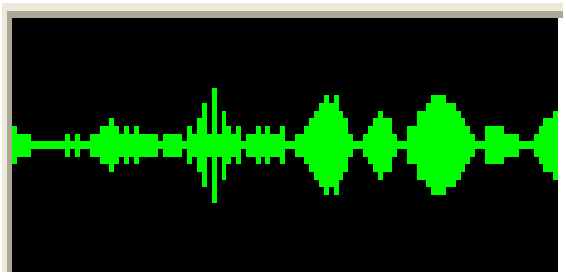


Fig. 8 The wave pattern of a ring from a mobile phone

Finally, assume that the sound is disposed in 45° 100cm away from the tested systems, in comparison with the proposed semi-spherical and plane types of microphone array, Fig. 9 shows that the former (the average of errors is almost 3°) is little better than the later (the average of errors is almost 4.2°) in plane direction estimation of sound. However, it is noted that the proposed system is good at vertical location and distance estimations, but the plane type is more inferior to those in these two aspects.

IV. Conclusions

The paper aims to design the semi-spherical configuration of 12 microphones array and the sound localization system composed of three fuzzy subsystems to estimate the plane direction, the vertical location, and the distance of the given sounds. A simple but useful method is adopted without any complex frequency computation. From the experimental results, it is seen that the proposed scheme has satisfactory recognition results. Moreover, the proposed system is able to be applied as the acoustic system of mobile robots. It results in that the robot can face to the main speaker based on which 3D location concluded by the proposed sound localization system.

Acknowledgment

This work was supported by the National Science Council, Taiwan, under grant number NSC96-2221-E-218-039-MY3.

References

[1] I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. on Speech and Audio Processing*, Vol. 11, Issue 6, pp. 709-716, Nov. 2003.

[2] B. Yegnanarayana, S.R.M. Prasanna, R. Duraiswami, and D. Zotkin, "Processing of reverberant speech for time-delay estimation," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, Issue 6, pp. 1110-1118, Nov. 2005.

[3] D. Mahmoudi, "Speech source localization using a multi-resolution technique," in *Proceedings of 1998 IEEE International Conference of Interactive Voice Technology for Telecommunications Applications*, IVTTA '98, Lausanne, Switzerland, pp. 161-165, 29-30 Sept. 1998.

[4] X.Y. Zhao, Z. Ou, M. Chen and Z.Y. Wang, "Closely Coupled Array Processing and Model-Based Compensation for Microphone Array Speech Recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP '05, Vol. 1, pp. 417-420, 18-23 March 2005.

[5] M.S. Brandstein, J.E. Adcock and H.F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Trans. on Speech and Audio Processing*, Vol. 5, Issue 1, pp. 45-50, Jan. 1997.

[6] P. Smaragdis and P. Boufounos, "Position and Trajectory Learning for Microphone Arrays," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP '07, Vol. 15, pp. 358-368, Jan. 2007.

[7] Z.Y. Li, R. Duraiswami, "Headphone-Based Reproduction of 3D Auditory Scenes Captured by Spherical/Hemispherical Microphone Arrays," *Proceedings of IEEE International Conference on Speech and Signal Processing*, Vol. 5, pp. V337-V340, 14-19 May 2006.

[8] Z.Y. Li, R. Duraiswami, "A Robust and Self-Reconfigurable Design of Spherical Microphone Array for Multi-Resolution Beamforming," *Proceedings of IEEE International Conference on Speech, and Signal Processing*, Vol. 4, pp. IV1137-IV1140, 18-23 March 2005.

[9] H. Alghassi, S. Tafazoli, P. Lawrence, "The Audio Surveillance Eye," *Proceedings of IEEE International Conference on Video and Signal Based Surveillance*, pp.106–111, Nov. 2006.

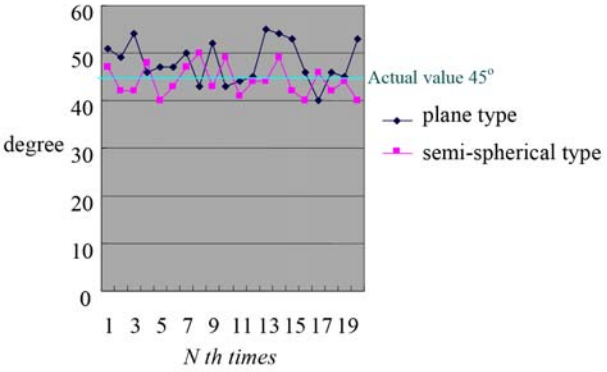


Fig. 9 The comparison of the tests when the sound is of 45° that are based on the plane and the semi-spherical types of microphone arrays.

Table 4. The comparison of the tests when the sound in front/rear of the proposed system

Direction (°)	Nth test	Result	Error	Average
0° (front)	NO1	3.1	3.1	3.3°
	NO2	-2.4	2.4	
	NO3	-2.1	2.1	
	NO4	2.6	2.6	
	NO5	-3.5	3.5	
	NO6	1.3	1.3	
	NO7	-6.9	6.9	
	NO8	-6.9	6.9	
	NO9	2.5	2.5	
	NO10	1.5	1.5	
180° (rear)	NO1	183.1	3.1	3.2°
	NO2	186.2	6.2	
	NO3	178.5	1.5	
	NO4	174.4	5.6	
	NO5	181.2	1.2	
	NO6	184.6	4.6	
	NO7	177.5	2.5	
	NO8	177.2	2.8	
	NO9	182.6	2.6	
	NO10	181.7	1.7	

Table 5. The comparison of the tests when the sound are departed from the proposed system 100cm and 200cm

Real departed distance	*N*th test	Voltage	Estimated distance	Error	Average
100 cm	NO1	1.87V	105.9cm	**5.9cm**	4.39cm (4.39%)
	NO2	1.90V	103.4cm	3.4cm	
	NO3	1.89V	104.2cm	4.2cm	
	NO4	1.89V	104.2cm	4.2cm	
	NO5	1.91V	102.5cm	2.5cm	
	NO6	1.90V	103.4cm	3.4cm	
	NO7	1.85V	107.6cm	**7.6cm**	
	NO8	1.89V	104.2cm	4.2cm	
	NO9	1.88V	105.1cm	5.1cm	
	NO10	1.90V	103.4cm	3.4cm	
200 cm	NO1	0.91V	207cm	7cm	9.07cm (4.535%)
	NO2	0.88V	210.5cm	10.5cm	
	NO3	0.92V	205.8cm	5.8cm	
	NO4	0.87V	211.6cm	**11.6cm**	
	NO5	0.91V	207cm	7cm	
	NO6	0.88V	210.5cm	10.5cm	
	NO7	0.87V	211.6cm	**11.6cm**	
	NO8	0.92V	205.8cm	5.8cm	
	NO9	0.89V	209.3cm	9.3cm	
	NO10	0.97V	211.6cm	**11.6cm**	