Journal of Computers Vol. 31 No. 5, 2020, pp. 198-211
doi:10.3966/199115992020103105015

Constructing the Microphone Array Monitoring System for Smart Concert Hall: A Case in Cello Section Rehearsal

Yang Liu[1*], Feng Tong[2]

[1] Music Department, Art College, Xiamen University, China

[2] Key Laboratory of Underwater Acoustic Communication and Marine Information Technology of the Ministry of Education (Xiamen University), China
123120356@qq.com, ftong@xmu.edu.cn

Received 8 June 2019; Revised 3 August 2019; Accepted 6 January 2020

Abstract. Design and evaluation of a microphone array monitoring system for smart music hall is addressed from a multidisciplinary viewpoint, with cello section rehearsal as a specific case. It is recognized that, for a typical music hall a successful performance of symphonies, string music, or ensembles depends upon the quality of each instrument section and their perfect cooperation. Therefore, sectional rehearsal is the basic requirement of such performances. Professional orchestras generally can obtain professional guidance from conductors and section leaders to conduct standardized sectional rehearsals. However, it is generally difficult, if not impossible, for unprofessional symphonies orchestras organized by amateurs such as schools, government entities or enterprises to find good rehearsal guidance, thus hindering the further popularization of symphonic performance among music lovers of various communities. Thus there is an urgent lack of quantitative guidance way for amateur orchestras. With an advantage of improving acoustic signal processing, the technology of microphone array has been widely applied in sound source localization, noise reduction and sound amplification. Thus, in view of the lack of a qualitative sensing way for teaching assistance of sectional rehearsal, a microphone array monitoring system is designed and evaluated for smart rehearsal teaching of the cello section. By optimizing the parameters of the microphone array according to the acoustic characteristics of the cello, it aims to realize real-time monitoring of the volume, pitch and other parameters of the cello for the smart teaching. The experimental results are provided to show the effectiveness and feasibility of microphone-array-assisted smart teaching of cello sectional rehearsal.

Keywords: cello, ensemble, microphone array, sectional rehearsal, smart living technology and innovations, smart music hall

1 Introduction

The music hall is a professional place for playing and listening of music, typically symphony. The word “symphony” has an etymological root in the ancient Greek word symphonia, meaning “harmony” [1-2]. As one of the most widely used forms of music, symphony first appeared in the first half of the eighteenth century as the opera overture. Symphony in its modern sense refers to instrumental ensemble played by large orchestras. As the highest form of musical performance, symphony is a very expressive form to explore great themes, rich feelings and profound thinking, which demands high skills both for composition and performance. Therefore, it is regarded as a measurement of a country or nation’s musical levels.

As the most typical playing unit in music hall, a symphony orchestra consists of strings, pipes, and percussion, also known as an orchestra. According to the requirements of compiling performance works, the size of the compilation can be divided into three types according to the arrangement of woodwind music in the and: single band, double band and three band. A symphony orchestra is usually composed of

* Corresponding Author

90 to 100 people. In order to ensure the balance of timbre, volume and sound between each part of the band, the instrument configuration of each part has a certain proportion. Take the establishment of standard band as an example: the string part accounts for about 60%, including the first violin part, the second violin part, the viola part, the cello part and the double bass part. The wind section accounts for about 30%, including flute, piccolo, clarinet, oboe, clarinet, bassoon, horn, trumpet, tuba and trombone. The percussion part accounts for about 10%, including: timpani, tambourine, tambourine, trigon, xylophone, cymbals, and clavichord, etc. There are others like the harp. Of course, the conductor is the soul of the band, is to ensure that the band accurate, harmonious, complete performance of the fundamental. And the violin family occupies quite heavy weight, so generally in the position closest to the conductor, the conductor's left hand is generally the violin, the right hand is the cello.

In the process of symphony orchestra performance, the cooperation of the whole team is of great significance. Everyone is a member of the team and an indispensable part of the whole team. Everyone must have the sense of team cooperation to ensure the smooth progress of the whole performance. Although the conductor has overall control over the performance of the whole symphony, leading and guiding the correct direction of strength, expression and music, if a certain musician in a certain province has serious intonation inaccuracy or wrong rhythm, it will have a serious impact on the performance of the whole symphony. For symphonic or string instrumental performance, the rehearsal of the sub-part is very important, especially during the early stage of the rehearsal where difficult parts of the entire work are dealt with. This makes sectional rehearsal a must for a successful symphonic performance. For example, a symphony requires a separate rehearsal of the violin part or the cello part, and for a chorus, the tenor part, and so on. In a professional orchestra a conductor and a section leader plays a leading and guiding role during the rehearsal process [3-4].

With the establishment of a harmonious society and the improvement of the living standard in China, art education is drawing more attention. Currently a large number of non-professional entities such as schools, armies, factories and mines, enterprises have organized their own amateur symphony orchestra, greatly contributing to the popularity of symphony as a form of collective musical enjoyment. In the symphonic teaching of music education, the spirit of team cooperation should run through the whole symphonic teaching, which is not only helpful to cultivate students' profound world cultural background, broaden the knowledge field and build a reasonable knowledge structure, but also helpful to cultivate students' strong team spirit and sense of cooperation. This also promoted the emergence of a large number of symphony orchestras, string orchestras and wind orchestras of various levels in the school. However, in these orchestras, a lot of basic rehearsals are very often in short of professional guidance, which greatly degrades the quality of rehearsals, thus restricting the orchestra's rehearsal level, weakening the final artistic expression and appeal of the orchestra. Without professional guidance, it is difficult for these type of amateur symphony orchestras to achieve the basic level that required to express the music ideological content.

Specifically, it has been recognized that a successful performance of symphonies, string music, or ensembles depends upon the quality of each instrument section and their perfect cooperation. Therefore, sectional rehearsal is the basic requirement of such performances. Professional orchestras generally can obtain professional guidance from conductors and section leaders to conduct standardized sectional rehearsals. However, it is generally difficult, if not impossible, for unprofessional symphonies orchestras organized by amateurs such as schools, government entities or enterprises to find good and professional rehearsal guidance, thus hindering the further popularization of symphonic performance among music lovers of various communities.

Under the framework of the concept 'smart living', the rapidly increasing 'smart' technology such as artificial intelligence and machine learning plays more and more importance role in improving the level of living for modern society, such as in the field of family, health, education, environment protection. Consider that for the basic practices of the symphonies orchestras, one key factor that is important for coordinating of the whole orchestra is to identify the presence of simple errors such as 'abnormal volume' and 'abnormal pitch'. Actually, for this kind of basic guidance that happens frequently in the initial practices of amateur symphonies orchestras, it is possible to replace the professional rehearsal guidance with a 'smart' teaching way. Thus there is an urgent lack of quantitative guidance way for amateur orchestras in terms of the research and development of 'smart' teaching technology, which offers the initial motivation of the present work in applying 'smart teaching' in unprofessional symphonies orchestras to enable "smart music hall.

While the cello (Violoncello) is a western stringed instrument that has a long history, Cello section is a critical part in the orchestra [1]. Its tone is deep and melodious, its range wide and deep, demanding delicate playing skills. Cello solos feature strong and intense feelings with softness and grace. When accompanied by other instruments, the cello often plays a role of connecting and transitioning. Having a high resemblance with human voice, it can play both low and high pitch section. It is also one of the most suitable western musical instruments that can be joined by traditional Chinese instruments. In addition to solo, it enjoys an irreplaceable place in symphony, orchestral, chamber music, ensemble, accompaniment and concerto. Thus, in this paper the cello section rehearsal was chose as a specific case for constructing the microphone array monitoring system for smart concert hall.

2 Literature Reviews

2.1 Smart Living

In recent years, with the development of big data, cloud computing, Internet, Internet of things and other technologies, artificial intelligence has crossed the technology gap between science and application and entered the explosive growth period. “intelligence +” has become an innovative paradigm and penetrated into all walks of life [21-23]. At present, artificial intelligence has reached or exceeded the human level in visual image recognition, speech recognition, text processing and other fields, and has made a breakthrough in the fields of visual art and programming. It has achieved remarkable success in image classification, automatic driving, machine translation, gait movement and question answering system.

Artificial intelligence refers to the science and technology that replaces human beings with machines to realize functions such as cognition, recognition and analysis. Its essence is the simulation of human consciousness and thinking process, and it is an interdisciplinary subject that integrates computer science, physiology and philosophy. In a brief history of the future, Israeli historian yuval hariri said that artificial intelligence based on big data and complex algorithms is making the world go through a huge leap from homo sapiens to “god man”, which is more revolutionary than the transition from ape to man. Such times enable artists to acquire unusually rich, diverse, broad and profound thinking materials, life practices and life experiences, thus brewing new soil for the epoch-making artistic peak. It is an important task for contemporary artists to give full play to the aesthetic and artistic perception, imagination, shaping and penetration brought by artificial intelligence.

2.2 The Technology of Microphone Array

Microphone array technology is a common processing method in the field of speech recognition. It is designed to improve spatial acoustic processing by making use of microphones organized into a certain geometric array so as to obtain the sound source location, reduce background noise and enhance the effect of the sound signal, smart living technology and innovations. In recent years, the technology of microphone array has been widely used in automotive systems environment, video conferencing, hearing aids, speech recognition front-end system, large-scale recording system at meeting place, robot navigation and other civilian and military products for the purpose of speech enhancement, source localization and noise mitigation [5-7].

The microphone array, which consists of sets of microphone sensors arranged in specific spatial patterns, has drawn extensive attention from various fields such as speech recognition, speaker recognition, man-machine interact, robot in practical noisy environments [1-5]. In dramatically increasing investigations, received signals of microphone array are popularly processed for the purpose of speech enhancement, noise mitigation or multiple source separation.

As a classic algorithm [8-17], the steered response power (SRP) microphone array beamforming algorithm achieves DOA estimation by maximizing the power output of the spatial filter. As the spatial resolution of the classic SRP method is limited by large peaks of the response power function, the SRP phase transform (SRP-PHAT) algorithm considers only the phase information for normalization to yield narrower response power peaks. While the Multiple Signal Classification algorithm (MUSIC) has been developed and used for frequency domain high resolution DOA estimation [10], the classic MUSIC algorithm is in most case suitable for narrow band signals.

Meanwhile, under the harsh conditions of reverberation and background noise, the traditional DOA estimation approaches are generally subject to substantial performance degradation. By transforming the estimation of DOA into the problem of sparse reconstruction, the compressed sensing (CS) microphone arrays sound source localization algorithm has been derived in recent years [20-21]. Compared with the traditional sound source localization algorithm such as steered response power phase transform (SRP-PHAT) algorithm, previous investigations indicated that the compressed sensing-based sound source localization algorithm is capable of achieve high estimation accuracy.

However, classic CS DOA methods always assume the sources to be spatially located under a classic sparse pattern to guarantee the performance of sparse recovery. In other words, in spatial domain a dominate direction bin associated with a source is required to be centered by adjacent direction bins with zero magnitude. Thus, under the reverberant environment, the CS DOA method will still suffer from performance degradation as the sources are located in clustered sparse manner due to reverberation or a large beam-width source.

In this paper, considering that there is a lack of similar research to apply the microphone array technology in the area of smart music education, different microphone array DOA estimation algorithms as introduced previously are considered. Consider that the ‘smart’ teaching system for the proposed symphonies orchestra should be a ‘real-time’ and portable system from the viewpoint of engineering application, the classic SRP-PHAT algorithm is adopted for obtain the spatial information of the sectional rehearsal due to its low computational complexity compared to those need high computational load such as MUSIC or CS type algorithms.

2.3 Acoustical Characteristics of Cello

The volume of the cello sound depends on the “amplitude” of the vibration of the cello body. When a string is bowed (plucked), the string vibrates and sound is produced. The bigger the amplitude, the louder the sound produced by the instrument. When the cello is played, strings are repeatedly bowed to produce high-frequency vibration. The vibrations of the strings are transmitted through the bridge to the hollowed body of the cello so that more air is moved to produce louder sound. When the strings are gently bowed the amplitude of the sound is smaller and the volume is lower.

From the musical point of view, the pitch of the sound depends on the speed of the instrument vibration. Musical instruments produce a “treble” when vibrating frequency is high, and “bass” is produced when vibration frequency is low. The “frequency” of the sound is called the “pitch” in musical terms. In general, smaller instruments produce faster vibration, therefore generating higher-pitched, sharper sound, while the larger musical instruments produce slower, longer and more melodious sound. For example, violins produce higher-pitched sound than the cello. In addition to the size of the hollowed body, the material of the strings, chord length (tension), and the string pressure (bow pressure) also affect the pitch of the sound that an instrument produces. In general, a string pressed by a finger produces higher-pitched sound than an empty string; the pitch of thinner strings is higher than thicker ones, and it increases as the strings are tightened.

Within the frequency range of the sound produced by the cello [2]: 100hz-200hz affect the sound fullness, 3Khz affect the brightness of the sound. The lowest bass is a large group C, and the frequency is 64/s; the highest pitch is up to three groups of e, and the frequency is 1280/s. But these figures are only the deduced theoretical frequency, which may not always be consistent with our sense of hearing. Due to the characteristics of the human ear structure, low bass sounds lower and treble sounds higher to our ears.

Although different voice boxes have different volumes and different strings have different pitches, still different instruments and human are able to produce a vast range of sounds of different pitch and volume. The reason is that almost all of the vibration is compounded. For example, when a cello string is vibrating, it is not only about the full-length of the string, each part of string is vibrating too, generating sub-vibrations. Although the sounds produced by sub-vibrations are not easily detected by our human ears, they are all integrated into the overall sound effect.

The abnormality of the volume and the pitch of each part affect the auditory experience of the performer and the audience. That explains why each part requires well-organized sub-part rehearsal to achieve the best effect of the resonance, so that a better mix of sound is to be conveyed to the ears of the audience [3-6]. Take the cello part as an example, the spacing between cellos is generally 1m in cello rehearsals. A medium-sized orchestra has 8 cellos, a large orchestra 12, and a string orchestra 4. From the musical point of view, each performer should produce sound of the same volume and pitch so as to achieve good resonance and satisfactory performance of the symphony. [7-9].

2.4 Contribution and Originality of the Proposed System

While the technology of "smart teaching'" has been applied in various field such as interactive course learning and teaching of preliminary education, middle school education and higher education [24], previously there was few report on the investigation of 'smart' technology in the area of amateur symphony orchestras. It is worth pointing out that, as most of the current research on microphone array is carried out for the purpose of man-machine interaction, speech recognition, video conference or intelligent speech device, microphone array for the purpose of 'smart' music teaching will need special design in the system configuration and algorithm algorithms to adapt to the different scenarios. In other words, most of the current microphone array systems are designed according to the acoustic characteristics of human speech, so special investigation and research should be performed to take the acoustic features of music instrument such as cello into account for the proposed 'smart' music teaching system.

In this paper, the microphone array technology is introduced into the amateur symphony orchestra's sectional rehearsal for the purpose of smart teaching assistance from a multidisciplinary viewpoint. The concept of 'smart teaching' here means that no professional manual guidance is needed in the teaching process, which can be fulfilled by smart systems consisting of computer and sensors. Specifically, the microphone array is adopted in this paper to design a smart music teaching system from a multidisciplinary viewpoint.

The number of violin parts is large in the process of symphony rehearsal, and the vocal part is of great significance. Only strict and correct performance can accurately express the music connotation of the whole symphony. We selected the rehearsal of the cello part in the symphony orchestra to conduct a smart teaching experiment, taking into account of the need for cello section rehearsal and the distinctive acoustic characteristics of the cello. The microphone array parameters is specifically designed and adjusted for cellos. A microphone-array-assisted smart teaching experiment is conducted in the basic rehearsing of the cello unit to evaluate the effectiveness of the proposed system.

3 Direction of Arrival Estimation (DOA) Based on Microphone Array [8-11]

As the principle of microphone DOA has been provided extensively, we just give a brief introduction in this section. The signal model of the narrow band far-filed array can be briefly introduced as follows: With a source at far-field, signal s(t) that received by an M element uniform linear array (Uniform Linear Array, ULA) is expressed as:

$$s(t) = \tilde{s}e^{jw_0 t} \quad \textbf{(1)}$$

$$\tilde{s} = u(t)e^{\varphi(t)} \quad \textbf{(2)}$$

Where: w, u(t), $\varphi(t)$ is the angular frequency, magnitude and phase of s(t) respectvely, $x_i(t)$ is the signal received by the ith element with the first element as reference, $i = 0,1,2,...,M-1$. Thus, we have:

$$x_0(t) = s(t) = \tilde{s}e^{jw_0 t} \quad \textbf{(3)}$$

$$x_i(t) = x_0(t - \tau_i) = s(t) \quad \textbf{(4)}$$

where τ_i is the time delay between the ith element and the reference. The model of the ULA is provided as in Fig. 1.

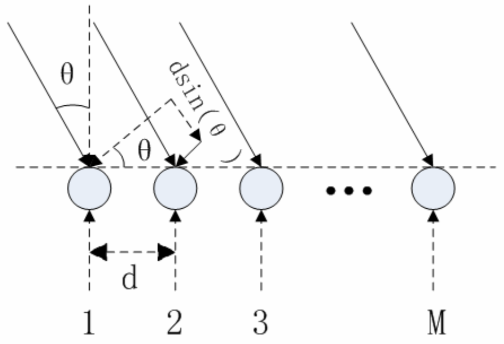


Fig. 1. Brief illustration of ULA

The delay compensation for the ith element can be expressed as [10]:

$$\tau_i = \frac{id\sin\theta}{v} = i\tau \quad \textbf{(5)}$$

Where: d is the space of adjacent element, θ is the incident angle of speech. Under the condition that the spatial sampling principle is satisfied, the parameter d will determine the delay and phase difference between different elements. A larger d leads to a higher spatial resolution of array. Because $s(t)$ is narrow band signal and $\tilde{s}(t)$ is slowly time-varying, we have:

$$\tilde{s}(t) = \tilde{s}(t-\tau_i) \quad \textbf{(6)}$$

$$x_i(t) = s(t-\tau_i) = \tilde{s}(t-\tau_i)e^{j\omega_0(t-\tau_i)} \approx \tilde{s}(t)e^{j\omega_0(t-\tau_i)}e^{-j\omega_0\tau_i} \quad \textbf{(7)}$$

Thus, we get:

$$x_i(t) = s(t)e^{-j\omega_0\tau_i} = s(t)e^{-j\omega_0 i\tau} \quad \textbf{(8)}$$

Defining the spatial phase as:

$$\phi = \omega_0\tau = \frac{2\pi\omega_0 d\sin\theta}{v} = \frac{2\pi d\sin\theta}{\lambda} \quad \textbf{(9)}$$

Where: ω_0, λ are the frequency and wavelength of the received signal respectively, then the spatial phase difference between the received signal at the ith element and that at reference element is $i\ \phi$. As a result, the received signal vector can be expressed as:

$$\begin{bmatrix} x_0(t) \\ x_i(t) \\ \vdots \\ x_{M-1}(t) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\phi} \\ \vdots \\ e^{-j\phi(M-1)\phi} \end{bmatrix} s(t) \quad \textbf{(10)}$$

Where:

$$X(t) = [x_0(t)x_1(t)\cdots x_{M-1}(t)]^T \quad \textbf{(11)}$$

$$A(\phi) = [1\ e^{-j\phi}\cdots e^{-j(M-1)\phi}]^T \quad \textbf{(12)}$$

By converting the equation (12) into vector form, we get:

$$X(t) = A(\phi)s(t) \quad \textbf{(13)}$$

Suppose the noise vector is:

$$N(t) = [n_0(t) \quad n_1(t) \cdots n_{M-1}(t)]^T \quad \textbf{(14)}$$

The equation (13) is further written as:

$$X(t) = A(\phi)s(t) + N(t) \quad \textbf{(15)}$$

For the scenario with only one source, the direction of which can be obtained by：

$$\phi_{\max} = \arg\max_{\phi} \sum_{i=1}^{L} X(t)_i^2 \quad \textbf{(16)}$$

If there are multiple sources, the spatial distribution can be obtained as:

$$f(\phi) = \sum_{i=1}^{L} X(t,\phi)_i^2 \quad \textbf{(17)}$$

Meanwhile, different tones produced by the cello correspond to different frequency. Thus, by designing band-pass filter with different center frequency, the direction information of the cello sounds can be located by microphone array signal processing. In Fig.2 the corresponding flowchart of the signal processing is provided.

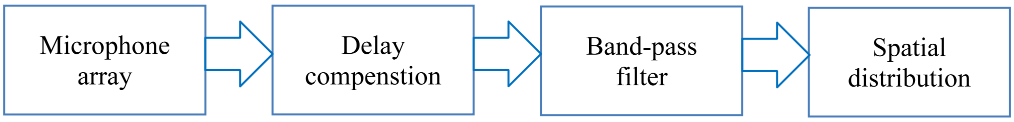


Fig. 2. The corresponding signal processing flowchart of ULA

Under the framework of the microphone array DOA algorithm, there have been several popular and complicated algorithms as introduced in the previous section such as MUSIC or CS type algorithm [10], which is capable of achieving promising DOA performance but unfortunately needs a high computational burden for implementation in terms of real-time calculation capability and memory space. Thus, for the proposed ‘smart teaching’ purposed, these types of algorithm are not suitable to be adopted in this paper as the DOA algorithm.

4 Design of Microphone Array Monitoring System for Cello Concert Smart Teaching

For the microphone array system, the algorithmic parameters play an important role in determining the theoretical performance of the signal processing. According to the acoustic feature, spectral range and concert spatial configuration of the cello, the parameters of the microphone array used for assisting the cello concert are designed as follows.

The parameter d is an important factor of the array, which determines the performance of the source localization. Generally speaking, the value of d is proportional to the spatial resolution of the array, with a large d corresponding to a high spatial resolution. From the viewpoint of theory, d is generally set at half of the wavelength of the interested signal. As the frequency range of the cello is at 100-3000Hz, the relationship of d can be expressed as: $d = \frac{\lambda}{2} = \frac{c}{2f}$, where λ is the wavelength of the signal, c is the velocity of sound in air, namely 340m/s; f is the frequency of the signal with a range of 100-3000Hz. It is apparent that, to meet the need of the signal with the highest and the lowest frequency, the value of d is determined by： $\frac{c}{2f_{\max}} \le d \le \frac{c}{2f_{\min}}$. Accordingly, d is set to be 0.25m in this paper.

Shown in Fig. 2 is the block chart of the proposed microphone array ‘smart teaching’ system for assisting the cello concert. As revealed in Fig. 2, the proposed system consists of a microphone array, signal acquisition, signal conditioning, DOA processing, quantitative output and the parameter configuration modular. Function of each modular is briefly introduced as:

The microphone array is used for multiple channel signal collection with the specially designed array geometry. In the proposed system designed for 'smart teaching' of cello concert, a uniform line array is used to ensure the DOA performance along the side of cello players. In the practical application of cello concert scenarios, different microphone array geometry can also be used to meet the requirement of different purpose or different hall. As to the number of microphone in the array, while a large microphone number will lead to improved performance of DOA estimation, it simultaneously increases the computational complexity of the signal processing.

The signal acquisition system is used for control the multiple-channel ADC to enable the signal processing in processor. In the proposed system, multiple-channel ADC of the adopted processor is directly used for signal acquisition to simplify the system design.

The function of the signal conditioning modular is to amplify and filter the collected signal to facilitate the following microphone signal processing to obtain the spatial information. Specifically, in the proposed system the integration operational amplifier chip MAX9814 is directly utilized for signal conditioning. Note that, as the MAX9814 is a single channel amplifier, the same number of MAX9814 is needed for construction of signal conditioning system of a microphone array system consisting of M microphone.

The DOA processing modular is the key part of the proposed system, in which the DOA algorithm is running real-time to obtain the spatial information of the cello concert to facilitate the correct identification of abnormal events during the cello concert practice. As the fixed-point digital signal processing will lead to loss of calculation precision, the float-point digital signal processor offers a better selection.

The quantitative output modular acts as a graphic indicator for the 'smart teaching', which provides the quantitative results of the DOA processing for the purpose of 'smart teaching'. For example, consider the detection of 'abnormal volume' or 'abnormal pitch' as designed in the proposed work, the presence of these two types of events as well as the specific direction of the abnormal events will be used for quantitative output to facilitate 'smart teaching'. Namely, with the help of quantitative output, existence of typical abnormal events will be real-time identified and thus facilitate the efficient teaching. To be specific, according to the principle of cello concert, automatic threshold can be also designed in the quantitative output modular to enable automatic indication of the abnormal events in cello concert in the forms of "warning sounds" or 'warning flash light'.

The parameter configuration modular is used for setting the parameter of the microphone array system to ensure the monitoring performance of the 'smart teaching' system. The parameters include the geometry of the array, the sampling rate, the DOA estimation calculation parameters and the form for quantitative output. The suitable selection of the system parameters will be dependent on the design principle as introduced in the beginning of this section.

Note that, the system block chart shown in Fig. 3 just offers a general and open implementation framework of the proposed "smart teaching'" concept. From the viewpoint of practical application, the popular intelligent terminal such as cellphone can be also adopted to act as the platform for DOA processing, parameter configuration and quantitative output.

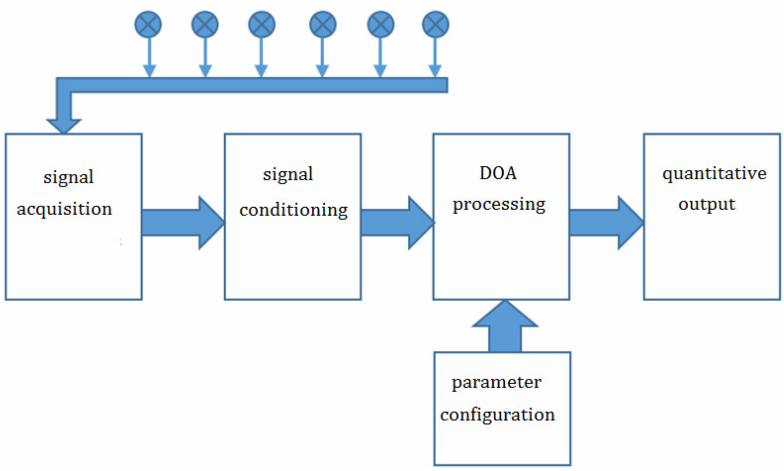


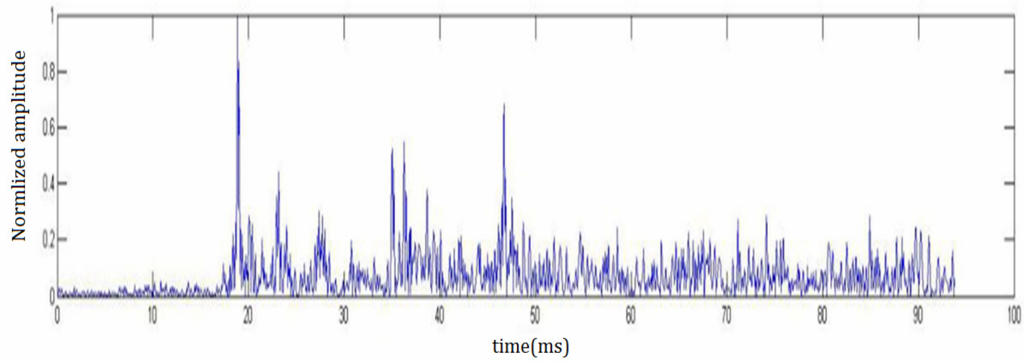
Fig. 3. Illustration of the proposed microphone array based 'smart teaching' system

5 Experiment

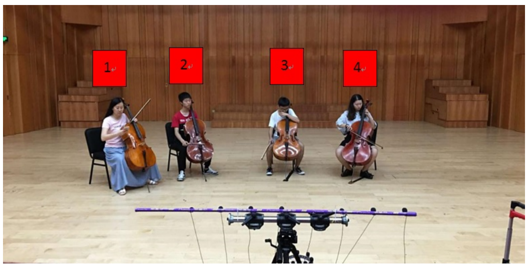
5.1 Experimental Setup

To verify the concept of the ‘smart teaching’ in cello section rehearsal as well as the effectiveness of an experimental ‘smart teaching’ system designed according to the combination of microphone array DOA algorithm and music cello section rehearsal theory, experiments are performed with the help of practical players and hardware/software implementation. The experimental venue is Xiamen University Concert Hall (Li Chengyi - Zhang Zhihua Concert Hall), which is located in the teaching building of Xiamen University Art College. Covering an area of 548 square meters and equipping with 365 seats, it can accommodate a maximum of 420 spectators.

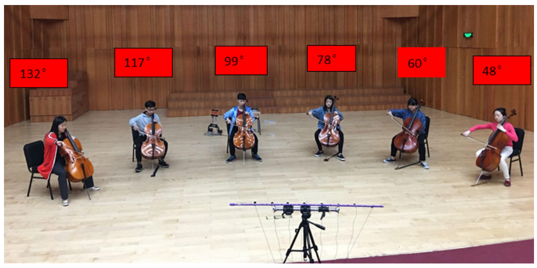
As a large size space that is specially selected to simulate one for symphonies orchestra operation, the room impulse response of the experimental space is measured with the help of linear frequency modulation (LFM) signal played in the location of cello players as shown in Fig. 4(b). As indicated from the Fig. 4(a), there are significant reverberation components in the room impulse response of the hall which lasting a range of about 30ms [20]. Thus, the experimental results will be obtained under a practical reverberant environment.



(a) Room impulse response of the experimental hall



(b)



(c)

Fig. 4. Experimental room impulse response and the 4-player and 6-player cello concert scenario

Under the concept of ‘smart teaching’, the microphone array monitoring system adopted in this paper is an 8-element ULA with the d parameter of 0.25m. The multi-channel data acquisition system is based on an STM32F407 microprocessor with a sampling rate of 16ksps and a quantization width of 16 bit to enable real-time implementation of the DOA algorithm. To guarantee the quality of signal, the low-noise microphone amplifier chip MAX9814 was selected for pre-processing of the multi-channel signal collected by microphone array, which provides a gain setting range of up to 60dB and the auto gain control (AGC) capability.

By achieving a good balance between the float-point calculation performance and the system overhead such as energy and size, the STM32F407 processor has a clock speed of up to 168MHz and has the advantages of short cycle times and low power consumption. Compared to other processing platform such as PC, laptop or professional DSP chips, the STM32F407 is chosen as real time processor because of its low cost and small size, which offer the potential to be developed into a portable 'smart teaching' device. Meanwhile, the quantitative output of the 'smart teaching' monitoring result can be implemented by the LED directly connected to the main processor, or be further extended to the screen of intelligent terminal such as cellphone via blue-tooth link.

Because of the adoption of STM32F407 processor, the multi-channel signals collected by the microphone array can be processed real-time to produce the quantitative output. In other word, there is no need to collect and save the dataset for off-time processing. Thus the proposed system provides the potential for practical implementation.

Consider the design purpose and the 'smart teaching' concept of the proposed system, 4-player and 6-player experimental scenario is designed for performance evaluation of the cello concert teaching assistance respectively. In the 4-player rehearsal experiment 4 cellos are performed simultaneously to simulate a typical small scale cello section rehearsal for data collection, with the distance between adjacent cellos at 5m. The direction associated with each cello is 60°, 75°, 90°, 99° respectively as shown in Fig. 4(b). For the 6-player scenario, the direction of 6 cellos is 48°, 60°, 78°, 99°, 117° and 132° respectively as shown in Fig. 4(c).

For the purpose of analysis and comparison, cello sounds produced with difference abnormal volume and different abnormal pitches are collected by the microphone array system. To mimic the usual errors of the cello beginners in section rehearsal, 'abnormal volume' and 'abnormal pitches' are selected as the detection target for the experimental 'smart teaching' system. In the practical practice of amateur symphonies orchestras, these types of usual errors frequently happened and thus seriously affect the quality of orchestra practice.

Fig. 5 shows the cello ensemble signal and its spectrogram collected by the experimental microphone array based 'smart teaching' system. It tells us the spectral distribution of cello sound, featuring a number of spectral lines composed of discrete spectrum below 4500Hz. For the cello ensemble, in order to ensure the success of performances, it is required that the cellos in the ensemble play the same volume and pitch with accurate time consistency. Using the microphone array technology to monitor the parameter of cello sound volume and pitch will enable amateur players and orchestras to greatly improve their rehearsal quality, thus facilitate the smart teaching.

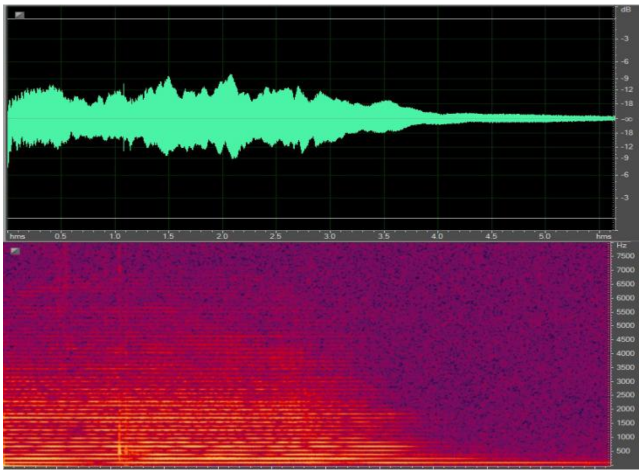


Fig. 5. Cello ensemble waveform (above) and spectrum (below) collected by microphone array monitoring system

5.2 Experimental Results and Discussion

5.2.1 Monitoring of "Abnormal Volume"

From the viewpoint of cello concert smart teaching, abnormal volume is a common problem for beginners in cello ensemble rehearsal, and the obvious volume abnormality in ensemble will seriously affect the ensemble rehearsal coordination. In this section 4-player and 6-player cello rehearsal practice experiment has been conducted to test the effectiveness of the smart teaching system. Shown in Fig. 6 is the spatial energy distribution obtained by the microphone array in the sample rate bandwidth when the

4-player cell volume is normal. As illustrated by Fig. 6, four obvious beams are shown from four directions respectively. The beam peaks correspond to the four cellos located at four different directions. When four cellos are in coordination, variation of the amplitude of four beams is within a limited range.

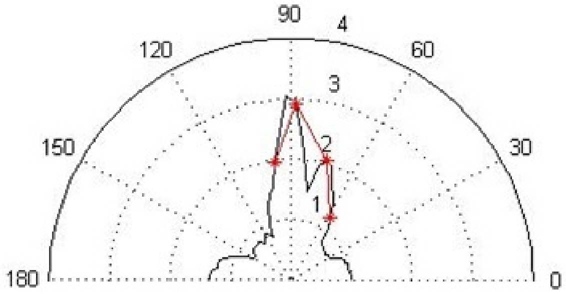
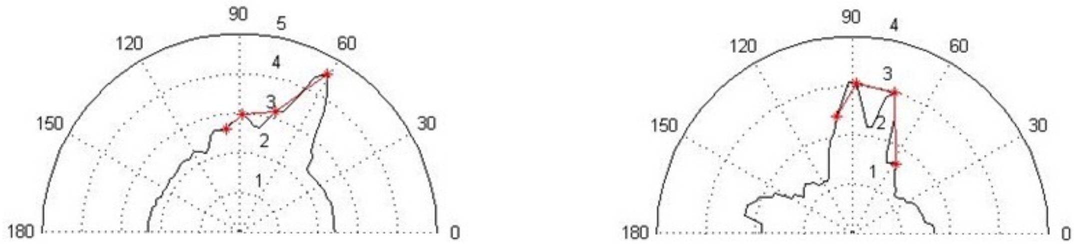


Fig. 6. Spatial distribution pattern of the norm 4-player cello rehearsal

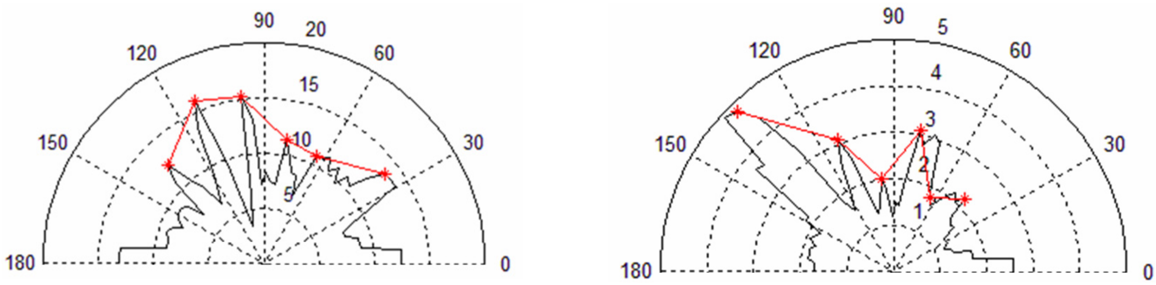
For the 4-player rehearsal, Fig. 7(a) and Fig. 7(b) provides the beam energy distribution when the cello sound from 60-degree direction is abnormally large or small. As one may observe from Fig. 7, the volume abnormality can be easily detected and monitored through the corresponding beam peak collected by the microphone array. The abnormal beam peak is obviously higher or lower than the other ones. Thus, based on the result of microphone signal processing, it is possible to detect the abnormal volume during the cello part rehearsal from the viewpoint of 'smart teaching'.



(a) (b)

Fig. 7. Spatial distribution pattern of the 4-player cello rehearsal with abnormal volume in 60 degree

In Fig. 8(a) spatial distribution pattern of the normal 6-player cello rehearsal was provided, from which one may notice six beams corresponding to six players. As the volume of 6-player cello rehearsal is normal, magnitude of six beams exhibits relatively small variation. As shown in Fig. 8(b), the spatial distribution pattern of the 6-player cello rehearsal with abnormal volume in 132 degree clearly displays an abnormal beam with much higher magnitude in the 132 degree.



(a) (b)

Fig. 8. Spatial distribution pattern of the normal 6-player cello rehearsal (a) and abnormal volume in 132 degree

5.2.2 Monitoring of "Abnormal Pitch"

As a popular error for the cello beginners, the problem of abnormal pitch also leads to less-than-ideal performance effect and disharmony of the cello rehearsal. Consider that the different pitches of cello correspond to different frequency, the band-pass filter is adopted to select suitable frequency band to facilitate the detection of abnormal tone.

In the experiment, a high band-pass filter and a low band-pass filter with the center frequency at 1920-2880Hz and 3360-4320Hz are adopted respectively for monitoring the "abnormal pitch". After the band-pass filtering, the spatial distribution pattern of the 4 cellos playing D tone is provided in Fig. 9, which reveals that all the 4 peaks associated with 4 cellos approximately exhibit similar magnitude.

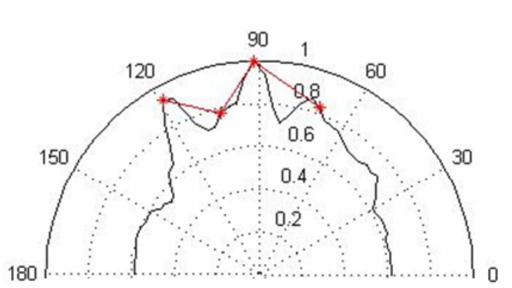


Fig. 9. Spatial distribution pattern of the 4-player cello rehearsal playing D pitch

To artificially simulate the presence of abnormal pitch in 4-player cello rehearsal, the cello at 120 degree abnormally plays G tone, and A tone respectively while all the other 3 cellos play D tone. After high band-pass filtering and low band-pass filtering, the spatial distribution pattern is provided in Fig. 10(a) and Fig. 10(b) respectively, from both of which we can tell that the peak at 120 degrees exhibits much higher magnitude compared with the other 3 peaks. The experimental results demonstrate that the employment of microphone array is capable of detecting the abnormal pich in the cello rehearsal.

Furthermore, the abnormal pitch in 6-player cello rehearsal was also evaluated, with the cello at 132 degree abnormally playing G tone, and the others playing D tone. After high band-pass filtering, the spatial distribution pattern of the 6-player rehearsal with normal pitch is provided in Fig. 11, which indicating that the proposed system is capable of identifying the abnormal pitch.

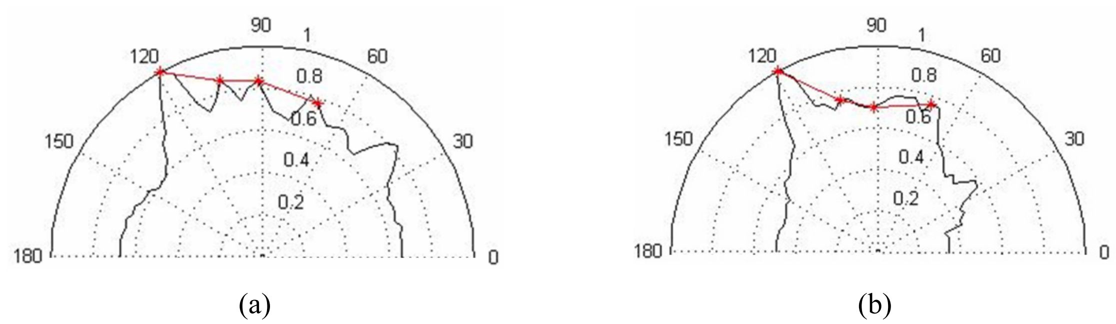


Fig. 10. Spatial distribution pattern of the 4-player cello rehearsal with abnormal pitch in 120 degree

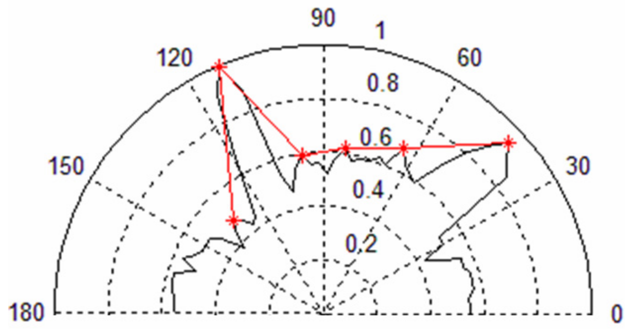


Fig. 11. Spatial distribution pattern of the 6-player cello rehearsal with abnormal pitch in 132 degree

6 Conclusion

To address the lack of professional guidance for the unprofessional symphonies orchestras organized by amateurs, we reported the research and development of the concept ‘smart music hall’ with the symphonies orchestra as a case by constructing microphone array monitoring system, with the purpose to utilize the DOA estimation capability of microphone array to identify some basic and preliminary errors for the beginners. An experimental system is designed and implemented with the real time signal processing capability and graphic quantitative result output to verify the effectiveness and performance enhancement of the concept of ‘smart music hall’, which has the possibility of being applied in the practical scenarios.

Specifically, in this paper the microphone array monitoring technology based on simple and effective SRP-PHAT algorithm is introduced as a quantitative monitoring way to enable the smart teaching of cello ensemble rehearsal. To meet the need of the cello acoustics and cello ensemble rehearsal, the parameters of microphone array are optimized accordingly. The 4-player and 6-player cello rehearsal experimental results show that the system can effectively monitor and locate the abnormal volume and pitch cello sounds during the rehearsal, therefore providing a better solution for cello players rehearse without professional conductors.

With the assistance of the combination of microphone array technology and theoretic music principle, teachers, conductors and section leaders are also able to improve the quality and efficiency of the rehearsal by the means of ‘smart teaching’. Therefore, compared to the classic artificial teaching ways, the microphone array technology based smart teaching technology has the potential to be used by extensive amateur cello players for their part rehearsal.

Finally, it is also worth pointing out that, as the current experiment and signal analysis is still limited, while the general effectiveness of the concept ‘smart music hall’ can be verified, there are much more research that need to be performed in the further to provide more detailed signal processing algorithm, parameter configuration and abnormality detection method to improve the performance of the proposed quantitative microphone array monitoring system.

Acknowledgements

The authors appreciate the funding of the Education Research Grant of Young and Middle-Aged Teacher of Fujian Province (No. JAS170012) for supporting the present research.

References

[1] P. Laird, The baroque Cello Revival: An Oral History, Sarecrow Press, 2004.

[2] R. Stowell, The Cambridge Companion to the Cello: Cambridge Companions to Music, Cambridge University Press, 2007.

[3] D. Homuth, Cello Music Since 1960, Sarecrow Press, 1994.

[4] E. Cowling, The Cello, Charles Scribner’s Sons, 1975.

[5] K. Audhkhasi, A. Rosenberg, A. Seth, B. Ramabhadran, B. Kingsbury, End-to-end ASR-free keyword search from speech, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, 2017.

[6] M. Wolf, C. Nadeu, Channel selectionmeasures formulti-microphone speech recognition, Speech Commun. 57(2014) 170-180.

[7] Z. Huang, S.M. Siniscalchi, C.-H. Lee, Bayesian unsupervised batch and online speaker adaptation of activation function parameters in deep models for automatic speech recognition, IEEE/ACM Trans. Audio, Speech, Lang. Process 25(1)(2017) 64-75.

[8] J. Chen, K. Yao, R. Hudson, Source localization and beamforming, IEEE Signal Processing Magazine 19(2)(2002) 30-39.

[9] X. Wan, Z. Wu, Sound source localization based on discrimination of cross-correlation functions, Applied Acoustics 74(1)(2013) 28-37.

[10] B. Jacob, J. Chen, Y. Huang, Microphone Array Signal Processing, Springer-Verlag, 2008.

[11] L. Vladimir, J. Robert, C. Lakhmi, V. Ljubo, Smart Education and E-learning, Springer, 2018

[12] M. Wax, T. Kailath, Optimum localization of multiple sources by passive arrays, IEEE Transactions on Acoustics Speech & Signal Processing 31(5)(1983) 1210-1217.

[13] D. Bechler, K. Kroschel, Reliability criteria evaluation for TDOA estimates in a variety of real environments, in: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.

[14] B. Krim, M. Viberg, Two decades of array signal processing research, in: Proc. IEEE Sig. 1996.

[15] L. Nunes, W. Martins, M. Lima, L. Biscainho, M. Costa, F. Goncalves, A. Said, B. Lee, A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays, IEEE Transactions on Signal Processing 62(19)(2014) 5171-5183.

[16] J. Dmochowski, J. Benesty, S. Affes, A generalized steered response power method for computationally viable source localization, IEEE Trans. Audio, Speech, Lang. Process 15 (8)(2007) 2510-2526.

[17] Y. Tan, Sound localization method using modified SRP-PHAT algorithm, Journal of Electronics & Information Technology 28(7)(2006) 1223-1227.

[18]D. Malioutov, M. Cetin, A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. on Signal Process 53(8)(2005) 3010-3022.

[19] A. Gurbuz, V. Cevher, J. McClellan, Bearing estimation via spatial sparsity using compressive sensing, IEEE Trans. Aerosp.Electron. Syst. 48(2)(2012) 1358-1369.

[20] J. Li, Y. Zhang, F. Tong, A Microphone-array De-reverberation Method Using Sidelobe Enhancement, Journal of Xiamen University (Natural Science) 56(5)(2017) 711-717.

[21] M. Mokhtari, B. Abdulrazak, H. Aloulou, Eds., Enhanced quality of life and smart living, in: Proc. 15 the International Conference ICOST 2017, 2017.

[22] M. Snider, Psychological Toolkit for Smart Living, 2nd edition, Mill City Press, 2014.

[23] J. Li, J. Jin, D. Yuan, M. Palaniswami, K. Moessner, EHOPES: data-centered fog platform for smart living, in: Proc. International Telecommunication Networks and Applications Conference (ITNAC), 2015.

[24] R. Raghunath, C. Anker, A. Nortcliffe, Are academics ready for smart learning, British Journal of Educational Technology 49(1)(2018) 182-197.