OPEN ACCESS Freely available online

IJSSAI International Journal of Social Sciences and Artistic Innovations

ISSN 2737-5293
Volume 1, Issue 1
https://www.iikii.com.sg/journal/IJSSAI
International Journal of Social Sciences and Artistic Innovations

Article

From Diagnosis to Realization of Virtual Choiring System: On the Way to the Creation of an Immersive and Versatile Practicing and Performing Platform based on New-Generation Communication Technologies

Ching-Mao Chen, Chia-Jen Lee, Chiung-Yuan Lin, Tsung-Fu Yang and Sau-Gee Chen *

Department of Electronics Engineering and Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan
* Correspondence: sgchen@nctu.edu.tw

Received: Apr 20, 2021; **Accepted:** Jun 30, 2021; **Published:** Sep 30, 2021

Abstract: This paper reviews the possible problems when realizing a virtual choir system. Under the common computer specifications, sound cards contribute its delay significantly. Applying Audio Stream Input Output (ASIO) reduces the delay to 6 ms in the proposed real-time virtual choiring system which is a huge improvement. With the existing technology, ASIO significantly suppresses the delay, which helps to realize real-time virtual choiring. The experiments suggest that a few Mbps network bandwidth be adequate to the audio transmission in a real-time virtual choir, but still easily available when considering that video streaming requires much higher bandwidth by using 5G or WiFi 6.

Keywords: Virtual choir, ASIO, sound card

1. Introduction

Since the outbreak of a global pandemic in 2020, learning (e-learning) (Anderson 2008), conferencing (Angeli 2003), and shopping online (Zhou 2007) have received much attention and usages. The global pandemic allows lots of online conferencing and video chatting platforms to have surged greatly due to people's staying at home around the world. In the emerging industry of the online platforms that stood out, Table 1 shows well-known platforms that are used widely (OWL Labs 2020). These applications are for education, business, and the consumer sectors which have given people more choices in user experience than before. As a direct by-product of the emergence of the Internet, online learning (Anderson 2008) has given students and people regardless of money, distance, and resource to learn with. Nowadays, there are many MOOCs (massive open online courses) that are provided by universities and around the world. YouTube, Masterclass, and several platforms have been providing a supporting role in online-learning. Some of them require payments but many do not charge to access them. Thus, online learning has made decent contributions to the ongoing effort to balance the inequality of knowledge and learning resources.

A choir is a musical ensemble of singers. A choir is a well-known form of a musical group and requires people to sing together by joining the singing according to their parts. An assignment of parts is based on the high and low of the voices such as soprano, alto, tenor, and bass. The key technique in a choir is the correct timing as the beauty of the choir lies in the base of fusion between different parts. Thus, not only in performing but also in training, timing is always an important issue when practicing.

Table 1. Summary of functionalities in existing video-conferencing platforms.

Zoom	720p video Up to 100 participants One-on-one and group meetings Screen/Audio sharing (about 1s delay for audio sharing) Multi sharing at same time Scheduled meetings Private and group chat Host controls
Cisco Webex	Up to 6 participants HD video Sharing desktops and documents Integrated communication with CISCO software Cross-platform compatibility: Windows/iOS, desktop/mobile Data protection: unparalleled secure data encryption, conference password protection, and network security protocols
Google Hangouts	Up to 25 participants Video and audio call Private and group chat Video meeting recordings Screen sharing Audio sharing
Skype	Call phone Up to 50 people on a group meeting call HD video calling Screen sharing Audio sharing (about 0.5−1s delay) Video call recording Interactive video chats Real-time subtitles (support Chinese and English)

Virtual Choir is a new concept in recent decades. The concept is based on having singers around the world record and upload their singing videos. Then, people synchronize the videos and combine them into a single performance. At TED 2013, Eric conducted a real-time live virtual choir, performing live streaming from 28 countries around the globe, alongside 100 singers on stage, which was streamed through Skype. Eric and the crew have been the most successful virtual choir team that held an event called Virtual Choir 6 that combines 40000 singers from 145 countries around the world in response to the global pandemic.

In 2020, the choir at Harvard University proposed a method (The Harvard Choruses 2021) to train during the national quarantine in the US in an online conferencing platform, ZOOM. Another similar effort is the HGC (Harvard Glee Club) Alumni Virtual Choir (Harvard Glee Club Alumini 2021) that proposes a way to perform the so-called “virtual choir”. It provides instructions on how each part is recorded and uploaded. However, they still require the post-process of the collected videos to calibrate the time and combine them into a whole choir song, which is still a big task that becomes the major problem of being automatic and even real-time.

Thus, the platform of virtual choiring needs real-time performance especially as timing is critical in signing and practicing. However, previous ways to perform online choiring used a current online conferencing platform (The Harvard Choruses 2021). For example, Zoom provides acceptable quality for talking. However, when it comes to music, certain notes, especially high notes from instruments are often filtered and muted. Moreover, it appears to have a delay of the sound. This gives a bad restoration of the original performance. For these challenges, the following technological views need to be considered.

High notes often disappear in existing platforms, as they seem to be filtered either by its compression and decompression process or by the noise reduction operation. Such disappearance of certain frequencies of the original music is not acceptable for virtual choirs. In sound quality, the dynamic range needs to be considered, too. The existing conferencing platforms use the dynamic ranges that are decided by the platforms for compromising the quality and transmission speed. This sometimes causes the delay related to the whole path of the whole process (A. Carôt 2009). Significant delays in the whole path of digital audio signals are caused by input-output blocking and driver buffering and transmission (Fig. 1).

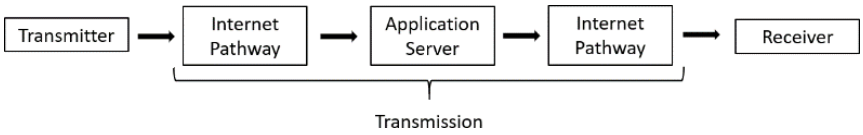

Fig. 1. The whole path of the transmission of digital audio signals.

The transmission depends on internet speed and the server's processing ability since it goes through the Internet pathways. The speed of the internet pathways depends on the Internet bandwidth and its speed, while that of the application server relies on the processing unit. As a scheduler in a computer system schedules a certain moment to process the data, it takes time to transfer audio data to the userspace. Therefore, the system collects data and processes them together, which is called device blocking. The number of collected data for the process is called a block size. The block size depends on the performance of a users' system (Jean 2013). The blocking implies a trade-off between stability and delay. A larger block requires less computational power with more stable system behavior. However, the delay increases as it needs time to generate an appropriate number of data. Increased audio sampling rate leads to better audio quality but causes dealy.

Audio encoding also affects the delay. It is the process of compressing audio data for a better throughput rate in transmission. For different encoding methods, researchers have developed several audio coding formats. Table 2 shows audio coding formats using voice over internet protocol (VoIP). Among the formats, OPUS is a common-used coding format that Skype adopted as it has the lowest delay compared to other codecs. OPUS has a low delay and supports a large range of bit rates. Thus, it has many different sound qualities to be selected for different applications.

Table 2. Audio coding formats commonly used for VoIP platforms.

Skype	OPUS (6kbps - 510kbps) SILK (6kbps - 40kbps)
Messenger	iSAC (10kbps - 52kbps)
Others	G.726 (16 kbps, 24 kbps, 32 kbps, 40 kbps) G.729 (8kbps) G.723 CELP CELT

By considering the problems of existing online platforms for video conferencing, we carried out a study for proposing an appropriate way of virtual choiring that overcomes the problems of the platforms. Therefore, focusing on online learning, we aim to propose online chorus practicing in this study. The proposed method provides an effective way of virtual choiring.

2. Methods and Experiments

2.1. Methods

Architecture

We used OPUS as the hardware architecture (Xiph 2013) through the high-level overview of OPUS that is the hybrid of constrained energy lapped transform (CELT) and SILK. The proposed system switches to one of them depending on the situation and application, which the virtual choring operates at a wide range of bitrates with popular VoIP codecs. As the delay of OPUS depends on the frame size, the algorithmic delay is equal to the summation of frame size and lookahead theoretically. The lowest delay that OPUS achieves is 5ms (2.5ms each for frame size and lookahead). A small sampling interval has a smaller delay. Also, the "Core Audio" of the Apple Macintosh operating system OS X requires the so-called "safety offset" of at least 64 audio samples up to even 292 audio samples due to its general architecture. In the "Core Audio", the CPU load increases as the buffer size decreases and the latency decreases as the buffer size decreases. For the lowest CPU load and the lowest latency, a trade-off needs to be made.

Data Rates

Table 3 shows a list of data rates in current and future wireless communication systems. A communication system with higher bandwidth assures availability. The other way that people often use is to add more servers to make sure that each server's loading does not become a bottleneck for the whole performance, which gaming companies often use when the number of players is increasing.

4G LTE (long-term evolution) is the fourth-generation standard for mobile wireless broadband communication. It is being gradually replaced by its successor 5G that allows better performance in speed and latency to help emerging applications such as the internet of things (IoT), self-driving cars, artificial intelligence (AI), and big data. Upon the rise of 5G commercialization, the 6G whitepaper is under drafting. Each generation stays for about a decade and speeds up about 10 to 100 times faster than the previous one, fulfilling lower-latency applications. In contrast to mobile wireless communication, wireless access points are used in wireless local area networks (WLAN also known as Wi-Fi), which is based on the IEEE 802.11 standard. The Wi-Fi version of the current mainstream is IEEE 802.11ac (also known as Wi-Fi 5), coexisting in the market with the rising, 2019- released Wi-Fi 6. From Wi-Fi 5 to the future successor Wi-Fi 7 (IEEE 802.11be), each generation improves in terms of bandwidth and latency, just like the mobile cases.

Table 3. A list of current and future wireless communication technologies.

4G-LTE	Theoretical max downlink rate: 100Mbps Theoretical max uplink rate: 50Mbps
	Latency: 10–100ms
5G	The minimum requirements for downlink peak data rate: 20 Gbps The minimum requirements for uplink peak data rate: 10 Gbps Target downlink “user experienced data rate”: 100 Mbps Target uplink “user experienced data rate”: 50 Mbps Latency: 4ms
6G	Latency: 0.1ms Expected data rate: 100~1000 Gbps “About 100 times faster than 5G” Will start to draft in 2026 Will be commercially launched in 2030
WiFi-6	Theoretical maximum data rate: 9.6 Gbps
WiFi-7	Theoretical maximum data rate: 30 Gbps Commercially launched no sooner than 2024

Necessary Functions

(HD) videos: High definition (HD) video has 1280 x 720 pixels abbreviated in 720p. Since in a choir practicing, for everyone being able to see the choir conductor’s hand movement is an essential need. Thus, the virtual choiring system provides the functionality of video capturing then playing it live on users’ screens. As the minimum requirement is to see the hand movement of the choir conductor, lower resolution is acceptable, such as 480p.

Good audio quality (maybe 128 kbps for music, 16kbps for speech): In virtual choir practicing, hearing other member’s voices, and deciding when to chime are basic needs. Therefore, providing good audio quality for users is necessary. From practical uses in audio applications and a decent, acceptable audio file, a minimum data rate of 16 kbps and 128kbps is required for speech or music, respectively. Thus, these values were used for system development. The value of sampling and data rate may be changed to any value based on the application and the need for audio quality.

Multiple participants: Similar to a video conferencing platform, a virtual choir platform serves multiple participating users simultaneously. The system is required to accommodate multiple users at the same time.

Host control: Similar to video conferencing platforms, host control for virtual choir needs to have good functionality. A choir conductor is an essential person in a choir. Being able to mute or add a certain group of users’ voices is important so the system needs to provide it.

Screen Sharing: Just as in Skype, sharing the current screen of the user makes it easier for others to see what the current user is doing in their platform. This makes communication much easier in some cases other than just try to describe it in words.

Document Sharing: Aside from screen sharing, document sharing is also needed since the sharing screen is only for seeing other users on the platform. Document sharing allows file exchange.

Screen/Audio Recording: The files may be videos or audio. Each user needs to record their parts and send them to others for teachings and mutual interactions.

Real-time Volume Weighting for Different Choir Groups: When it comes to the choir conductor, aside from being able to mute or add a certain group of voices into the real-time playing, changing the volume of each group at the same time is required. Thus, a basic real-time online mixing functionality is required for the choir conductor to adjust if certain groups are recording in high volume.

Offline Mixing: The choir conductor needs to mix even when the practice is over. The conductor is allowed to mix recorded parts after practicing.

A Good User Interface and A Touchscreen: A good user interface with a touchscreen is needed so that the user uses the functions easily and benefits from the virtual choiring system.

2.2. Experiments

In the experiment, the network is chosen for the transmissions part to be a control variable that is fixed during different experiments. The dependent variables include operating systems of transmitting or receiving devices and the sound card each possesses. The importance and role of input blocking, driver buffering, output blocking, and driver buffering are investigated during the experiment.

To understand the unknown factor in the conferencing platforms and test the proposed hypothesis, we conducted experiments. Figure 2 shows the system that was used for the experiment.

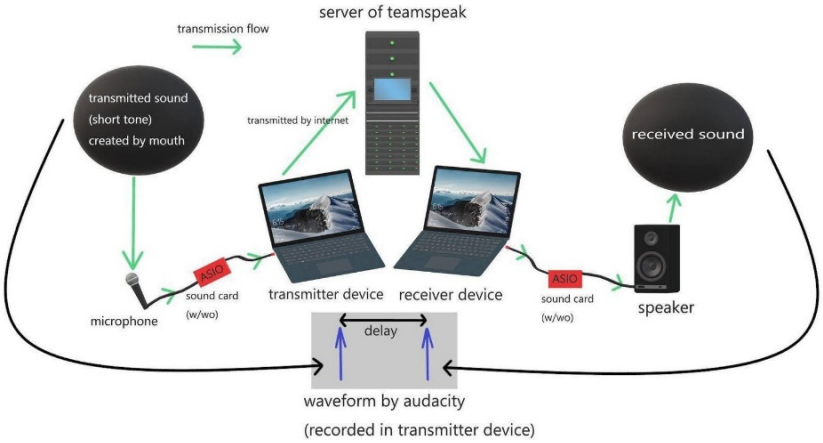

Fig. 2. The experiment setup in this study.

3. Results and Discussions

3.1. Total Delay Time of Conferencing Platforms—Skype and Zoom

We investigated if there were huge delays in the conferencing platform. Thus, we set up a testing system that comprised a PC, a laptop, or a mobile phone (Table 4).

Table 4. The specifications of testing devices used in experiments.

	Audio Device	**Network Card**	**CPU**	**Internet Speed by NTU Speed (Mbps) (Download/Upload)**
PC	Realtek high definition audio	Realtek PCIe GbE Family Controller	Intel (R) Core (TM) i5-9400F CPU @ 2.90GHz	58.11/18.81
Laptop (ASUS UX430u)	Realtek high definition audio	Intel (R) Dual Band Wireless- AC 8265	Intel (R) Core (TM) i5-8250U CPU @ 1.60GHz	10.05/46.57
Mobile (ASUS ZS551KL)	Qualcomm Snapdragon 835		Qualcomm Snapdragon 835	63.17/37.28

Then, we connected a microphone for checking the sound sharing function, as playing a 1 Hz triangular sound wave. In the receiver, the microphone was turned off with the music sound on to hear the sound wave received from the sound sharing function from the source. At the same time, we recorded two sound files into the same file and analyzed the delay time by observing the difference between the crest and trough of the sound wave. The result showed that Skype was faster than Zoom to receive the sound wave in all circumstances (Table 5, Fig. 3). Using mobile caused latency since the hardware computing power was inferior to a PC or a laptop.

Table 5. The delay time of using the sound sharing in Skype and Zoom (in second).

Source \ Destination	PC	Laptop	Mobile
PC		0.302/1.121	0.742/1.005
Laptop	0.345/0.42		0.686/1.085

We did not test the mobile phone as the source as the mobile version of Skype and Zoom did not support sound sharing.

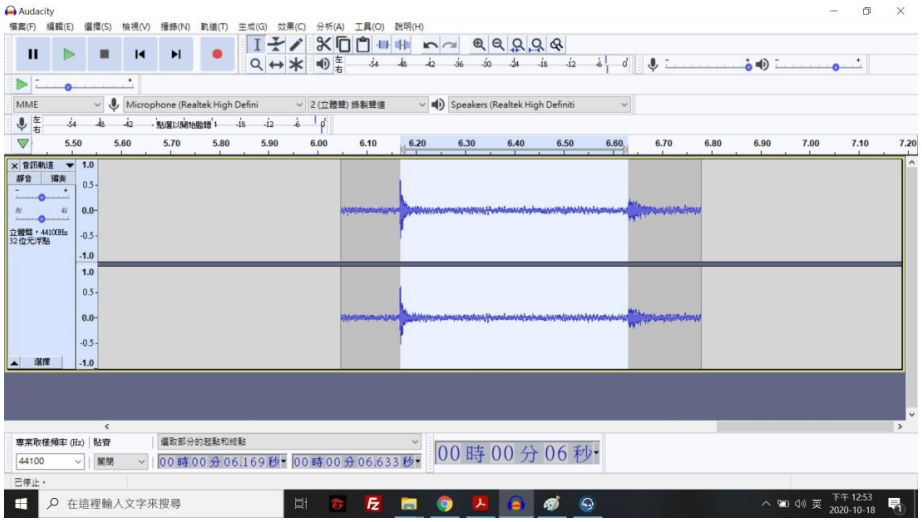

Fig. 3. Computing delay through audacity.

3.2. Delay Time Tests of a Protocol Application Software—TeamSpeak 3

We tested if the handmade platform achieved a shorter delay time result than the existing platforms, such as Skype and Zoom. In this case, we chose a VoIP application that is called "TeamSpeak 3" (Teamspeak 2021) to build a simulating system. Audio stream input/output (ASIO) is the protocol for a digital audio sound card driver, providing low latency and high fidelity between software and hardware. How the sound driver technology affected the overall delay of the sound transmission in TeamSpeak 3 was tested with the delay time on the windows and mac OS (Tables 6 and 7). The delay time was tested with the sound card using ASIO technology.

Table 6. The delay time of TeamSpeak 3 with Windows.

		Test1	Test 2
Transmitter	Hardware	Windows laptop w/o ASIO	
	Internet	NCTU Wi-Fi	
Receiver	Hardware	Android mobile phone w/o ASIO	
	Internet	NCTU Wi-Fi	
VoIP server		VoIP server 1	VoIP server 2
Overall delay time		0.449 (s)	0.384 (s)

Table 7. The delay time of TeamSpeak 3 with mac OS.

		Test1	Test 2	Test3
Transmitter	Hardware	MacBook w/o ASIO		
	Internet	NCTU Wi-Fi		
Receiver	Hardware	Android mobile phone w/o ASIO	MacBook w/o ASIO	
	Internet	4G LTE	NCTU Wi-Fi	
VoIP server		VoIP server 1		VoIP server 2
Overall delay time		0.342 (s)	0.143 (s)	0.134(s)

3.3. Delay Test for Windows with On-Board Sound Card

On Windows OS, the default sound driver does not use ASIO technology. Thus, the test was carried out without ASIO. We chose two VoIP servers and a default channel. Having the source (a laptop on windows OS) and the receiver (a mobile phone) in the same room, the source made a monotone sound. Then, in the receiver, we turned off the microphone and run the sound player which played the voice from the source. At the same time, an audio analyzing platform called "Audacity" recorded both sounds. Then, we computed the time difference. For tests 1 and 2, we used a Windows laptop and Android mobile phone as the transmitter and receiver. The only difference between the two tests was different VoIP servers (VoIP servers 1 and 2). The specifications were detailed in Table 8.

Table 8. The specification of VoIP servers 1 and 2.

	VoIP Server 1	VoIP Server 2
CPU	Intel Xeon Silver 4214 2.2GHz 12 core×2	Core i3-7350k-4.2GHz×1
Memory	192GB DDR4	ADATA DDR4 2400 16G RAM
Hard Drive×2	2.5" 1.2TB 10k rpm SAS hot swap drive	Micron Crucial MX300 275GB SSD×1

The VoIP Server 2 had a better CPU and RAM than the VoIP Server 1. However, the improvement in delay time was not significant, which proves the speculation that the computing power of CPU and RAM are not critical factors in the delay time. We set the CODEC of TeamSpeak3 in OPUS Music at Quality 10 (in the scale of 1–10), which decides the level of compression rate.

3.4. Delay Time for Mac OS with On-Board Sound Card

For the three tests, we used a MacBook without ASIO as the transmitter with NCTU Wi-Fi as the Internet. The only difference was the choice of receiver and VoIP server as shown in Table 8. From tests 1 and 2, changing the receiver from android mobile to mac shortened the delay time significantly. Therefore, one of the major reasons for the delay time was the sound driver technology. The average delay in test 3 was slightly shorter than that in Test 2, which is consistent with the fact that the hardware affects the delay time. However, the difference in the time delay between tests 2 and 3 was not large due to the inevitable measurement error.

3.5. Delay Test for Sound Card with ASIO Technology

For tests 1 and 2, we used a Windows laptop connected with ASIO on the Internet of NCTU Wi-Fi as the transmitter and chose VoIP Server 2 as the application server. We selected a Windows PC with ASIO on the Internet of NCTU cable wire, an iPhone without ASIO on the Internet of NCTU Wi-Fi as the receiver of tests 1 and 2, respectively. The result of Test1 showed that the average delay was 6.3ms, which is almost the speed of sound transmission. Thus, we concluded that the main bottleneck in the overall delay of sound transmission on TeamSpeak 3 was due to different sound driver technology. Since "BlasterX G5" is a sound card with ASIO technology, changing the receiver from PC with BlasterX G5 to an iPhone made the delay longer, since the hardware of the sound driver in iPhone was worse than the "BlasterX G5" sound card.

Table 9. The delay time of TeamSpeak 3 by introducing the ASIO technology.

		Test1	Test 2
Transmitter	Hardware	Windows laptop w/ ASIO	
	Internet	NCTU Wi-Fi	
Receiver	Hardware	Windows PC w/ ASIO	iPhone w/o ASIO
	Internet	NCTU cable wire	NCTU WIFI
VoIP server		VoIP server 2	
Overall delay time		0.006 (s)	0.250 (s)

3.6. Compression Rate Test of TeamSpeak 3 on OPUS Music at Quality 10

To test the compression rate of the TeamSpeak simulation platform, we compared the data size difference in two sides, having audacity recording and speaking a sentence on TeamSpeak. The uncompressed audio file size was 14.6 MB, and the total bytes transferred during the TeamSpeak test was 1.01 MB. TeamSpeak compressed the file to 6.88% of the original size to send out.

3.7. Bandwidth Usage Rate Test of TeamSpeak 3

To test the bandwidth usage rate, we used the software "iPerf" in the full bandwidth of the system. The source had about 19 Mbits/s in a cabled Wi-Fi and about 28 Mbits/s in 4G Wi-Fi (Table 10). The data transmitting rate on TeamSpeak 3 was about 88 Kbits/s in both cases. The result revealed that TeamSpeak did not use the full bandwidth to transmit. On the contrary, the bandwidth usage rate was about 0.1%.

Table 10. The bandwidth of source and receiver in the cabled Wi-Fi.

Cabled Wi-Fi	**Bandwidth**
source	19.1 Mbits/sec
receiver	18.9 Mbits/sec
4G wireless Wi-Fi	Bandwidth
source	28.4 Mbits/sec
receiver	28.4 Mbits/sec

4. Conclusions

This paper reviewed and tested the possible problems in a virtual choir system. Under the common computer specifications, sound cards contributed the most to the delay time. Applying ASIO achieved a 0.006ms lower delay in the virtual-choiring experiments. The results imply that using this ASIO provides good performance in overcoming the delay in a virtual choiring system which is a difficult task in an existing communication system. The compression rate of TeamSpeak 3 on OPUS Music at Quality 10 was 6.88%. The bandwidth usage of the TeamSpeak 3 test was about 88 Kbits/s. The results indicate that the audio transmission in a real-time virtual choir does not require a high-bandwidth internet. However, according to the reported speed performance in real user environments, the 4G mobile network exhibits an overall download speed of no more than 40 Mbits/s (Sam 2020), which barely meets the data rate required for low-delay video transmission. On the other hand, the 5G mobile provides superior user experiences at a typical download speed of about 250Mbits/s (Ian 2020) and serves as a good network of low-delay video transmission.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Anderson, Terry, ed. The theory and practice of online learning. Athabasca University Press, 2008.
2. Angeli, Charoula, Nicos Valanides, and Curtis J. Bonk. "Communication in a web-based conferencing system: The quality of computer-mediated interactions." British Journal of Educational Technology 34.1 (2003): 31-43.
3. Zhou, Lina, Liwei Dai, and Dongsong Zhang. "Online shopping acceptance model-A critical survey of consumer factors in online shopping." Journal of Electronic commerce research 8.1 (2007).
4. OWL Labs. March 2, 2020. "The 10 Best Free Video Conferencing Tools to Choose From" OWL Labs. https://www.owllabs.com/blog/video-conferencing-tools
5. The Harvard Choruses. 2021. "Choiring During Quarantine." The Harvard Choruses. Accessed Jan. 18, 2021. https://www.singatharvard.com/choir-during-quarantine.html
6. Harvard Glee Club Alumni. 2021. "HGC ALUMNI VIRTUAL CHOIR - '00S" Harvard Glee Club Alumni. Accessed Jan. 18, 2021. https://hgcalumni.org/virtual-choir-00s
7. A. Carôt, ''Musical Telepresence – A Comprehensive Analysis Towards New Cognitive and Technical Approaches,'' Ph.D. dissertation, Inst. Telematics— Univ. Lübeck, Lübeck, Germany, 2009.
8. Jean-Marc Valin, Gregory Maxwell, Timothy B. Terriberry, Koen Vos. "High- Quality, Low-Delay Music Coding in the Opus Codec" October 17–20, 2013. (PDF). www.xiph.org. New York, USA: Xiph.Org Foundation. p. 2.
9. Xiph.Org Foundation, "The Opus Codec" 135th AES Convention 2013 October 17–20 New York, USA
10. Teamspeak. 2021. "Home | TeamSpeak" Teamspeak. Accessed Jan. 18, 2021. https://www.teamspeak.com/en/
11. Sam Fenwick. 2020. "TAIWAN Mobile Network Experience Report December 2020" Opensignal. Accessed Mar. 23, 2021. https://www.opensignal.com/reports/2020/12/taiwan/mobile-network-experience
12. Ian Fogg. 2020. "TAIWAN 5G User Experience Report December 2020" Opensignal. Accessed Mar. 23, 2021. https://www.opensignal.com/reports/2020/12/taiwan/mobile-network- experience-5g